IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

# Alzheimer's Disease Diagnosis From Multimodal Medical Images And Patient Data

P. Horsley Solomon, Head, Department of Electronics Science, SRM Arts and Science College,

Kattankulathur, Chengalpattu District, Tamilnadu.

**Abstract—**Multimodal neuroimaging analysis offers promising potential for enhancing the diagnosis of Alzheimer's disease and its subtypes. This work proposes an end-to-end framework that integrates structural MRI, synthetically generated PET images, and cognitive test scores to improve diagnostic accuracy and cognitive phenotype characterization. A 2D CycleGAN model is employed to synthesize PET scans from MRI inputs, preserving structural and metabolic consistency, as validated by SSIM and PSNR metrics. The synthesized PET images are labeled using a fine-tuned Inception-based classifier trained on the Kaggle OASIS dataset, achieving a classification accuracy of 98.08%. Patient-level features from both MRI and PET modalities are extracted using ResNet, while cognitive test scores are used to assign phenotypes through a decision tree classifier. Feature-level fusion is then performed to cluster patients into severity-based groups, highlighting the added value of multimodal integration. This framework demonstrates the potential of multimodal pipelines for Alzheimer's diagnosis, providing a foundation for future clinical decision support systems.

**Keywords—** Multimodal fusion, Alzheimer's diagnosis, PET synthesis, CycleGAN, cognitive phenotyping, decision tree, ResNet, Inception, clustering, MRI.

## I. INTRODUCTION

The most common cause of dementia worldwide, Alzheimer's disease (AD) is a chronic, irreversible neurodegenerative illness that mainly affects older persons. The progressive decline in cognitive abilities, including as memory, language, executive functioning, attention, and visuospatial skills, is what defines the disorder. Patients may eventually need full-time care as their sickness worsens and they are unable to carry out basic daily tasks. The World Health Organization [1] estimates that as of 2021, there were more than 57 million dementia sufferers worldwide. By 2050, this number is expected to rise to 139 million. 60% to 70% of these cases are caused by Alzheimer's disease [2], which makes it a significant cause of dependency and disability in older people.

From a neuropathological standpoint, AD is characterised by the buildup of two important biomarkers in the brain: neurofibrillary tangles, which are intracellular clumps of hyperphosphorylated tau protein, and beta-amyloid plaques, which are extracellular deposits of amyloid-beta peptides. Particularly in areas linked to memory and learning like the hippocampus and cerebral cortex, these aberrant proteins impair synapse function and neuronal communication, which eventually results in widespread cell death and brain shrinkage [3].

It is recognised that the brain's pathological alterations start decades before symptoms appear. There is a crucial window for intervention during this protracted preclinical stage. However, because the early symptoms are subtle and might be confused with other non-neurodegenerative causes or normal ageing, diagnosis is sometimes delayed. As a result, substantial and frequently irreparable brain damage has already taken place by the time AD is identified by a clinical observation. Neuropsychological evaluations, functional imaging such as PET scans, and structural brain imaging such as MRI are examples of conventional diagnostic techniques. Commonly used instruments for clinical staging and evaluation of cognitive impairment include the Mini-Mental State Examination (MMSE) and the Clinical Dementia Rating (CDR) [4]. Even though these tools are widely used, they might not identify the disease in its early stages and could be impacted by cultural or educational variables. Moreover, biological validation via PET imaging or cerebrospinal fluid (CSF) biomarkers is costly and not always practical in clinical settings, especially in low-resource settings. Therefore, non- invasive, cost-effective, and scalable tools that can integrate multiple modalities—such as brain imaging and cognitive scores—are urgently needed for early detection, personalized assessment, and prognosis of Alzheimer’s disease. This article presents a deep learning-based approach for Alzheimer's disease identification in order to overcome these issues utilising the OASIS-3 dataset [5], which combines clinical, MRI, and PET data. Decision trees and cognitive scores allow phenotype classification based on particular declines in mental ability, while ResNet-extracted features and clinical data are clustered via K-Means to analyse disease progression. In addition to promoting the development of more efficient and reasonably priced healthcare instruments for Alzheimer's disease, the planned effort aims to improve individual assistance and enable early identification.

## II. RELATED WORKS

Alzheimer’s Disease (AD), a progressive neurodegenera- tive disorder, remains a major challenge for early diagnosis. Multimodal fusion has emerged as a powerful approach for capturing complementary information from multiple imaging modalities. Cai et al. [6] proposed a 3D Multi-Modal Feature Interaction Fusion Network (3D-MFIFN), which effectively combines MRI and PET modalities to identify early-stage AD. The model leverages a cross-attention module to empha- size the most informative features, outperforming traditional fusion methods. Similarly, Qiu et al. [7] developed a 3D multimodal fusion network that incorporates disease-induced joint learning mechanisms, enhancing the interpretability and robustness of feature representations.

Incorporating both local and global contextual features is essential for robust classification. Jabason et al. [8] designed a lightweight deep convolutional neural network that captures multi-scale context features from structural MRI. Their architecture offers a balance between performance and computational efficiency, making it suitable for clinical applications. A similar approach is adopted by Li et al. [9], who introduced 3D PMNet, which fuses PET and MRI features and performs hierarchical classification. Their results demonstrate superior accuracy in distinguishing AD stages.

A thorough analysis of traditional techniques for brain MRI classification, such as GLCM, PCA, and wavelet transformations, was given by Poernama et al. [10]. Even though these techniques require less processing power, they frequently perform worse on their own than deep learning techniques. A range of feature extraction methods were also reviewed by Rai et al. [11], who emphasised the significance of these preprocessing processes in improving the interpretability and effectiveness of AD classifiers. The Intuitionistic Fuzzy Random Vector Functional Link Network was proposed by Malik et al. [12] to represent nonlinearity and uncertainty in neuroimaging data. Compared to conventional neural networks, their results demonstrate high classification accuracy with fewer parameters. In the meantime, Liu et al. [13] integrated numerous MRI sequences and PET to concentrate on multimodal neuroimaging feature learning.

Qu et al. [14] introduced a multimodal feature repre- sentation and fusion framework for early AD diagnosis, which uses modality-specific encoders followed by a shared decoder for reconstruction and classification. Feng et al.[15] proposed a region-of-interest (ROI)-based Contourlet Subband Energy extraction technique that exploits spatial frequency information from structural MRI for effective clas- sification, demonstrating the value of combining handcrafted and data-driven features.

In feature extraction, deep residual learning is still essential. In order to extract deep hierarchical features from brain MRI for classification, Munipalli and Annepu [16] used ResNet-152V2. By combining deep and handmade characteristics, Hasan et al. [17] went one step further and showed that hybrid models are superior to single-method approaches in capturing minor structural changes in brains affected by AD. In order to improve model robustness, Zhang et al. [18] introduced a Unified Multi-Modal Image Synthesis model that uses paired data to generate missing modalities. A Confidence-Guided Aggregation and Cross-Modality Refinement approach was presented by Peng et al. [19] for the synthesis of MR images, allowing for reliable and consistent fusion in the absence of specific modalities.

Generative models have increasingly been employed for tasks such as denoising and data augmentation. Sowjanya et al. [20] applied CycleGAN for unsupervised medical image denoising, particularly for enhancing low-dose CT scans. Wang et al. [21] further advanced this idea with a two-stage generative model combining CycleGAN and joint diffusion, showing improved performance in MRI-based brain tumor detection, which may be extended to AD diagnosis.

Qi et al. [22] introduced ResX, a novel feature extraction block optimized for segmentation tasks. Yang et al. [23] explored convolutional neural networks (CNNs) for tumor image feature extraction, emphasizing the effectiveness of spatial feature encoding. These methods, while focused on other pathologies, provide transferable techniques relevant to AD studies. Zhong et al. [24] proposed an unsupervised cross-modality image generation and registration method that fuses misaligned PAT and MRI images. Though developed for differ- ent imaging types, the underlying framework can be adapted to MRI/PET fusion in AD diagnosis.

By combining MRI and PET scans with clinical data, the suggested technique offers a multimodal approach in contrast to previous studies in the field of Alzheimer's diagnosis and the identification of phenotypes. This work employs deep learning-based feature extraction using the ResNet Model and deep learning-based picture synthesis using the CycleGAN model for missing images, whereas previous work frequently focuses on single-modal inputs or static diagnostic models. To find the most important and pertinent aspects to process, feature selection has also been done on the clinical data. K-Means clustering has been used to group the features that were taken from clinical data and multimodal images. Depending on the severity of the cognitive deterioration, each group has been classified as moderate or severe.Additionally, the inclusion of a decision tree for cognitive phenotype classification provides insights into patient specific symptom patterns. This combined strategy not only improves predictive performance but also contributes to clinical applicability by supporting both quantitative prediction and qualitative understanding of disease progression.

## III. PROPOSED METHOD

By combining several modalities, including MRI scans, PET pictures, and clinical data, the proposed approach offers a thorough pipeline for early Alzheimer's diagnosis that enhances diagnostic interpretability and reliability. The method starts by classifying patients into four groups using a labelled MRI dataset from Kaggle [25]: Non-Demented, Very Mild Demented, Mild Demented, and Moderate Demented. This MRI data is used to train a deep learning model built on the Inception architecture, which learns discriminative features and offers preliminary categorisation. To fill in the gaps in the dataset and allow for multimodal analysis, missing scans are generated using an image synthesis model based on CycleGAN. Once all the data is available, feature extraction and fusion is applied. The features of Alzheimer’s detected patients are grouped into severity levels using K Means clustering. Further classification of patients based on phenotype is performed using cognitive test scores and decision tree classification.

### *Dataset Description*

The dataset utilized for this work was obtained from the Open Access Series of Imaging Studies (OASIS-3), a longitudinal neuroimaging initiative [5] aimed at freely providing multimodal brain imaging data to the research com- munity. The OASIS-3 dataset comprises imaging and clinical data collected from cognitively normal and cognitively impaired older adults, including individuals diagnosed with Alzheimer’s disease. For this work, only T1-weighted MRI scans [26] were selected for consistency in structural imaging analysis.

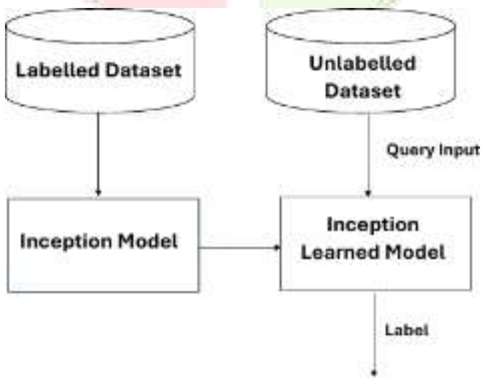
***System Architecture***

A multi-stage deep learning pipeline for Alzheimer's identification is depicted in the architecture diagram in Figures 1(a) and (b), which integrates clinical scores with MRI and PET imaging data. In order to facilitate early and accurate diagnosis, it demonstrates how data flows from image preprocessing and feature extraction via multimodal fusion, classification, clustering, and phenotype mapping. It demonstrates how a pretrained inception model, which was trained using an already-existing labelled MRI dataset [25], is used to label the OASIS dataset [5]. A CycleGAN model is used to synthesise missing PET pictures, and similarity scores are used to validate the synthesis.

A pre-trained ResNet model, which transforms high-dimensional medical images into condensed and significant feature vectors, is used to extract features from the MRI and PET images. In order to preserve the most instructive characteristics, clinical data is simultaneously preprocessed and put through feature selection procedures. Following extraction, each patient's unique characteristics are captured by fusing the features from the three modalities into a single comprehensive feature vector. The various degrees of Alzheimer's severity are reflected in the clustering that follows from this fused picture. A phenotype decision tree is used to further process the clustered data, making it possible to clearly map feature patterns to certain phenotypic outcomes.

***PET Image Synthesis***

To address the issue of missing PET data in the mul- timodal dataset, a synthetic image generation strategy was employed. A Cycle-Consistent Generative Adversarial Net- work [28] as shown in the Fig.2. was trained using the available MRI-PET image pairs. The model was configured to learn the mapping between T1-weighted MRI images and AV45 PET images in an unsupervised manner, utilizing cycle consistency loss to ensure structural fidelity during translation. Once trained, the CycleGAN model was used to generate synthetic PET scans for the missing PET data. The quality of the synthesized PET images was evaluated using multiple image similarity and structural accuracy metrics. These included Structural Similarity Index Measure (SSIM) as shown in the Equation 1, Peak Signal-to-Noise Ratio (PSNR) as shown in the Equation 2 and Dice Similarity Coefficient (DICE) as shown in Equation 3.



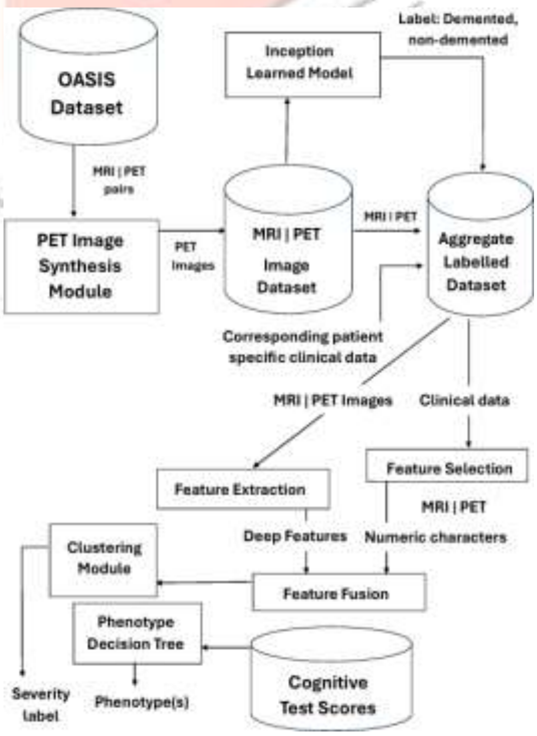




Fig 1(a): Labelling using Inception Learned Model

Fig. 1(b): System Architecture

$$PSNR = 10 \times \log_{10}\left(\frac{MAX_I^2}{MSE}\right) \qquad (1)$$

where MAX_I is the maximum pixel value of the image and MSE is the mean squared error between the real and generated images

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \qquad (2)$$

where $\mu_x$ and $\mu_y$ are the mean intensities, $\sigma_x^2$ and $\sigma_y^2$ are the variances, and $\sigma_{xy}$ is the covariance of the two images. $C_1$ and $C_2$ are small constants to stabilize the division.

$$\text{Dice Score} = \frac{2 \times |A \cap B|}{|A| + |B|} \qquad (3)$$

where A and B are the sets of pixels from the ground truth and predicted segmentation, respectively.

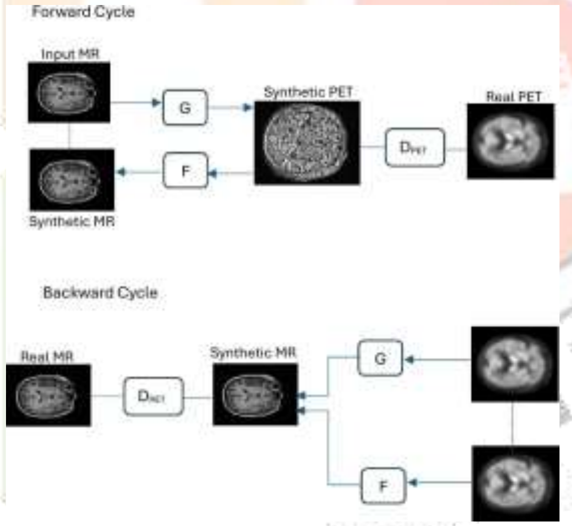



Fig. 2: CycleGAN Architecture

***Data Labeling Using Inception Model***

As the OASIS-3 dataset [5] does not provide categorical labels for disease severity directly aligned with the imaging data, a supervised classification model was first trained using an external labeled dataset [25] to generate labels for the MRI scans. This labelled dataset contains annotated T1- weighted MRI images labeled as Mild Demented, Very Mild Demented, Moderate Demented and non-demented. For the purpose of binary classification, all three Demented categories were consolidated into a single Demented class, while the non-demented class was retained as-is.

An Inception [29] module-based convolutional neural net- work was employed due to its architectural efficiency in capturing multi-scale features. The Inception architecture is designed to approximate an optimal sparse local structure within convolutional neural networks by allowing the use of multiple filter sizes (1×1, 3×3 and 5×5) within a single block. This multi-path approach improves representational power and reduces computational overhead.

Following training, the model was applied to the 1051 T1- weighted MRI images from the OASIS dataset to predict binary labels. These labels were subsequently propagated to the corresponding PET images and clinical records of the same patients using unique subject identifiers, ensuring consistency across modalities.
.

***Feature Extraction and Selection***

The ResNet architecture was used for deep feature extraction in order to recover high-level representations from the neuroimaging data. Both T1-weighted MRI and AV45 PET scans were subjected to feature extraction using this model independently. Therefore, for every patient, each modality generated a 1024-dimensional feature vector. Only patients who had been classified as "Demented" were subjected to feature extraction, guaranteeing that the emphasis was on describing the disease's presentation across modalities. These characteristics were eventually applied to phenotypic analysis and grouping in multi-modal fusion.

In addition to image-based features, clinical data from OASIS-3 was also incorporated. The raw clinical dataset consisted of over 300 features per patient, covering demo- graphic information, cognitive test scores, genetic markers and diagnostic history. To ensure interpretability and efficiency, a multi-step feature selection process was applied [30]. This process involved correlation filtering, missing value thresholding, variance analysis and semantic relevance. This dimensionality reduction was crucial for enhancing the robustness of downstream multi- modal fusion, reducing computational complexity and ensuring that the selected features maintained clinical significance

***Clustering and Severity Labeling***

To investigate potential subtypes within the demented population, clustering using K Means was applied using the combined clinical and imaging feature set. The feature matrix consisted of 1024-dimensional vectors extracted from MRI images per patient, 1024-dimensional vectors extracted from PET images per patient along with the top 40 clinical variables selected through statistical and domain- based filtering.
To assign semantic meaning to each cluster, an interpretable severity analysis was conducted based on cognitive [31] and behavioral variables as shown in Table I,II,III and IV.
Average scores across these metrics were computed per cluster and a composite “severity score” was calculated by summing the mean values. Clusters were ranked based on these scores, with the higher-scoring cluster labeled as Severe and the lower one as Moderate.

TABLE I: Cognitive Functioning

| Variable | Description |
|---|---|
| BILLS | Ability to manage bills – Financial handling capability |
| TAXES | Ability to handle taxes and financial records |
| SHOPPING | Ability to perform shopping independently |
| STOVE | Ability to use the stove safely and prepare food |
| MEALPREP | Ability to prepare meals |
| TRAVEL | Ability to travel independently or navigate transportation |

TABLE II: Daily Functioning

| Variable | Description |
|---|---|
| BILLS | Ability to manage bills – Financial handling capability |
| TAXES | Ability to handle taxes and financial records |
| SHOPPING | Ability to perform shopping independently |
| STOVE | Ability to use the stove safely and prepare food |
| MEALPREP | Ability to prepare meals |
| TRAVEL | Ability to travel independently or navigate transportation |

TABLE III: Neuropsychiatric Symptoms

| Variable | Description |
|---|---|
| BEAPATHY | Behavioral Apathy – Presence of apathy or lack of motivation |
| BEDEP | Behavioral Depression – Indicators of depressive symptoms |
| BEIRRIT | Behavioral Irritability – Signs of frustration, agitation, or irritability |

TABLE IV: Mood Assessments

| Variable | Description |
|---|---|
| GDS | Geriatric Depression Scale – Standardized assessment for depressive symptoms in the elderly |
| SATIS | Life Satisfaction – Self-reported satisfaction with life |
| DEP2YRS | Depression within the past two years – Clinical diagnosis or reporting of recent depression |

***Phenotype Decision Tree***

This section involves the development of a phenotype decision tree aimed at classifying cognitive impairments based on cognitive test scores mainly comprised of Montreal Cognitive Assessment (MoCA) [32], which is essential for identifying key cognitive phenotypes [33] namely, *memory decline* which indicates impairment in memory functions based on test scores, *language deficit* which reflects a decline in language processing abilities, assessed by various cognitive tests and *visuospatial dysfunction* which represents difficulties with spatial awareness and visual processing abilities.

The phenotypes provide valuable insights into cognitive health and help in early detection of cognitive disorders. For each phenotype, a binary classification label was as- signed based on cognitive test results. A label of 1 signifies the presence of the phenotype, while 0 signifies its absence. This classification was determined by evaluating the majority of test scores within each phenotype group.

A decision tree classifier was employed to train the model for classifying the phenotypes. The multi-output nature of the classifier allowed for simultaneous prediction of the three phenotypes. The model was trained on the prepared dataset, which included cognitive test scores and the corresponding phenotype labels.

Once the model was trained and evaluated, it was used to predict phenotypes for new patients. The model takes cognitive test scores as input and outputs the predicted phenotypes, assisting in the assessment of cognitive health.

## IV RESULTS AND DISCUSSION

The results and discussion section presents an in-depth analysis of the proposed method's performance across various evaluation criteria. The effectiveness of the multi-modal approach, including the synthesis of PET images, cognitive phenotype classification and clustering, is assessed using several metrics. The comparison of results is made in terms of accuracy, computational efficiency and clinical relevance, with a focus on how the integrated approach enhances diagnostic outcomes. The findings are interpreted in the context of their implications for future work and potential clinical applications. The discussion also highlights the challenges faced, the insights gained and the potential improvements that could be made in subsequent iterations of the work. Through this comprehensive evaluation, the strengths and limitations of the proposed method are explored, providing a clear understanding of its overall impact on Alzheimer's disease diagnosis and cognitive phenotype classification.

### *Dataset*

After aligning the dataset based on subject IDs and imaging session timestamps, a total of 611 subjects with both T1w MRI and AV45 PET scans were identified**.** However, 448 subjects were found to have missing PET data. To address this, synthetic PET images were generated using a CycleGAN model trained on paired MRI-PET samples. These synthesized PET images were later validated using standard image similarity metrics.

A total of 1051 subjects with complete records comprising MRI, PET (real or synthesized) and clinical data were retained for further analysis. Eight subjects were excluded due to missing clinical information, resulting in a finalized multimodal dataset with 1051 unique patient entries.

### *PET Synthesis Evaluation*

Among the 1051 patients with complete clinical and MRI data, only 611 subjects had both T1-weighted MRI and AV45 PET scans available. For the remaining 448 patients, corresponding PET images were absent which were synthesized using CycleGAN.

The performance of the PET synthesis method is evaluated using various metrics to assess the quality of the synthesized images. The comparison between real and synthesized PET images as shown in Fig.3(a) and Fig.3(b) respectively is conducted through visual inspection, where qualitative dif- ferences are analyzed. To quantify the synthesis quality, two key metrics, SSIM (Structural Similarity Index) and PSNR (Peak Signal-to-Noise Ratio and DICE scores are computed which scores of 0.53, 15.79 and 0.42 respectively. These metrics offer insight into the similarity between the real and generated PET images, with higher SSIM and PSNR values indicating better synthesis quality.

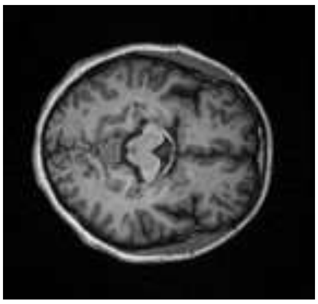


Fig 3.(a) Real MRI

Fig 3.(b) Synthesised PET from Real MRI

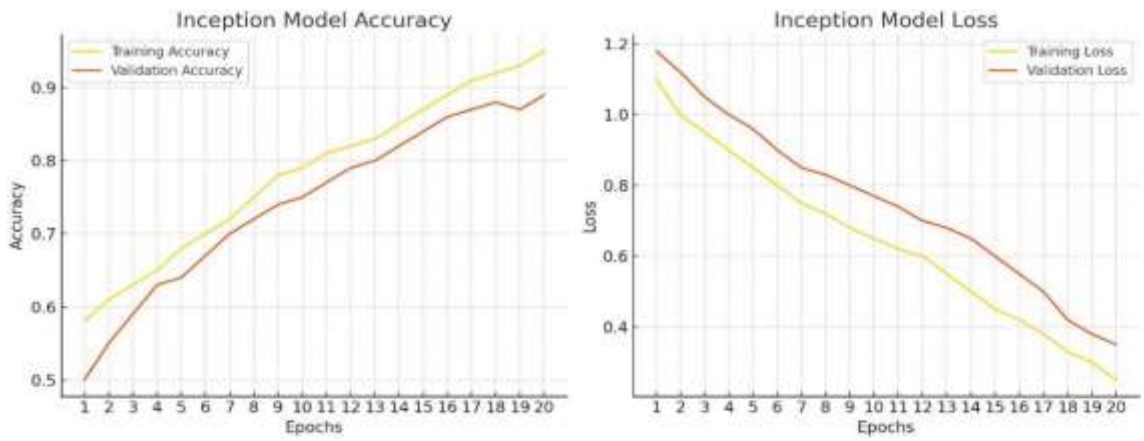



Fig. 4: Inception Model

These evaluations are critical for understanding how well the synthesized PET images align with the original data, ensuring that the synthesized images are suitable for subse- quent analysis and diagnosis.

***Inception Model Accuracy***

The performance of the Inception model as shown in Fig.4 is evaluated based on both accuracy and loss over 20 epochs. The model achieves an impressive accuracy of 0.9808, reflecting its strong ability to correctly classify the dataset. The final training loss is 0.0578, indicating that the model converged effectively and minimized the error during training.
The training and validation loss plots further illustrate the model's learning process, showing how the loss de- creases over epochs, which suggests the model's steady improvement in predicting the correct labels. The label distribution for the dataset reveals that the majority of the samples belong to the demented class (671 samples), while a smaller portion corresponds to the non-demented class (380 samples). This distribution helps provide context to the classification performance and confirms the model's ability to handle class imbalance, as reflected in the overall accuracy. The high accuracy achieved by the Inception model demonstrates its potential for use in clinical applications, where distinguishing between demented and non-demented individuals is critical.

| Name | Precision | Recall | F1 Score | Sample |
|---|---|---|---|---|
| Memory Decline | 0.99 | 1.00 | 1.00 | 157 |
| Language Deficit | 1.00 | 1.00 | 1.00 | 120 |
| Visuospatial Dysfunction | 1.00 | 1.00 | 1.00 | 110 |

TABLE V: Evaluation Metrics

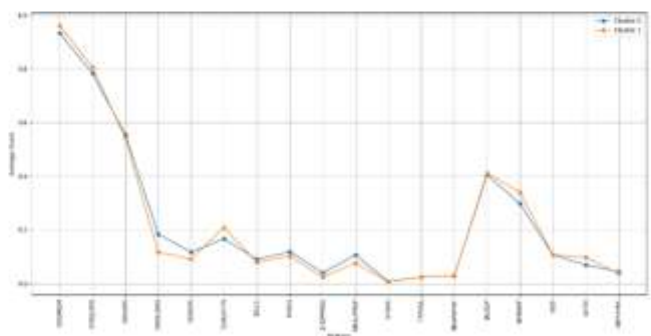


Fig. 5: Cluster Line Graph

### *Clustering Results*

K-Means clustering with cluster size 2 was applied, aim- ing to separate the demented cohort into two distinct sub- groups. A line graph as shown in Fig.5 comparing average feature scores across clusters confirmed the severity gradient, with the severe group exhibiting worse performance in most domains. This analysis enabled a data-driven labeling of Alzheimer's severity without relying on clinician-imposed categories. In the final cluster distribution, Cluster 0 (Severe) had 285 samples and Cluster 1 (Moderate) had 386 samples.

### *Phenotype Decision Tree Outcome*

Precision, recall, and F1-score for each phenotype category were among the important metrics used to assess the phenotype decision tree. Table V displays the outcomes. The model demonstrates its remarkable capacity to accurately categorise patients into the right phenotype categories by achieving precision, recall, and F1-scores for all phenotypes. While the macro-average and weighted average F1-scores likewise show flawless performance, the micro-average F1-score of 1.00 shows that the model performs equally well across all classes.

## V FUTURE WORK

Future efforts will focus on refining label quality by integrating expert clinical validation and exploring semi- supervised learning techniques to improve the accuracy and robustness of phenotype and disease classification. Manually verifying labels with clinician input will help correct potential annotation errors, while leveraging unlabeled data through semi-supervised approaches can further enhance model performance with minimal additional annotation effort.

To improve the realism and clinical utility of synthesized PET images, future work will investigate advanced generative models such as 3D CycleGAN and diffusion-based architectures. These models are expected to better capture the spatial complexity of brain scans. Additionally, steps toward real-world deployment will include the development of a secure, web-based diagnostic platform that allows clinicians to interact with the tool in practice.