Dual-3DM^3−AD: Mixed Transformer Based Semantic Segmentation and Triplet Pre-Processing for Early Multi-Class Alzheimer's Diagnosis

Arfat Ahmad Khan, Rakesh Kumar Mahendran, Kumar Perumal, *Member, IEEE*, and Muhammad Faheem

Abstract—Alzheimer's Disease (AD) is a widespread, chronic, irreversible, and degenerative condition, and its early detection during the prodromal stage is of utmost importance. Typically, AD studies rely on single data modalities, such as MRI or PET, for making predictions. Nevertheless, combining metabolic and structural data can offer a comprehensive perspective on AD staging analysis. To address this goal, this paper introduces an innovative multi-modal fusion-based approach named as Dual-3DM^3-AD. This model is proposed for an accurate and early Alzheimer's diagnosis by considering both MRI and PET image scans. Initially, we pre-process both images in terms of noise reduction, skull stripping and 3D image conversion using Quaternion Non-local Means Denoising Algorithm (QNLM), Morphology function and Block Divider Model (BDM), respectively, which enhances the image quality. Furthermore, we have adapted Mixed-transformer with Furthered U-Net for performing semantic segmentation and minimizing complexity. Dual-3DM^3-AD model is consisted of multi-scale feature extraction module for extracting appropriate features from both segmented images. The extracted features are then aggregated using Densely Connected Feature Aggregator Module (DCFAM) to utilize both features. Finally, a multi-head attention mechanism is adapted for feature dimensionality reduction, and then the softmax layer is applied for multi-class Alzheimer's diagnosis. The proposed Dual-3DM^3-AD model is compared with several baseline approaches with the help of several performance metrics. The final results unveil that the proposed work achieves 98% of accuracy, 97.8% of sensitivity, 97.5% of specificity, 98.2% of f-measure, and better ROC curves, which outperforms other existing models in multi-class Alzheimer's diagnosis.

Index Terms—Alzheimer's diagnosis, multi-modalities, MRI, PET, semantic segmentation, mixed transformer, multi-scale feature extraction.

I. Introduction

ALZHEIMER'S disease, an inexorable and series neurological problem, causes brain shrinkage and ranks among the most prevalent causes of mortality in the elderly population [1], [2], [3]. It progressively erodes memory and cognitive faculties, eventually rendering even the simplest tasks insurmountable, disrupting daily life [4]. The primary culprit behind the disease is the accumulation of abnormal proteins in and around brain cells [5]. Amyloid protein aggregates to form plaques around the brain, while tau protein forms tangles within. Diagnosing Alzheimer's disease can be challenging, especially in older individuals [6], [7]. Consequently, Magnetic Resonance Imaging (MRI) helps medical professionals in the detection of this illness. Image analysis stands out as a prominent method for diagnosing Alzheimer's disease, as modern medical imaging equipment yields a plethora of data about the under-examination patient. T1-weighted structural MRI scans and 18F 2-Fluoro-2-deoxy-D-Glucose Positron Emission Tomography (FDG-PET) offer spatial insights into atrophy and hypometabolism, respectively [8], [9], [10], [11].

The pathophysiological processes behind Alzheimer's disease inflict damage upon brain tissues and disrupt their normal metabolic functions [12]. FDG-PET can pinpoint areas with impaired functions by visualizing metabolic irregularities. The regional hypoperfusion/hypometabolism, particularly in biparietal and bitemporal distributions, strongly correlates with the clinical detection of the disease [13], [14]. PET scans are capable of identifying diseases even before the emergence of discernible symptoms or warning signals by scrutinizing biological functions through metabolic processes [15]. Similarly, MRI scans can gauge variations in the volume of recognizable brain regions, allowing the observation of the gradual brain atrophy caused by AD-related neurodegeneration [16]. This atrophy is attributed to losses in dendrites and neurons. The atrophy measurements from MRIs can be employed to estimate cumulative neuronal damage, as there

exist a robust correlation between atrophy and cognitive decline [17], [18], [19].

Detecting Alzheimer's disease with the help of MRI images involves many key stages, such as pre-processing, extraction of features, segmentation, and classification. In the initial stage of pre-processing, MRI images undergo essential adjustments to address their susceptibility to noise as well as non-brain tissue existences (such as the skin, scalp, dura, muscles, fat, eye, etc.) [20], [31], [32]. It is worth noticing that some previous studies omit skull stripping and overlook noise reduction (including salt and pepper noise, Gaussian noise, and Rician noise, etc.), ultimately compromising their classification accuracies. To enhance the classification accuracy and computational efficiency, segmentation follows pre-processing. Segmentation is a crucial process that involves distinguishing the cerebrospinal fluid, white matter, and gray matter, yielding essential information for subsequent categorization [33]. Interestingly, some prior research neglects segmentation altogether, while many rely on automated image analysis tools like Statistical Parametric Mapping (SPM), FreeSurfer, and FSL-FAST4 [34]. However, the use of such automated tools can substantially increase the computation time, potentially impacting the efficiency of the segmentation process [35]. It is important to highlight that the automated methods for estimating volume yield inaccurate results without the proper validation. Automated tools often rely on intensity comparisons with atlases to guide the segmentation process, which can introduce potential errors and complexities in the analysis [36].

The prevailing approach in current research involves employing deep learning-based methods with the aim of classification and extraction of useful features. However, these algorithms typically extract only individual features or small datasets, which proves to be insufficient in terms of classification in an accurate way. The existing studies draw upon a repertoire of techniques, containing Machine Learning (ML), neural networks, and Deep Learning (DL) [37]. ML methods including K-nearest neighbours, decision trees, SVM and random forests are frequently utilized. However, their training complexity tends to increase due to the generation of many trees during the extraction of features, and these methods do not perform well in terms of handling extensive datasets. On the other hand, deep learning, which relies on neural networks for classification and the extraction of features, encompasses various models like convolutional neural networks, multilayer perceptrons, and radial basis functions. Deep learning surpasses the shortcomings of conventional ML methods. However, this approach often involves numerous hidden layers, substantial convergence weights, and extended computation times, leading to the heightened complexity and a potential reduction in classification accuracies [38], [39], [40]. To address these challenges, researchers have turned to Mixed transformer-based semantic segmentation to overcome the hurdles faced by automated tools during the segmentation process. Additionally, a multi-scale feature extraction with an effective Dual-3DM^3−AD architecture has been employed to mitigate the issues arising from high complexity and elevated false positive rates encountered during the feature extraction.

Research Contribution: The diagnosis of Alzheimer's disease faces several notable drawbacks, particularly in the context of neuroimaging and image analysis. Alzheimer's, a relentless and debilitating neurological condition, is marked by significant challenges in its diagnosis. MRI and PET scans have become integral tools for identifying the disease, and they are not without limitations. One significant drawback is the high cost and resource-intensive nature of these imaging techniques, making them less accessible for many patients and healthcare facilities. Furthermore, these methods primarily provide structural or metabolic insights into the brain, often lacking the ability to diagnose the disease in its early stages when structural changes may not yet to be apparent. Additionally, the process of image analysis, involves pre-processing, segmentation, and classification, is susceptible to errors and variations. Although automated tools are convenient, they can compromise accuracy and introduce complexities. The prevailing use of neural networks, machine learning, and deep learning methods exhibits good performances. However, they often demand substantial computational resources, resulting in the increased complexity and potentially reduced diagnostic accuracy. These challenges highlight the need for ongoing research and the development of more accessible and precise diagnostic methods for Alzheimer's disease. Henceforth, we focus on an accurate and earlier Alzheimer diagnosis using multi-modalities. To achieve this, we have contributed several novelties explained as follows:

- This paper introduces a novel approach that combines multiple data modalities, specifically MRI and PET scans, to enhance Alzheimer's Disease (AD) diagnosis. This fusion-based approach offers a holistic perspective on AD staging analysis.
- The research incorporates advanced preprocessing techniques, including noise reduction, skull stripping, and 3D image conversion, achieved through the QNLM, Morphology function, and BDM. These processes significantly enhance the quality of the image data, ensuring more reliable analysis.
- To reduce complexity and improve the accuracy of the analysis, the study employs a Mixed-transformer with Furthered U-Net architecture for semantic segmentation. This step aids in identifying and isolating relevant regions within the images.
- Dual-3DM3-AD model includes a multi-scale feature extraction module, which extracts pertinent features from both segmented images. This module ensures that the critical information from images is effectively captured. The extracted features are then aggregated using the DCFAM. This aggregation process maximizes the utilization of information from both MRI and PET scans, further enhancing the accuracy of the diagnosis. The multi-head attention mechanism helps to reduce the feature dimensionality. This step actually aids to streamline the data, while retaining essential information.

II. Literature Survey

The prevalence of big data analytics and the enhanced computational power offered by GPU clusters have firmly established Deep Learning (DL) as a prevalent and influential technique, extending its reach into numerous domains. Presently, it has become common to leverage DL models

for various recognition applications in the realm of medical image analysis. In recent times, researchers have increasingly turned to MRI and PET modalities to embrace DL for the development of Alzheimer's Disease (AD) diagnosis models. Remarkably, a high-resolution T1-weighted MRI scan possesses the capability to identify atrophies in distinct brain regions by providing critical structural insights into the brain. Wei et al. [21] explore the application of Bi-directional Empirical Model Decomposition (BEMD) for the automated detection of Alzheimer's disease. BEMD, a signal processing technique, is employed for feature extraction from medical data. This approach leverages BEMD's potential in revealing hidden patterns in multi-modal data sources to enhance the early diagnosis of Alzheimer's disease. The novelty of this work lies in its innovative application of BEMD for the automated Alzheimer's disease detection, potentially improving diagnostic accuracy. Zaina et al. [22] introduce a novel feature extraction method called Exemplar Pyramid for Alzheimer's disease classification. The study focuses on extracting discriminative features from neuroimaging data, particularly MRI scans, to aid in the accurate detection of Alzheimer's disease. The innovation lies in its novel approach of utilizing exemplar pyramid feature extraction, which enhances the accuracy and effectiveness of Alzheimer's disease classification. Basheera et al. [23] present a classification method for Alzheimer's disease based on Convolutional Neural Networks (CNNs), and the enhanced Independent Component Analysis (ICA) is applied to segmented gray matter in MRI images. By combining deep learning and feature extraction from MRI scans, this study aims to advance the accuracy and efficiency of Alzheimer's disease detection. The paper's contribution lies in introducing a novel Alzheimer's disease classification method that combines CNN and hybrid enhanced ICA segmentation, improving the accuracy of diagnosis using MRI data. Murugan et al. [24] propose a deep learning model for the early diagnosis of Alzheimer's disease and dementia using MR images. This research leverages the power of deep neural networks to automatically extract relevant features and classify patients based on neuroimaging data. The aim of this work is the development of a deep learning model for early and accurate diagnosis of Alzheimer's disease and dementia, potentially advancing early intervention and treatment. Febietti et al. [25] delve into early detection by utilizing cortical and hippocampal Local Field Potentials (LFPs) and ensemble machine learning models. By incorporating electrophysiological data, this study explores an alternative approach to Alzheimer's disease detection. The contribution of this work is the development of an ensemble machine learning approach for early Alzheimer's disease detection using neural signals, potentially advancing early diagnosis and intervention.

Dwivedi et al. [26] focus on the development of a multimodal fusion-based deep learning network for the effective diagnosis of Alzheimer's disease. It addresses the importance of integrating data from various sources, such as neuroimaging, genomics, and clinical assessments, to enhance diagnostic accuracy. Yu et al. [27] explore the application of high-order pooling and Generative Adversarial Networks (GANs) for assessing Alzheimer's disease. The research introduces innovative techniques for feature extraction and data representation

by tensorizing GANs. The approach aims to improve the accuracy and efficiency of Alzheimer's disease assessment using advanced data manipulation. The effectiveness of this paper is the innovative integration of high-order pooling and GAN techniques to enhance the assessment of Alzheimer's disease, potentially improving diagnostic accuracy and early detection. Song et al. [28] delve into the application of the Random Forest algorithm for diagnostic classification and biomarker identification in Alzheimer's disease. It emphasizes the importance of interpretable machine learning methods in uncovering relevant biomarkers for diagnosis. Bron et al. [29] investigate the generalizability of machine learning models for Alzheimer's disease diagnosis across different cohorts. It addresses the challenge of model transferability by examining the performance of deep learning and conventional machine learning models on diverse datasets. The effectiveness of this research is demonstrated through its robust ability to generalize and accurately diagnose Alzheimer's disease across multiple cohorts, showcasing its potential for broad clinical application. Etmanani et al. [30] introduce a 3D deep learning model for predicting the diagnosis of various neurodegenerative disorders, including dementia with Lewy bodies, Alzheimer's disease, and mild cognitive impairment. The use of brain 18F-FDG PET scans and deep learning techniques underscores the potential of non-invasive imaging in early diagnosis and differentiation of these conditions. The effectiveness of this work is evidenced by its accurate prediction of various neurodegenerative conditions through the analysis of 3D PET scans, providing valuable diagnostic support.

III. Dual-3Dm^3-AD Framework

In this study, we primarily concentrate on the detection of Alzheimer's disease with the help of mathematical modelling. With the help of pre-processing, extracting features, and segmenting, the suggested approach increases the classification accuracy. We use the Alzheimer's Disease Neuroimaging Initiative (ADNI) database's T1-weighted MRI and PET images. The three phases of the proposed work are as follows:

A. Data Acquisition

In this research, we have utilized neuroimaging data acquired from Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset (https://www.kaggle.com/datasets/madhucharan/alzheimersdisease5classdatasetadni). The main intention of ADNI team is the neuropsychological calculation for evaluating the improvement of MCI to initial AD and for AD supplemented via research of resultant of combined several biomarkers, utilizing Cerebos Spinal Fluid (CSF) data, MRI and PET. The cases are chosen from ADNI dataset cohort to our experiment prerequisite, having the visit of both consequent and screening. The cases age ranges from 55 to 89 years old, containing both female and male. We chose 100 normal, 100 MCI and AD cases. For every case, the 18-FDG-PET images and T1-weighted MRI are adapted in this research. Here, the PET images are obtained by the constructor model of SIEMENS along with 2.4mm slice thickness. For that, the radiopharmaceutical 18F-FDG is utilized which consists of

63 slices. Besides, MRI images are acquired by 1.5 T scanners. The slice thickness is 1.2mm with 160 slices, where the size of each slice is 192×192 of 3D images.

B. Data Pre-Processing

The pre-processing approach is optimally prejudiced by the consequent processing algorithm with image format defined as:

1) Noise Reduction: Initially, the noise present in both MRI and PET scans is removed for enhancing image quality. To do, we have utilized Quaternion Non-local Means Denoising Algorithm (QNLM). As the QNLM denoising technique leverages the inherent high-degree self-similarities within images for noise suppression, the choice of a similarity metric among image patches plays a pivotal role in the algorithm's noise reduction effectiveness. We have introduced a novel approach by replacing the traditional Euclidean distance with the QNLM technique as a metric for evaluating similarities between image patches. Meanwhile, the image information constantly contains certain repeatability, as self-resemblance forms during the distribution of noise is arbitrary. Hence, the target of QNLM is to make utilize of self-resemblance forms to overwhelm the noise. Henceforth, the QNLM improves the denoising process from the level of pixel to patch. The noisy MRI image is modeled as $\mathfrak{Y} = \mathfrak{X} + \mathfrak{N}$, and then the denoised image $\hat{\mathfrak{X}}$ by QNLM is mathematically expressed as:

$$\hat{\mathfrak{X}}_{(\rho)} = \frac{\sum_{q \in \delta_\rho} \varpi\,(\rho, q) \times \mathfrak{Y}\,(q)}{\sum_{q \in \delta_\rho} \varpi\,(\rho, q)} \tag{1}$$

where δ_ρ is denoted as the search window along with center ρ, and the weight $\varpi\,(\rho, q)$ is defined as:

$$\varpi\,(\rho, q) = \exp\left(-\frac{\eth\,(\rho, q)\,/\alpha_n^2}{\mathfrak{h}^2}\right) \tag{2}$$

Here, $\eth\,(\rho, q)$ indicates the Euclidean distance among two image patches along with center ρ and q in δ_ρ. Likewise, the PET image scans are denoised for image betterment.

2) Skull Stripping: Following denoising, the skull stripping is performed by utilizing morphology. The skull stripping is a preprocessing step performed in Alzheimer's disease diagnosis using brain imaging techniques, such as MRI and PET scans. It involves the removal of non-brain tissues, including the skull, scalp, and other extraneous structures, from the acquired images. This step is crucial because it helps isolate the brain region of interest, reducing noise and interference caused by surrounding tissues. By effectively stripping away non-brain elements, the processes of subsequent image analysis and feature extraction become more accurate, allowing for a clearer focus on the brain's structural and metabolic changes associated with Alzheimer's disease. The skull stripping enhances the overall quality of the images and aids in the reliable and precise detection of Alzheimer's-related abnormalities. For this purpose, the proposed technique is mathematically integrated with Erosion and Dilation operators. Furthermore, the proposed technique utilized global thresholding continued by morphological functions. The thresholding value is evaluated as per intensity distribution knowledge of brain scans. Initially, the image ($\mathfrak{I}$) is read, and RGB is converted as

grayscale profile ($\mathfrak{I}_1$). Here, the grayscale scan is eroded ($\mathfrak{I}_2$) by structuring element of disk-handed (x) in size 4 that is continued by Dilation ($\mathfrak{I}_3$) of outcome image utilizing same structuring element (x). By adapting thresholding scheme, the acquired image is then binarized ($\mathfrak{I}_4$). The acquired binary image is transmuted to unit of 8 format ($\mathfrak{I}_5$) and that is subtracted ($\mathfrak{I}_6$) from the grayscale profile comprising skull portion alone. By subtracting the image of ($\mathfrak{I}_7$) from grayscale, the skull portion is removed and then, the region of brain is acquired, which is written as:

$$\mathcal{E}\,(f) = f \oplus x = \left\{ \gamma \,|\,(x)_z \cap f^{\complement} = \varnothing \right\} \tag{3}$$

$$\mathfrak{D}\,(f) = f \oplus x = \left\{ \gamma \,|\,(x)_z \cap f = \varnothing \right\} \tag{4}$$

3) 3D Image Conversion: As 3D image facilities a better navigation in terms of multiple perspectives, we transfigured the images to 3D with the skull stripping. As 3D images allow us to navigate from multiple perspectives in the quest for skull stripping, the transformation of two-dimensional (2D) MRI scans into three-dimensional (3D) images is undertaken. This transformation is driven by the inherent limitation of 2D images, which provides a flat and single-perspective view, while 3D images enable navigation from multiple angles, offering richer and more diverse viewpoints. To achieve these enhanced 3D images, a Block Divider Model (BDM) is employed, significantly reducing the time required to obtain precise depth details by segmenting the 2D images into blocks. The process begins with the creation of a depth map through node and link formation. During the conversion from 2D to 3D images, the depth gradient hypothesis assigns depth values to individual blocks. This hypothesis encompasses depth gradients, validating accuracy within the detected area, culminating in the generation of depth maps. Furthermore, the identification of shifts in the scene allows the examination of linear scene perception, facilitated by the Hough Transform Line Detection Algorithm (HTLDA). The mathematical formulation of the depth gradient hypothesis is as follows:

$$\mathrm{Dep}\,(\mathcal{D}) = 128 + 255\left\{\sum_{\mathrm{pixel}(a,b)} W_{lr} + W_{td}\frac{b - \frac{height}{2}}{heigt}\right\}$$
$$/\mathrm{pixel}_{num}(\mathcal{D}) \tag{5}$$

Where $|w_{lr}| + |w_{td}| = 1$

$$\mathrm{Dep}_{(x_i)} = \frac{1}{P\,(a_i)}\sum_{a_j \in \Omega(a_i)} e^{-0.5\left[\frac{|a_j - a_i|}{\gamma_x^2} + \frac{|v(a_j) - v(a_i)|^2}{\gamma_\xi^2}\right]}\mathrm{Dep}\,(a_j) \tag{6}$$

$$P\,(a_i) = \sum_{a_j \in \Omega(a_i)} e^{-0.5\left[\gamma_a^2 |a_j - a_i| + \gamma_\xi^2 |v(a_j) - v(a_i)|^2\right]} \tag{7}$$

A higher depth value indicates that the pixel is closer to the observer. Here, the intensity values are scaled from 0 (black) to 255 (white), with intermediate shades of gray are representing different signal strengths in the image. The following equation illustrates that the center of gravity is represented by the depth value within a block group, where the pixels belong to the same group share the same depth value. The $|w_{lr}|$ and $|w_{td}|$ values are controlled to control the depth gradient horizontally as well as vertically. Once the depth map is generated by grouping regions into blocks, it may exhibit blocky artifacts.

To address this issue, the cross-bilateral filter is employed to smoothly refine the depth map while preserving object boundaries. Afterward, the depth map is further improved through pixel value adjustments and hole filling using the QNLM filter, resulting in the creation of 3D representations. The preprocessing of the depth image primarily involves applying a smoothing filter. However, this filter, combined with the transition of sharp horizontal features, can create significant holes. To mitigate this problem, the QNLM filter is utilized to reduce the occurrence of large holes.

We then execute 3D image warping, and the 3D image warping scheme repositions pixels according to their depth values. The formulation of 3D image warping is as follows:

$$e_l = e_m + \left(\frac{d_{gx}}{2}\frac{f}{Z}\right) \tag{8}$$

$$e_r = e_m - \left(\frac{d_{gx}}{2}\frac{f}{Z}\right) \tag{9}$$

where, the horizontal positions are expressed as e_l, e_r and e_m with respect to the left, right and interposed positions, respectively. The value of depth in the current pixel is represented by Z. The distance of eye and the focal length is represented as d_{gx} and f, respectively. Moreover, we use QNLM with the aim of filtering holes to generate a 3D image.

C. Transformer Based Semantic Segmentation

Following the pre-processing, both pre-processed images are utilized for segmentation. Here, transformer based semantic segmentation is executed for acquiring pixel-level information effectively. For that, Mixed-transformer is used for getting features, including cortical thickness, colour, texture and boundary details from images. The densely connected feature aggregator model is then employed for collecting the features from multi-modalities and segment the ROI, which is detailly described below as follows:

1) Mixed Transformer: The core architecture of the network is based on an encoder-decoder framework, with the incorporation of skip connections during the decoding phase to retain essential low-level features. Notably, in an effort to optimize computational resources, we selectively apply Multi-Head Transformer Modules (MTMs) exclusively to the deeper layers with reduced spatial dimensions. For the upper layers, we maintain the use of conventional convolutional operations. This distinction is deliberate, as the initial layers contain higher-resolution features, and our focus is on capturing local relationships within them. Furthermore, the utilization of convolutional operations in the upper layers enables us to introduce structural priors into the model, a valuable feature particularly when working with relatively small medical image datasets. It is worth noting that a 2-stride convolutional/deconvolutional kernel is uniformly employed across all Transformer modules to facilitate channel expansion, compression, and down/up sampling. MT comprises of Local Global Gaussian-Self Attention (LGG-SA) and Dense Allied Feature Accumulation (DAFA). LGG-SA is constructed to model long-range and short-range dependencies along with diverse granularity. This technique is designed to substitute the encoder of traditional transformer for minimizing time
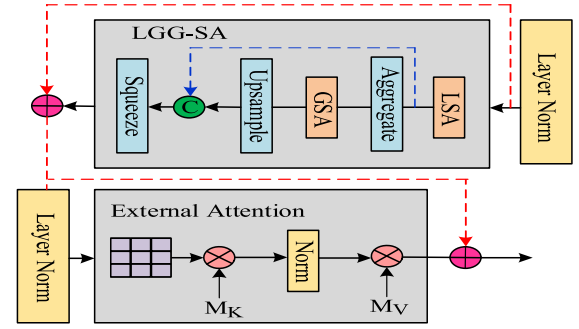

Fig. 1. LGG architecture.

complexity as well as providing better performance. LGG-SA modules are detailed below as follows:

a) Local-global self-attention: Initially, the SA tends to extract the interconnectedness among the entire entities of both MRI and PET image inputs individually. To identify the target, SA adapts three matrices that are key ($\mathfrak{K}$), query ($\mathcal{Q}$) and value ($\mathfrak{V}$). These three matrices are defined as input linear transforms $\mathcal{X}$. Besides, we introduce LGSA, as shown in fig.1, for enhancing the significance of correlations. Here, the local SA evaluates self-sympathies inside every window. Next, the tokens inside every window are accumulated as global tokens. For the accumulation operations, we apply max pooling, stride convolution, and other techniques of that Lightweight Dynamic Convolution (LDC) execute effectively. Following the overall features of down-sampled, we execute Global SA with minimal expense. For $\mathcal{X} \in \mathbb{R}^{\mathcal{H} \times \mathcal{W} \times \mathcal{C}}$, if we fix window size to $\mathcal{P}$, then the entire process is mathematically expressed as:

$$z_{loc} = LSA(\mathcal{X}) \tag{10}$$

$$z_{glo} = GSA(LDC(z_{loc})) \tag{11}$$

$$z = Concat(z_{loc}, Up_{sample}(z_{glo})) \tag{12}$$

where z indicates the output, LSA is local self-attention, and GSA is equivalent global functions.

b) Gaussian-weighted axial attention: Contrasting Local Self-Attention (LSA) utilizing default SA, we designed Gaussian Weighted Axial Attention (GWAA) which improves every query perception of adjacent via determinable Gaussian matrix, and meanwhile minimal time complexity as per axial attention. Let $\mathcal{Q} \in \mathbb{R}^{\frac{\mathcal{H}}{\mathcal{P}} \times \frac{\mathcal{W}}{\mathcal{P}}}$ signifies the queries acquired from accumulation step, for query $q_{i,j}$ in $\mathcal{Q}$, we describe $\mathfrak{D}_{i,j}$ as Euclidean distance among $q_{i,j}$ and it is equivalent to $\mathfrak{K}_{i,j}$ and $\mathfrak{V}_{i,j}$, where $\mathfrak{K}_{i,j}$ and $\mathfrak{V}_{i,j}$ are represented as matrices computed from tokens on ith row and jth column after accumulation. Assume the similarity among q and $\mathfrak{K}$ existence $\mathfrak{P}(q, \mathfrak{K})$ and then weight of Gaussian being $e^{-\frac{\mathfrak{D}_{i,j}^2}{2\varphi^2}}$, the output of final in position (i, j) can be depicted as:

$$z_{i,j} = e^{-\frac{\mathfrak{D}_{i,j}^2}{2\varphi^2}} softmax(\mathfrak{P}(q_{i,j}, \mathfrak{K}_{i,j})) \mathfrak{V}_{i,j} \tag{13}$$

Meanwhile, we need the variance φ to be determinable and then aforementioned equation can be also denoted as:

$$z_{i,j} = softmax\left(-\frac{1}{2\varphi^2}\mathfrak{D}_{i,j}^2 + \mathfrak{P}(q_{i,j}, \mathfrak{K}_{i,j})\right) \mathfrak{V}_{i,j} \tag{14}$$
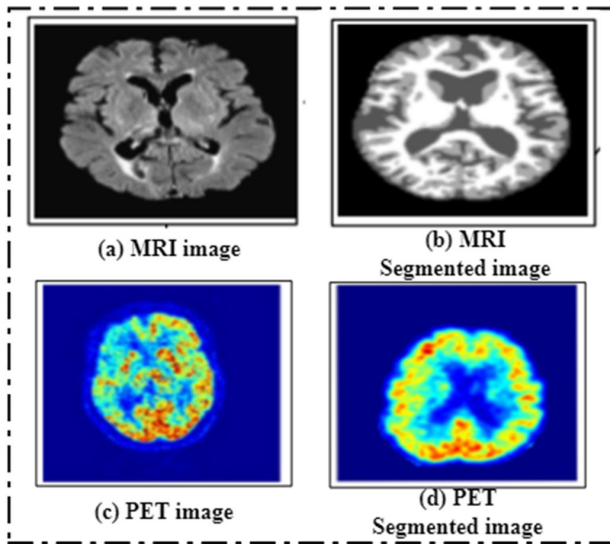
Fig. 2. Representation of multi-modalities segmented image.

Here, we generally utilize ω to denote the factor of coefficient before $\mathfrak{D}^2_{i.j}$, $\omega\mathfrak{D}^2_{i.j}$ further play as bias of correlative position, which can underline the position information of MT. It enhances the model performance for obviously affording correlative relations, and it is the usual embedding of utter positional. At last, the EA is introduced for solving the issues which cannot exploit correlations among diverse images.

2) Semantic Segmentation Using Furthered U-Net: After extracting features with the Mixed Transformer, we employ the Furthered U-Net (FU-Net) Algorithm to segment white matter, grey matter, and cerebrospinal fluid. This segmentation effectively breaks down the infected areas, as depicted in Figure 2. In contrast to the traditional U-Net approach, our work incorporates Batch Normalization (BN) to enhance training stability and mitigate gradient vanishing issues. This optimization enhances the segmentation performance, further aiding model convergence. The mathematical evaluation of the rational formula proceeds as follows:

$$\Lambda = \frac{\psi}{\sqrt{Var\,[s] + \varepsilon}} \cdot x + \left(\xi - \frac{\psi.\zeta\,[x]}{\sqrt{Var\,[s] + \varepsilon}} \right) \quad (15)$$

In the equation above, 'x' represents the input features, 'Λ' denotes the standardized feature with values close to zero. The parameters 'ψ' and '$\mathfrak{m}$' are training parameters that are updated during the process. Subsequently, the loss function (cross-entropy) is used in the training phase. The Adam optimizer is utilized for the optimization tasks, The updating of parameters within the algorithm can be expressed as follows:

$$s_\mathfrak{m} = s_0 - \frac{\eta\mathfrak{w}}{\sqrt{\mathfrak{v}_\mathfrak{m}}} \quad (16)$$

$$\mathfrak{m}_\mathfrak{m} = b_1 \times \mathfrak{m}_0 + (1 - b_1)\,f'\,(q_0) \quad (17)$$

$$\mathfrak{v}_\mathfrak{m} = b_2 \times \mathfrak{v}_0 + (1 - b_2)\left[f'\,(q_0)\right]^2 \quad (18)$$

where b_1, b_2 denoted as loss rate, η is the learning rate, the parameters $\mathfrak{v}_\mathfrak{m}$ and $\mathfrak{v}_0$ are the old and new parameters. $\mathfrak{m}$ represents the morphology differs. Moreover, the algorithm can compute the learning rates range in repetition to assure the parameter stability and efficiency of high computational.

D. Multi-Modality-Based Alzheimer's Diagnosis

Once the segmentation is completed, the segmented image is fed into proposed Dual-3DM^3−AD model. In that, the appropriate features are extracted in multi-scale, and dimensionality is minimized by using the multi-head attention mechanism, which is elaborated as follows:

1) Multi-Scale Feature Extraction: We utilize two parallel ResNet-51 blocks as encoders for extracting the feature maps from both MRI and PET segmented 3D images separately. For the utilization of encoder input, we direct the MRI and PET images in three channels by repeating their information in single-channel. The encoder is convolution integration, Rectified Linear Unit (ReLU), batch normalization and max pooling (CRBM) followed through an alternate integration of ResNet block (RB) and Evolution Down sampling Block (EDB). We extract the feature F_{MRI} such as textural, statistical, structural, edge, blobs, color and contour are extracted using the multi-scale feature extraction model. Additionally, the PET images are extracted F_{PET} after every ResNet block. From encoders, we extract F_{MRI} and F_{PET} features at 1/4, 1/8, 1/16 and 1/32 scales in size of original image. After that, the multi-scale features are acquired in elementwise addition.

2) Densely Allied Feature Accumulation: In order to aggregate the features from MRI and PET, we adapted DAFA module for feature representation. Specifically, we introduce Collective Spatial Attention (CSA) and Collective Channel Attention (C2A) for improving the spatial-wise and channel-wise representation of semantic features. Here, the main intention of utilizing CSA and C2A is to perform multi-scale features in diverse scales. To be more specific, both CSA and C2A comprise of convolutional filters, query, value and key functions which provide appropriate weights for individual features to accumulate precisely. Additionally, the features from multi- modalities are combined by utilizing downsample association and upsample association of large-filed for enhancing the multi-scale illustration. The DAFA accumulates features of MRI and PET as F_{MRI} and F_{PET}.

a) Upsample connections: The upsampling connections $\cup^j_i$ (ɕ) aim to pass information from one layer to another, while maintaining or even enhancing spatial resolution. In which, both MRI and PET pass features information for enhancing the spatial resolution by integrating upsampling operations.

b) Downsample connection: The downsample connection tends to interlink with both MRI and PET features for fusion, and it can be expressed as:

$$\mathfrak{D}^j_i\,(ɕ) = \mathfrak{f}_{\shortmid}\;\left(\mathfrak{f}_\mu\,(ɕ) + \mathfrak{f}_\tau\,(\mathfrak{f}_\theta\,(ɕ))\right) \quad (19)$$

where ɕ denotes the input vector, $\mathfrak{f}_{\shortmid}$ is the ReLU activation function. The parameter $\mathfrak{f}_\mu$ and $\mathfrak{f}_\tau$ are 3×3 convolution layer along with 2 stride and $\mathfrak{f}_\theta$ is a 3×3 convolution layer along with 1 stride. Here, every convolution layer includes batch normalization technique. i and j are represented as channels of input and output, respectively.

c) Collective spatial attention: As per the mechanism of linear attention, we used the CSA to design the long-range addictions of spatial dimension, and it can mathematically be

defined as:

$$CSA\left(\mathfrak{h}\right) = \frac{\sum_n V\left(\mathfrak{h}\right)_{c,n} + \left(\frac{Q\left(\mathfrak{h}\right)}{\|Q\left(\mathfrak{h}\right)\|_2}\right)\left(\frac{K\left(\mathfrak{h}\right)}{\|Q\left(\mathfrak{h}\right)\|_2}\right)^T V\left(\mathfrak{h}\right)}{\mathcal{N} + \left(\frac{Q\left(\mathfrak{h}\right)}{\|Q\left(\mathfrak{h}\right)\|_2}\right)\sum_n \left(\frac{Q\left(\mathfrak{h}\right)}{\|Q\left(\mathfrak{h}\right)\|_2}\right)^T_{c,n}} \quad (20)$$

where, $Q\left(\mathfrak{h}\right)$, $K\left(\mathfrak{h}\right)$ and $V\left(\mathfrak{h}\right)$ indicate the convolutional functions to compute the query matrix $Q \in \mathbb{R}^{\mathcal{N} \times \mathfrak{D}_Y}$, key matrix $K \in \mathbb{R}^{\mathcal{N} \times \mathfrak{D}_Y}$ and value matrix $V \in \mathbb{R}^{\mathcal{N} \times \mathfrak{D}_Y}$, $\mathcal{N}$ denotes the number of pixels of input feature maps. n and c are the dimension of flattened spatial and channel dimension.

d) Collective channel attention: Likewise, CCA is modelled for extracting the long- range addictions between channel dimension that can defined as:

$$CCA\left(\mathfrak{h}\right) = \frac{\sum_c \mathfrak{R}\left(\mathfrak{h}\right)_{c,n} + \left(\mathfrak{R}\left(\mathfrak{h}\right)_{c,n}\left(\frac{K\left(\mathfrak{h}\right)}{\|Q\left(\mathfrak{h}\right)\|_2}\right)^T\right)\frac{Q\left(\mathfrak{h}\right)}{\|Q\left(\mathfrak{h}\right)\|_2}}{\mathcal{N} + \left(\frac{\mathfrak{R}\left(\mathfrak{h}\right)}{\|\mathfrak{R}\left(\mathfrak{h}\right)\|_2}\right)^T \sum_c \left(\frac{\mathfrak{R}\left(\mathfrak{h}\right)}{\|\mathfrak{R}\left(\mathfrak{h}\right)\|_2}\right)^T_{c,n}} \quad (21)$$

where $\mathfrak{R}\left(\mathfrak{h}\right)$ denotes the reshape function for flattening the spatial dimension. In summary, the primary difference lies in what actually these attention mechanisms focus on: spatial attention deals with the spatial positions within the data, while channel attention deals with the feature channels or dimensions. They can be used in combination to enhance the representation and performance of the proposed model, depending on the nature of the classification task.

e) Feature accumulation: At last, the features obtained from both MRI and PET features $\mathfrak{AF}_1$ and $\mathfrak{AF}_2$ are fused, which can be generated by the following mathematical equations:

$$\text{\textbardbl} = F_{MRI} + F_{PET} + U \quad (22)$$

Here, \text{\textbardbl} is the feature accumulation factor, F is the feature obtained from both MRI and PET indicated as F_{MRI} and F_{PET}. U is denoted as upsample function of bilinear interpolation and spatial enhancement along with 2 scale factors.

3) Multi-Head Attention Mechanism: Multi-head attention mechanism executes several linear transformations at feature matrix of input and determines the attention illustrations of image across diverse linear transformation; therefore, we acquire huge inclusive Alzheimer's information. This mechanism is fundamentally integration of several self-attention scheme, key ($\mathfrak{K}$), query ($\mathcal{Q}$) and value ($\mathfrak{V}$). The primary intention of the scheme is a Scaled Dot product Attention (SDA). The function of SDA is expressed as:

$$SDA\left(\mathcal{Q}, \mathfrak{K}, \mathfrak{V}\right) = softmax\left(\frac{\mathcal{Q}\mathfrak{K}}{\sqrt{d_\mathfrak{k}}}\right)^{\mathfrak{T}} \mathfrak{V} \quad (23)$$

The concept of multi-head attention is to utilize diverse parameters $\mathfrak{W}_i^{\mathcal{Q}}, \mathfrak{W}_i^{\mathfrak{K}}, \mathfrak{W}_i^{\mathfrak{V}}$ to execute linear transformations on $\mathcal{Q}, \mathfrak{K}, \mathfrak{V}$ matrices, and the result of input linear transformations as SDA. The estimation result is evaluated via $head_i$, which can be formulated as:

$$head_i = SDA\left(\mathcal{Q}\mathfrak{W}_i^{\mathcal{Q}}, \mathfrak{K}\mathfrak{W}_i^{\mathfrak{K}}, \mathfrak{V}\mathfrak{W}_i^{\mathfrak{V}}\right) \quad (24)$$

Next, we concatenate the evaluated results $head_1$ to $head_\hbar$ to create a matrix, and multiply it via parameter $\mathfrak{W}$ to conclude the final linear transformation:

$$Head = Multi_{head}\left(\mathcal{Q}, \mathfrak{K}, \mathfrak{V}\right) \quad (25)$$
$$= Concat\left(head_1, \ldots, head_\hbar\right)\mathfrak{W} \quad (26)$$

4) Output Layer of Alzheimer's Diagnosis: The average pooling is executed on $Head$ output matrix in multi-head attention layer to acquire the features vectors F_{MP}^{avg}. We pass the input F_{MP}^{avg} via fully connected layer to final softmax classifier to obtain final Alzheimer diagnosis as:

$$\text{\textbarh} = softmax\left(w_m F_{MP}^{avg} + b_m\right) \quad (27)$$

Here, w_m is depicted as weight matrix and b_m is bias. We utilize back propagation technique to optimize our proposed model, and the cross entropy is expressed as:

$$loss = \sum_{i=1}^{\mathbb{D}}\sum_{j=1}^{C} \hat{\text{\textbarh}}_i^j \ In \ \text{\textbarh}_i^j + \lambda\|\theta\|^2 \quad (28)$$

where, $\mathbb{D}$ is denoted as training data size, C is the number of data classes, $\text{\textbarh}$ is represented as predicted class, $\hat{\text{\textbarh}}$ is the actual class and $\lambda\|\theta\|^2$ is the default term cross-entropy.

IV. EXPERIMENTAL RESULTS

In this section, we demonstrate the effectiveness of the proposed Dual-3DM^3−AD model in terms of Alzheimer detection. This section is divided into three sub-sections including simulation setup, comparison analysis and research summary:

A. Simulation Setup

The entire model execution and evaluation are implemented by utilizing MATLAB 2020A. Moreover, we distributed the dataset as 90:10 ratio, and the 10-fold-cross validation is adopted. To diagnosis Alzheimer's using MRI and PET scans, the Dual-3DM^3−AD model is utilized as a classifier. We set 32 mini-batch size, 100 epochs to fair analysis in 0.00008 learning rate. Tab. I shows the hardware parameters.

B. Experiments

The proposed Dual-3DM^3−AD model performance is compared with the existing approaches with respect to sensitivity, accuracy, confusion matrix, specificity, and ROC curve. We performed the classification by Cognitive Normal (CN) vs AD, AD vs Mild Cognitive Impairment (MCI) and CN vs MCI. Accuracy affords us the true resultants proportion, which can be true negative or true positive. Sensitivity appearances the entire performance of proposed model. Specificity shows how effectively the model is recognizing CN condition. ROC curves and confusion matrices are visually characteristics perceptions regarding predictive analysis.
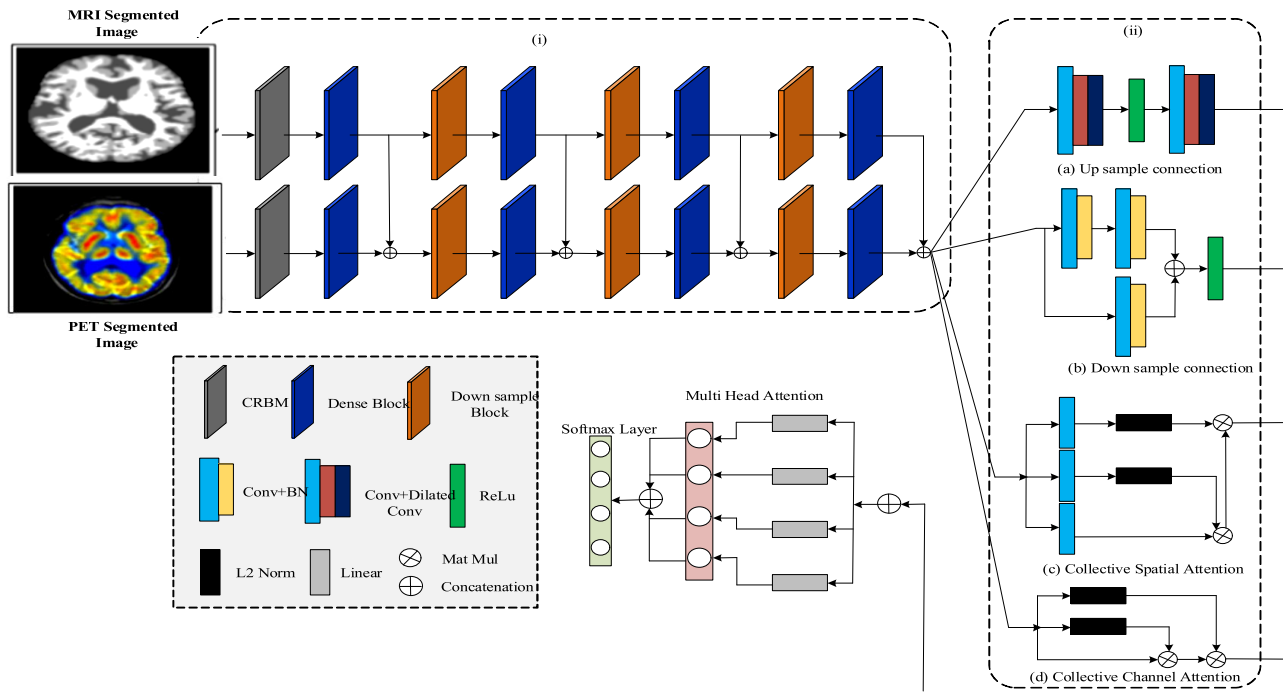
Fig. 3. Overall architecture of proposed dual-3DM^3−AD model.

C. Comparative Analysis

We elucidated the comparison between the proposed model and existing works, where we have contemplated with two existing works such as - The primary intention of this paper is to perform segmentation and Alzheimer diagnosis effectively.

1) Comparison With Diverse Modalities: For the comparative analysis between MRI, PET fused information, the Dual-3DM^3−AD model is utilized for each of those modalities. Fig 4(a)-(c) represents the confusion matrices and ROC curves of CN vs AD classification acquired from diverse modalities. In fig 5, class-1 illustrates CN, and class-2 illustrates AD. As defined, classification by the consideration of fused data provides ROC curve about to top-left recommending the fused data usefulness.

Table II shows the comparative analysis in terms of performance metrics, and it outlines that the fusion-based classification is more accurate than PET and MRI. Both MRI and PET data separately obtain minimal performance, which is justified through inefficiency of single modality to meet metabolic and structural modifications instantaneously. Whereas, the multi-modality fused data concentrates on these brain information. In pre-processing, the noise removal and skull stripping are performed, which removes the noise and unwanted tissues; therefore, contemplating the amount of computation cost. Moreover, the multi-head-based attention mechanism minimizes the complexities. Henceforth, the Dual-3DM^3−AD model testing utilizes 2 minutes on machine with one GPU, which articulating the algorithm's space complexity and optimum time.

2) Comparison With Diverse State-of-Art Approaches: The proposed Dual-3DM^3−AD model is compared with several state-of-the-art approaches to demonstrate the proposed model efficacy for AD classification. EPEE [22], Novel-CNN [23], DEMNET [24], EMLM [25], RELS-TSVM [26] and THS-GAN are the approaches utilized for the comparison purpose.
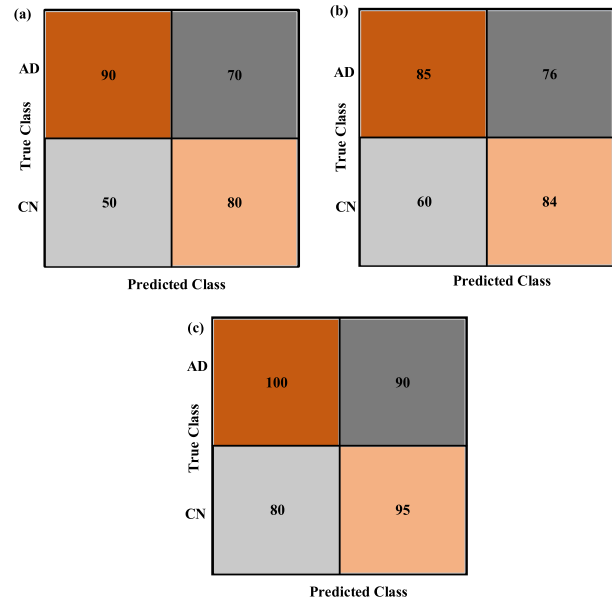

Fig. 4. Confusion matrix for proposed model (a) MRI, (b) PET and (c) Fused Data.

The comparison of Dual-3DM^3−AD model performance metrics with state-of-the-art approaches is unveiled in Table III. The baseline approaches are defined as follows:

[i] EPEE: A deep learning based approach using EPEE is proposed for Alzheimer diagnosis using MRI images, which performs better.

[ii] Novel-CNN: Early diagnosis of Alzheimer's classification is proposed by designing neural network-based novel-CNN using T2 weighted MRI scans.

[iii] DEMNET: DL model is proposed for diagnosing Dementia and Alzheimer's classification for handling unbalancing dataset.

TABLE II
PERFORMANCE ANALYSIS OF PROPOSED MODEL FOR ALZHEIMER DIAGNOSIS WITH DIVERSE MODALITIES

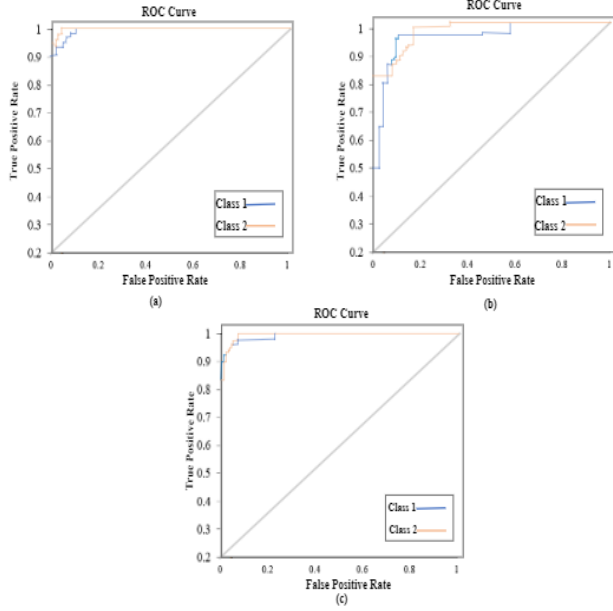| Modality	CN vs AD				CN vs MCI				AD vs MCI			
	Acc	Sen	Spe	f-m	Acc	Sen	Spe	f-m	Acc	Sen	Spe	f-m
MRI	92	91	90.7	91	91.7	91.4	92	91.8	92.3	91.8	91.2	91.5
PET	90	91	90.3	90.6	91.3	91.2	92	92	91	93	92	92.8
Fusion	**98**	**97.5**	**98**	**97.2**	**97.9**	**98.2**	**98**	**98**	**98.2**	**97.8**	**97**	**98**

Fig. 5. ROC curve for proposed model (a) MRI, (b) PET and (c) Fused data.

[iv] EMLM: An early detection for Alzheimer's based on hippocampal and cortical local filed is proposed by adapting EMLM model.

[v] RELS-TSVM: A DK based Alzheimer's detection is implemented by utilizing multi-modality data for obtaining accurate result.

[vi] THS-GAN: An MRI based classification model THS-GAN is proposed for the identification of multi-class Alzheimer's disease.

The proposed Dual-3DM^3−AD model exhibits superior performance with 98% of accuracy, 97.8% of sensitivity, 97.5% of specificity and 98.2% of f-measure for CN vs AD diagnosis. Figs 6-9 represent the performance metrics analysis of the proposed vs existing works (accuracy, sensitivity, specificity, and F-measure). The proposed Dual-3DM^3−AD model displays better convergence characteristics and persuasive accuracy. It is apparent that the Dual-3DM^3−AD's ROC curve is nearer to top-left corner, depicting best performance than any other existing approaches. Hence, the multi-modal fusion based Dual-3DM^3−AD model proves to be a betterment automatic classification method.

3) Comparison With Diverse Machine Learning Approaches: We compare the proposed Dual-3DM^3−AD model with various machine learning approaches. BEMD [21], RF [27] and SVM [28] are utilized as classifiers for Alzheimer's diagnosis. The comparison of Dual-3DM^3−AD performance metrics with the existing classifiers in terms of accuracy, sensitivity, specificity and f-measure is illustrated in Table IV. The RF
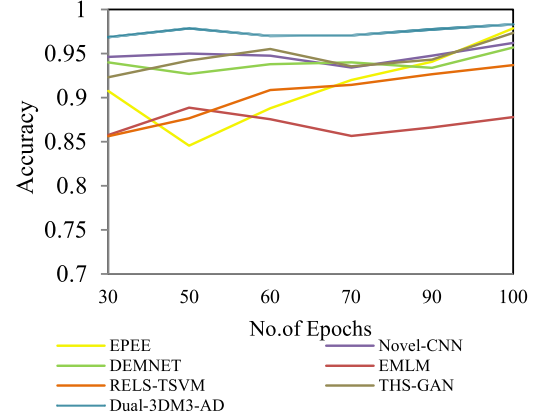

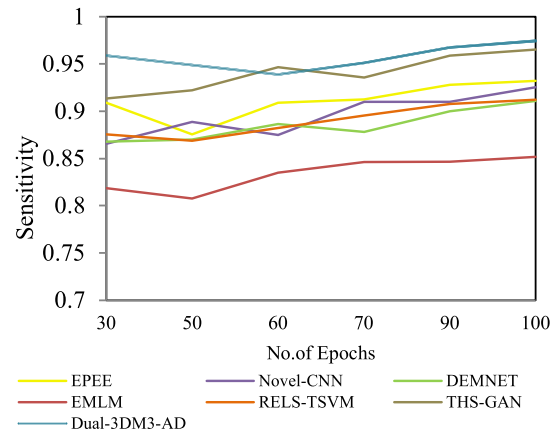
Fig. 6. Analysis of accuracy.

Fig. 7. Analysis of sensitivity.

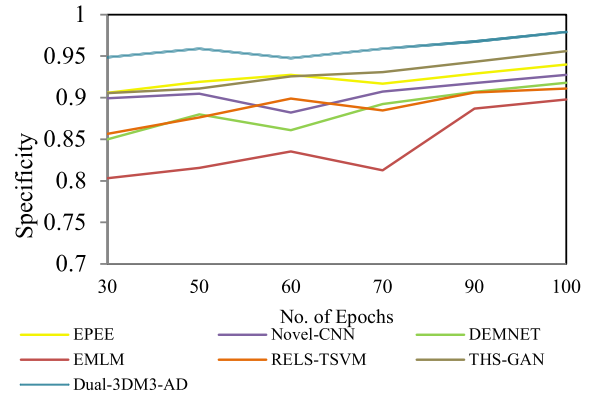Fig. 8. Analysis of specificity.

model performed better than SVM and NB as an Alzheimer's classification model on entire performance metrics. Also, the proposed work achieves maximum accuracy than other models. The reason for attaining lower accuracy by the machine learning approaches because they suffer from handling large dataset and being insufficient in terms of extracting appropriate features.

TABLE III
COMPARISON ANALYSIS OF PROPOSED MODEL FOR ALZHEIMER DIAGNOSIS WITH BASELINE APPROACHES

Modality	CN vs AD				CN vs MCI				AD vs MCI			
	Acc	Sen	Spe	f-m	Acc	Sen	Spe	f-m	Acc	Sen	Spe	f-m
EPEE	0.978	0.932	0.939	0.954	0.967	0.922	0.943	0.963	0.978	0.956	0.929	0.964
Novel-CNN	0.962	0.925	0.925	0.951	0.963	0.903	0.958	0.945	0.962	0.945	0.924	0.941
DEMNET	0.957	0.911	0.917	0.974	0.947	0.915	0.933	0.970	0.957	0.928	0.907	0.934
EMLM	0.878	0.851	0.898	0.88	0.847	0.862	0.843	0.883	0.878	0.867	0.888	0.868
RELS-TSVM	0.936	0.912	0.91	0.966	0.967	0.954	0.92	0.969	0.936	0.934	0.920	0.946
THS-GAN	0.973	0.965	0.955	0.976	0.970	0.966	0.938	0.971	0.973	0.945	0.965	0.966
Dual-3DM^3-AD	**0.983**	**0.974**	**0.978**	**0.98**	**0.98**	**0.979**	**0.987**	**0.981**	**0.986**	**0.98**	**0.978**	**0.985**

TABLE IV
COMPARISON ANALYSIS OF PROPOSED MODEL FOR ALZHEIMER DIAGNOSIS WITH ML APPROACHES

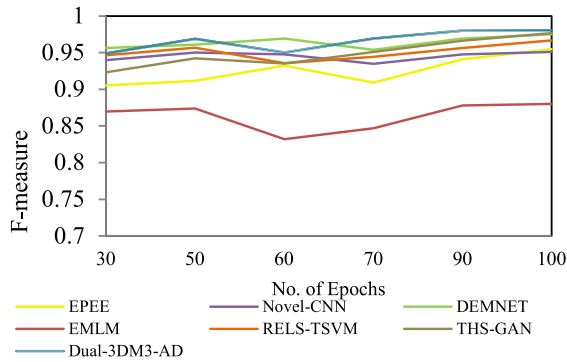| Modality	CN vs AD				CN vs MCI				AD vs MCI			
	Acc	Sen	Spe	f-m	Acc	Sen	Spe	f-m	Acc	Sen	Spe	f-m
BEMD	82	83.1	80.7	81.5	81.7	82.4	82	81.8	82.6	82.7	83.2	81.5
RF	87	87.5	88	87.2	87.9	88.2	86	86.3	87.2	87.8	86	87.3
SVM	80	81.5	80.15	80.6	81.8	82.5	82	82.4	81	83	85	82.8
Dual-3DM^3-AD	**0.983**	**0.974**	**0.978**	**0.98**	**0.98**	**0.979**	**0.987**	**0.981**	**0.986**	**0.98**	**0.978**	**0.985**

Fig. 9. Analysis of F-measure.

D. Evaluation of Proposed Dual-3DM^3−AD Model

To validate the Multi-level Capsule Network and Dual Vision Transformer based Attention Mechanism-Dual-Atten proposed framework, we accomplish ablation tests. For that, we have utilized SWLD-20K, Cresci-2017 and Cresci-2015 datasets to accord and understand the influence of every layer and component of our proposed Dual-3DM^3−AD model. The introduction of a multi-modal fusion-based approach is promising and indicates an effort to address the complex nature of AD diagnosis. Combining MRI and PET scans is a sound approach. The use of sophisticated techniques, such as QNLM, Morphology function, and BDM for image preprocessing is a positive aspect. These techniques can significantly enhance image quality, which is crucial for accurate diagnosis. The adoption of the Mixed-transformer with Furthered U-Net for semantic segmentation is a good choice, as it helps in identifying and isolating relevant regions within the images, which is critical for extracting meaningful features. The incorporation of a multi-scale feature extraction module DCFAM demonstrates a commitment to leveraging insights from both scans effectively. The use of a multi-head attention mechanism for feature dimensionality reduction is a suitable choice, as it can help managing the complexity of the data and concentrates upon the desired features. The application

of a softmax layer for multi-class Alzheimer's diagnosis is important for classifying the disease into different stages. This is a valuable contribution, as it provides clinicians with more detailed information.

To demonstrate its effectiveness, the proposed model has been compared to existing methods and benchmarked against them to establish its superiority. In conclusion, while the proposed work appears promising and comprehensive, its true effectiveness can only be determined through rigorous testing and validation on real-world data, and consideration of its practicality and ethical implications. In this experiment, the ADNI and radiopharmaceutical 18F-FDG dataset is distributed into training, validation and testing as 90%, 10%, and 15%, respectively. This is because we adapted large scale of dataset, where 10 % of data is adequate for estimation of test set or validation set. Besides, the utilization of large data in training can enhance the performance of deep neural network to train sufficiently.

We also tend to compare the evaluation of the proposed multi modal approach with the single modal approach in terms of accuracy, specificity, sensitivity, and F-measure. For a multi modal approach, the results we achieved are clearly depicted in fig (6)-(9). Whereas for the single modal scenario MRI and PET, the results acquired by the MRI is higher than the PET. Also, Tab. V unveils the utilized symbols.

V. DISCUSSION

The effectiveness of the proposed Dual-3DM3-AD model for Alzheimer's diagnosis was rigorously evaluated, and the results demonstrated its potential for accurate and early detection of the disease using both MRI and PET image scans. In the initial stages of the study, the extensive preprocessing techniques, including noise reduction, skull stripping, and 3D image conversion, were applied using state-of-the-art algorithms such as the QNLM, Morphology function, and BDM. These steps significantly enhanced the quality of the input images, ensuring that the subsequent analysis was performed on clean and accurate data. The model architecture

TABLE V
SYMBOL DEFINITION

Symbol	Definition
$\widehat{\mathfrak{x}}$	Denoised Image
$\mathrm{Dep}(\mathcal{D})$	Depth gradient hypothesis
e_l, e_r, e_m	Horizontal Coordinates
$\mathcal{Q}, \mathcal{V}, \mathcal{K}$	Query, Value and Key
$\mathcal{X}$	Input feature map
Λ	Standardized feature
$\mathfrak{m}$	Training parameters
$\mathfrak{y}$	Learning rate
$\mathfrak{D}_i^j$	Downsample connection
$\mathfrak{h}$	Input vector
F_{MRI}	Feature of MRI
F_{PET}	Feature of PET
F	Feature
ϕ	Feature Accumulation
$\mathbb{D}$	Training image size
U	Feature Upsampling
C	Number of Classes

itself was designed for optimal performance. The integration of a Mixed-transformer with Furthered U-Net for semantic segmentation effectively minimized complexity, allowing for the extraction of meaningful features from both MRI and PET scans. The multi-scale feature extraction module played a crucial role in capturing relevant information from the segmented images. The model further benefited from the DCFAM, which efficiently aggregated the extracted features, enabling the utilization of both modalities. The multi-head attention mechanism was employed for feature dimensionality reduction, enhancing the model's ability to distinguish key patterns associated with Alzheimer's disease. Our model overcome both underfitting and overfitting issues as:

Complexity Reduction With Mixed-Transformer and Furthered U-Net: The use of a Mixed-transformer and Furthered U-Net suggests an effort to create a model with increased representational capacity. This can help capture complex patterns in the data. By combining different transformer architectures and enhancing the U-Net, the model may be better equipped to handle intricate relationships within the images.

Dual-3DM3-AD Model: The Dual-3DM3-AD model is described as having a multi-scale feature extraction module. Multi-scale features can capture information at different levels of granularity, which may assist in handling both finer details and more global context in the images.

Feature Aggregation With Densely Connected Feature Aggregator Module (DCFAM): The DCFAM module is mentioned as a feature aggregator. Aggregating features from different scales or sources can help in capturing a comprehensive representation of the input data. Densely connected architectures often encourage feature reuse, which can be beneficial for learning informative representations.

Multi-Head Attention Mechanism for Dimensionality Reduction: The use of a multi-head attention mechanism is stated for feature dimensionality reduction. Attention mechanisms allow the model to focus on relevant parts of the input. In this context, reducing dimensionality may aid in preventing overfitting by promoting more efficient use of information.

Softmax Layer for Multi-Class Alzheimer's Diagnosis: The application of a softmax layer for multi-class Alzheimer's diagnosis indicates the usage of a common activation function for classification tasks. This is crucial for preventing underfitting or overfitting in the final classification layer.

VI. CHALLENGES AND LIMITATIONS OF PROPOSED WORK

The proposed Dual-3DM^3−AD model for Alzheimer's diagnosis presents several limitations for its practical implementation in real clinical environments. Firstly, the model's reliance on high-quality and diverse MRI and PET datasets may pose challenges in real-world settings, where data availability can be limited. Additionally, the computational demands of the model, including preprocessing and complex neural network architectures, may strain the resources of healthcare facilities. The lack of model interpretability hinders the understanding of how diagnoses are arrived at, potentially impacting trust among healthcare professionals. Variations in imaging standards and equipment in clinical settings must be addressed for the model to perform consistently.

VII. CONCLUSION AND FUTURE WORK

Lack of training/testing data consideration and ineffective segmentation are one of the major reasons for low Alzheimer diagnosis accuracy, which is still a crucial concern. To alleviate these issues, we presented a promising avenue for a more comprehensive understanding of AD staging. This paper introduced an innovative approach to address this challenge. We proposed the Dual-3DM^3−AD model, designed for accurate and early Alzheimer's diagnosis, by leveraging both MRI and PET image scans. Our methodology involved a series of preprocessing steps, including noise reduction, skull stripping, and 3D image conversion, performed using the QNLM, Morphology function, and BDM, respectively, to enhance the image quality.

Subsequently, we employed a Mixed-transformer with Furthered U-Net architecture for semantic segmentation, effectively reducing complexity. The Dual-3DM^3−AD model incorporated a multi-scale feature extraction module to extract pertinent features from the segmented images. These extracted features were then aggregated using the densely connected feature aggregator module to make the most of both information sources. Furthermore, we employ a multi-head attention mechanism to reduce feature dimensionality, followed by the application of a softmax layer for multi-class Alzheimer's diagnosis. Our proposed Dual-3DM^3−AD model was implemented in MATLAB 2020A and rigorously compared with several baseline approaches by using a range of performance metrics, including accuracy, sensitivity, specificity, f-measure, and ROC curve analysis. Remarkably, our work surpassed existing models in multi-class Alzheimer's diagnosis, underscoring its potential as a valuable tool in the early detection of this debilitating disease. In terms of future work, we have planned to propose an Explainable Artificial Intelligence (EAI) with computation reduction technique for better understanding of classification result with the aim of further reducing computational complexity and including feedback system.

REFERENCES

[1] Z. Wang, J. Song, Y. Wang, and W. Liu, "Alzheimer's disease classification detection based on brain electrical signal graph structure," in *Proc. 3rd Int. Conf. Frontiers Electron., Inf. Comput. Technol. (ICFEICT)*, May 2023, pp. 294–300.

[2] K. N. McFarland and P. Chakrabarty, "Microglia in Alzheimer's disease: A key player in the transition between homeostasis and pathogenesis," *Neurotherapeutics*, vol. 19, no. 1, pp. 186–208, Jan. 2022.

[3] R. Lathe and D. S. Clair, "Programmed ageing: Decline of stem cell renewal, immunosenescence, and Alzheimer's disease," *Biol. Rev.*, vol. 98, no. 4, pp. 1424–1458, Aug. 2023.

[4] P. Gruener, "Alzheimer's disease in American fiction," in *Beyond the Great Forgetting*, J. B. Metzler, Ed. Berlin, Germany: Springer, 2022, doi: 10.1007/978-3-662-66029-4_5.

[5] G. Plascencia-Villa and G. Perry, "Status and future directions of clinical trials in Alzheimer's disease," *Int. Rev. Neurobiol.*, vol. 154, pp. 3–50, Jul. 2020.

[6] Y. Zhang, H. Chen, R. Li, K. Sterling, and W. Song, "Amyloid β-based therapy for Alzheimer's disease: Challenges, successes and future," *Signal Transduction Targeted Therapy*, vol. 8, no. 1, p. 248, Jun. 2023.

[7] M. Mather, "Noradrenaline in the aging brain: Promoting cognitive reserve or accelerating Alzheimer's disease?" *Seminars Cell Develop. Biol.*, vol. 116, pp. 108–124, Aug. 2021.

[8] M. F. Ahmad, S. Akbar, S. A. E. Hassan, A. Rehman, and N. Ayesha, "Deep learning approach to diagnose Alzheimer's disease through magnetic resonance images," in *Proc. Int. Conf. Innov. Comput. (ICIC)*, Nov. 2021, pp. 1–6.

[9] M. B. T. Noor, N. Z. Zenia, M. S. Kaiser, S. A. Mamun, and M. Mahmud, "Application of deep learning in detecting neurological disorders from magnetic resonance images: A survey on the detection of Alzheimer's disease, Parkinson's disease and schizophrenia," *Brain Informat.*, vol. 7, no. 1, pp. 1–21, Dec. 2020.

[10] S. Iqbal, A. N. Qureshi, J. Li, and T. Mahmood, "On the analyses of medical images using traditional machine learning techniques and convolutional neural networks," *Arch. Comput. Methods Eng.*, vol. 30, no. 5, pp. 3173–3233, Jun. 2023.

[11] E. Guedj et al., "EANM procedure guidelines for brain PET imaging using [18F]FDG, version 3," *Eur. J. Nucl. Med. Mol. Imag.*, vol. 49, no. 2, pp. 632–651, Jan. 2022.

[12] B. R. Price, L. A. Johnson, and C. M. Norris, "Reactive astrocytes: The Nexus of pathological and clinical hallmarks of Alzheimer's disease," *Ageing Res. Rev.*, vol. 68, Jul. 2021, Art. no. 101335.

[13] J. Hong et al., "Image-level trajectory inference of tau pathology using variational autoencoder for flortaucipir PET," *Eur. J. Nucl. Med. Mol. Imag.*, vol. 49, no. 9, pp. 3061–3072, Jul. 2022.

[14] M. Solnik et al., "Imaging of uveal melanoma—Current standard and methods in development," *Cancers*, vol. 14, no. 13, p. 3147, Jun. 2022.

[15] H. Pleş et al., "Migraine: Advances in the pathogenesis and treatment," *Neurol. Int.*, vol. 15, no. 3, pp. 1052–1105, Aug. 2023.

[16] V. B. Gupta et al., "Retinal changes in Alzheimer's disease—Integrated prospects of imaging, functional and molecular advances," *Prog. Retinal Eye Res.*, vol. 82, May 2021, Art. no. 100899.

[17] S. Hashimoto et al., "Neuronal glutathione loss leads to neurodegeneration involving gasdermin activation," *Sci. Rep.*, vol. 13, no. 1, pp. 1–9, Jan. 2023.

[18] B. J. Matchett, L. T. Grinberg, P. Theofilas, and M. E. Murray, "The mechanistic link between selective vulnerability of the locus coeruleus and neurodegeneration in Alzheimer's disease," *Acta Neuropathologica*, vol. 141, no. 5, pp. 631–650, May 2021.

[19] Y. Blinkouskaya and J. Weickenmeier, "Brain shape changes associated with cerebral atrophy in healthy aging and Alzheimer's disease," *Frontiers Mech. Eng.*, vol. 7, pp. 1–17, Jul. 2021.

[20] V. Sathiyamoorthi, A. K. Ilavarasi, K. Murugeswari, S. T. Ahmed, B. A. Devi, and M. Kalipindi, "A deep convolutional neural network based computer aided diagnosis system for the prediction of Alzheimer's disease in MRI images," *Measurement*, vol. 171, Feb. 2021, Art. no. 108838.

[21] J. E. W. Koh et al., "Automated detection of Alzheimer's disease using bi-directional empirical model decomposition," *Pattern Recognit. Lett.*, vol. 135, pp. 106–113, Jul. 2020.

[22] H. S. Zaina, S. B. Belhaouari, T. Stanko, and V. Gorovoy, "An exemplar pyramid feature extraction based Alzheimer disease classification method," *IEEE Access*, vol. 10, pp. 66511–66521, 2022.

[23] S. Basheera and M. S. S. Ram, "A novel CNN based Alzheimer's disease classification using hybrid enhanced ICA segmented gray matter of MRI," *Computerized Med. Imag. Graph.*, vol. 81, Apr. 2020, Art. no. 101713.

[24] S. Murugan et al., "DEMNET: A deep learning model for early diagnosis of Alzheimer diseases and dementia from MR images," *IEEE Access*, vol. 9, pp. 90319–90329, 2021.

[25] M. Fabietti et al., "Early detection of Alzheimer's disease from cortical and hippocampal local field potentials using an ensembled machine learning model," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 2839–2848, 2023.

[26] S. Dwivedi, T. Goel, M. Tanveer, R. Murugan, and R. Sharma, "Multimodal fusion-based deep learning network for effective diagnosis of Alzheimer's disease," *IEEE MultimediaMag.*, vol. 29, no. 2, pp. 45–55, Apr. 2022.

[27] W. Yu, B. Lei, M. K. Ng, A. C. Cheung, Y. Shen, and S. Wang, "Tensorizing GAN with high-order pooling for Alzheimer's disease assessment," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 9, pp. 4945–4959, Sep. 2022.

[28] M. Song, H. Jung, S. Lee, D. Kim, and M. Ahn, "Diagnostic classification and biomarker identification of Alzheimer's disease with random forest algorithm," *Brain Sci.*, vol. 11, no. 4, p. 453, Apr. 2021.

[29] E. E. Bron et al., "Cross-cohort generalizability of deep and conventional machine learning for MRI-based diagnosis and prediction of Alzheimer's disease," *NeuroImage: Clin.*, vol. 31, 2021, Art. no. 102712.

[30] K. Etminani et al., "A 3D deep learning model to predict the diagnosis of dementia with lewy bodies, Alzheimer's disease, and mild cognitive impairment using brain 18F-FDG PET," *Eur. J. Nucl. Med. Mol. Imag.*, vol. 49, no. 2, pp. 563–584, Jan. 2022.

[31] C. S. Martinez, M. B. Cuadra, and J. Jorge, "BigBrain-MR: A new digital phantom with anatomically-realistic magnetic resonance properties at 100-μm resolution for magnetic resonance methods development," *NeuroImage*, vol. 273, Jun. 2023, Art. no. 120074.

[32] H. Kalantar-Hormozi et al., "A cross-sectional and longitudinal study of human brain development: The integration of cortical thickness, surface area, gyrification index, and cortical curvature into a unified analytical framework," *NeuroImage*, vol. 268, Mar. 2023, Art. no. 119885.

[33] A. Irimia, "Cross-sectional volumes and trajectories of the human brain, gray matter, white matter and cerebrospinal fluid in 9473 typically aging adults," *Neuroinformatics*, vol. 19, no. 2, pp. 347–366, Apr. 2021.

[34] N. Gharaibeh, A. A. Abu-Ein, O. M. Al-hazaimeh, K. M. O. Nahar, W. A. Abu-Ain, and M. M. Al-Nawashi, "Swin transformer-based segmentation and multi-scale feature pyramid fusion module for Alzheimer's disease with machine learning," *Int. J. Online Biomed. Eng. (iJOE)*, vol. 19, no. 4, pp. 22–50, Apr. 2023.

[35] M. Liu et al., "A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease," *NeuroImage*, vol. 208, Mar. 2020, Art. no. 116459.

[36] C. L. Saratxaga et al., "MRI deep learning-based solution for Alzheimer's disease prediction," *J. Personalized Med.*, vol. 11, no. 9, p. 902, 2021.

[37] R. A. Hazarika, A. K. Maji, S. N. Sur, B. S. Paul, and D. Kandar, "A survey on classification algorithms of brain images in Alzheimer's disease based on feature extraction techniques," *IEEE Access*, vol. 9, pp. 58503–58536, 2021.

[38] T. Wang and L. Cao, "Deep learning based diagnosis of Alzheimer's disease using structural magnetic resonance imaging: A survey," in *Proc. 3rd Int. Conf. Appl. Mach. Learn. (ICAML)*, Jul. 2021, pp. 408–412.

[39] J. Neelaveni and M. S. G. Devasana, "Alzheimer disease prediction using machine learning algorithms," in *Proc. 6th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS)*, Mar. 2020, pp. 101–104.

[40] A. Puente-Castro, E. Fernandez-Blanco, A. Pazos, and C. R. Munteanu, "Automatic assessment of Alzheimer's disease diagnosis based on deep learning techniques," *Comput. Biol. Med.*, vol. 120, May 2020, Art. no. 103764.