Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 21 No. 8 (2025) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v21i08.54363

PAPER

A Hybrid Model for Alzheimer's Disease Classification Based on Neural Network Architectures Enhanced by GAN Model

Iliass Zine-dine(✉), Jamal Riffi, Khalid El Fazazy, Ismail El Batteoui, Mohamed Adnane Mahraz, Hamid Tairi

Laboratory of Informatics, Signals, Automatics and Cognitivism (LISAC), Faculty of Sciences Dhar El Mehraz (FSDM), Sidi Mohamed Ben Abdellah University (USMBA), Fes, Morocco

iliass.zinedine@usmba.ac.ma

ABSTRACT

Alzheimer's disease (AD) is a neurodegenerative disorder marked by progressive cognitive decline, making early and accurate diagnosis vital for timely intervention. This study explores the efficacy of combining generative adversarial networks (GANs), convolutional neural networks (CNNs), and vision transformers (ViTs) for AD classification using magnetic resonance imaging (MRI) data. GANs were employed to generate synthetic brain images, addressing data scarcity by augmenting the dataset. CNNs were then used for feature extraction, accelerating model training, and mitigating overfitting. These extracted features were subsequently fed into ViTs, known for their ability to capture spatial dependencies in image data. Experimental results demonstrated that the proposed GAN-CNN-ViT fusion model achieved high accuracy (96%) and robustness, outperforming traditional machine learning (ML) and deep learning approaches. GAN-generated synthetic images enhanced dataset generalization, improving ViT performance in distinguishing AD patients from healthy controls. Comparative analyses validated the superiority of this approach over recent methods in AD classification. This framework underscores the potential of deep learning techniques in advancing neuroimaging-based disease diagnosis. It holds significant promise for early AD detection, ultimately contributing to improved patient outcomes and quality of life through the integration of cutting-edge computer vision and ML methodologies in medical applications.

KEYWORDS

Alzheimer's disease (AD), generative adversarial networks (GANs), VGG-16, resnet50, vision transformers (ViTs)

1 INTRODUCTION

Alzheimer's disease (AD) is a progressive neurodegenerative disorder marked by cognitive decline and memory loss, affecting millions of individuals worldwide [1].

Zine-dine, I., Riffi, J., Fazazy, K.E., Batteoui, I.E., Mahraz, M.A., Tairi, H. (2025). A Hybrid Model for Alzheimer's Disease Classification Based on Neural Network Architectures Enhanced by GAN Model. *International Journal of Online and Biomedical Engineering (iJOE)*, 21(8), pp. 23–40. https://doi.org/10.3991/ijoe.v21i08.54363

Article submitted 2025-01-12. Revision uploaded 2025-02-28. Final acceptance 2025-02-28.

According to statistics, it accounts for approximately 60–70% of all dementia cases and is one of the leading causes of disability and dependency among the elderly population. With the aging global population, the prevalence of AD is expected to rise significantly in the coming years, presenting substantial challenges for healthcare systems and caregivers. Early diagnosis and intervention are crucial for managing the symptoms and improving the quality of life for individuals affected by Alzheimer's disease [2].

Alzheimer's disease, a prevalent neurodegenerative disorder highlighted by the World Health Organization, presents a multifaceted challenge in public health. Within the broader spectrum of dementia with Alzheimer's, distinct subgroups delineate the progressive nature of the disease. These subgroups encompass mild cognitive impairment, often serving as a precursor to dementia, and mild dementia, characterized by subtle cognitive deficits affecting daily activities [3]. Moderate dementia poses more substantial challenges, necessitating enhanced support due to heightened cognitive impairments and behavioral changes. Severe dementia represents the advanced stage, where individuals may experience profound communication difficulties, loss of bladder control, and significant functional limitations, warranting comprehensive care strategies [4]. Understanding the nuanced progression and subgroups of AD is pivotal for tailored diagnostic approaches and effective management strategies.

Alzheimer's disease presents a hard challenge in research endeavors aimed at early prediction due to its intricate pathology and the structure of the nervous system's tissue and cells. However, the integration of artificial intelligence and sophisticated computer vision methods holds promise for addressing this challenge [5]. By harnessing the power of machine learning (ML) algorithms and advanced image analysis techniques, it becomes feasible to discern subtle patterns and features from medical imaging data. This amalgamation of artificial intelligence and computer vision methodologies offers a novel approach to enhance early detection and intervention strategies for AD [6].

Statistical ML techniques have shown early success in the automated detection of AD. However, the current trend leans towards the adoption of deep learning models and sparse auto-encoders for more robust performance. These approaches face several challenges in feature extraction and classification tasks, including the requirement for massive datasets to achieve optimal performance, the inherent complexity of deep learning models, and their sensitivity to poor-quality data. These drawbacks underscore the need for further research to address these limitations and enhance the efficacy of deep learning techniques for such tasks [7].

The task of classifying AD presents considerable challenges due to its dependence on various criteria related to the intricate structure of nervous system tissues and cells. However, advancements in technology and the digitization of medical tools have significantly enhanced the ability of healthcare professionals to accurately detect. The domain ML is dedicated to the pursuit of various computational tasks for many researchers [8]. Within this landscape, DL emerges as the forefront approach that has shown superior performance over traditional ML in identifying complex structures in complex, high-dimensional data, particularly in the field of computer vision. In this regard, transfer learning methods have garnered considerable interest in medical image classification due to their ability to leverage pre-trained models and adapt them to new tasks with limited labeled data, by transferring knowledge learned from large-scale datasets, transfer learning methods hold potential for enhancing diagnostic accuracy and clinical decision-making in healthcare applications [9]. Another recent classifier models namely capsule network methods, demonstrates promising capacity in image classification by capturing intricate spatial relationships and hierarchically organizing features, leading to enhanced diagnostic

accuracy [10]. Their ability to encode pose and instantiation parameters offers a robust representation of anatomical structures, facilitating precise disease detection and classification in medical imaging applications [11]. Recently, new research has been based on the feature extraction methods, such as convolutional neural networks (CNNs) and vision transformers (ViTs), which excel in capturing intricate patterns and semantic information from medical images, leading to improved classification accuracy. Additionally, auto-encoders facilitate unsupervised learning by extracting meaningful representations from raw image data, aiding in feature discovery and dimensionality reduction. Another technique, namely generative adversarial networks (GANs), offer a unique capability to generate synthetic medical images, enabling data augmentation and enhancing the diversity of training datasets for robust classification models in medical imaging tasks [12].

In this study, we present a novel approach for AD classification utilizing GANs, CNNs, and ViTs. Our approach aims to tackle key challenges in medical image analysis by leveraging GANs to generate synthetic images for data augmentation, CNNs to extract features, accelerate model training ViT and prevent overfitting, and ViTs to capture complex patterns. The contributions associated with our approach include robustness in the face of variability, improved classification accuracy, and the potential for early detection of Alzheimer's disease.

The subsequent sections of this paper are structured as follows: The second section provides an overview of related studies in the field. In the third section, we detail the methodologies and techniques employed in our approach. Following this, the fourth section presents the experimental results obtained from our approach. Finally, the fifth section concludes with a summary and outlines directions for future research endeavors.

2 RELATED WORK

The objective of this section is to critically examine prior research pertaining to the utilization of feature extraction, ML, deep learning models, auto-encoders, and transformers for the identification and classification of AD. Numerous studies within this domain focus on early disease prediction utilizing non-invasive computer-aided diagnostic (CAD) methods, thereby circumventing the need for surgical or invasive procedures. Illán et al. [13] introduced a fully automated CAD system aimed at enhancing the precision of early AD diagnosis. Their approach entails an initial automatic feature selection process, coupled with a combination of component-based support vector machine (SVM) classification and a classifier assembly vote passing technique SVM. This methodological framework represents a significant contribution to the field, offering potential improvements in the accuracy and efficiency of AD diagnosis. In this regard, there are several image-processing architectures are interested in detecting and classifying diseases. Our present study centers on the early prognostication of AD, aligning closely with the referenced work. These studies are situated within the same medical domain and employ comparable techniques and methodologies of computer vision and infographics for this purpose. Nonetheless, our proposed approach yielded compelling outcomes, outperforming numerous state-of-the-art experiments in the domain of AD classification in terms of accuracy.

Numerous deep CNNs have been previously trained and are being employed to extract profound features from magnetic resonance (MR) images. Waleed [14] proposed a deep learning solution employing two CNN models for diagnosing and classifying AD. In this regard, Deepanshi et al. [15] applied convolutional CNN models including VGG-19, Inception-V3, ResNet-50, DenseNet-169, and a custom CNN

utilizing transfer learning with pre-trained weights from the ImageNet model. This comparative analysis aimed to identify the most effective model in terms of performance for their specific task and present the corresponding results. Although these methods yield precise results, their performance tends to be suboptimal when confronted with extensive databases. Tooba Altaf et al. [16] used an approach based on feature extraction through various methods, including the gray level co-occurrence matrix (GLCM), scale-invariant feature transform (SIFT), local binary pattern (LBP), and oriented gradient histogram (HOG), to generate a hybrid feature vector to improve the effectiveness of texture. In computer vision, an alternative approach yielding improved classification outcomes involves combining methods such as concatenation and fusion of vectors derived from extracted features. Madusanka et al. [17] employed a fusion-based approach, combining texture and morphometric features, to explore a potential diagnostic biomarker AD, their study aimed to demonstrate improved classification performance through the integration of these features. Another approach, focusing on capturing direct relationships between images, might prove more effective for brain image analysis compared to CNNs. Yanjun Lyu et al. [18] utilized ViTs as the backbone architecture, first pretraining it on the ImageNet-21K dataset, and then transferring it to the brain imaging dataset to demonstrate improved classification performance. In this context, convolutional autoencoders represent a commonly utilized neural network architecture in image processing tasks. Their primary function involves capturing spatial relationships within the input data to facilitate various image-processing applications. Francisco J et al. [19] introduced a novel exploratory analysis of AD data employing deep convolutional autoencoders. These autoencoders are utilized to learn a compressed representation (encoding) of the input data, which is subsequently reconstructed (decoded) to closely resemble the original input. This methodology enables prediction and classification of the disease based on the learned representations of the data. Linfeng Liu et al. [20] presented an alternative approach utilizing transformers called Multi-Modal Mixing Transformer (3MT). This transformer model is specifically designed for disease classification tasks and is capable of handling multi-modal data as well as scenarios involving missing data.

Another architecture, namely capsule network models, focused on extracting richer features from image datasets, prioritize capturing hierarchical relationships within the data. Kruthika et al. [21] introduced a content-based image retrieval system for early Alzheimer's detection, employing a combination of 3D capsule network, 3D CNN, and pre-trained 3D autoencoder technology. Their approach underscores the importance of leveraging capsule networks to enhance feature extraction capabilities in medical imaging tasks.

The feature extraction and classification methods mentioned earlier exhibit limitations such as data dependency, limited interpretability, and computational intensity. To address these shortcomings, ongoing research efforts are required to develop computer vision methods that are more interpretable, data-efficient, and computationally lightweight for feature extraction from images. Given the aforementioned limitations, it is advisable to explore novel approaches that surpass these constraints and yield improved prediction scores.

The ability to deliver real-time performance is crucial in medical diagnostics, especially during emergencies where prompt and accurate decisions are vital for effective patient care and prognosis. Assessing a model's inference time and computational requirements is essential to determine its practicality for real-world applications. Such models must strike a balance between speed and accuracy to ensure dependable diagnostic results without sacrificing computational efficiency [22].

The primary contribution of this study can be outlined as follows: During the preprocessing stage, we applied standard computer vision and graphic techniques to streamline subsequent tasks. Next, we utilized two neural networks, VGG-16 and ResNet50, to extract meaningful and significant features, enhance model training, minimize overfitting, and improve the robustness of the classification framework. The extracted features were concatenated into a single vector and used as inputs for our ViT classifier. To validate the proposed methodology, we systematically evaluated the results at each stage of our process to identify areas for improvement. Our approach demonstrated significant potential in the early detection of AD, leveraging advancements in computer vision and ML technologies and their applications in the medical field.

3 RESEARCH METHODOLOGY

In this section, we introduce the overall architecture of our proposed method. Figure 1 illustrates the detailed architectural division of our proposed method for classifying AD. This approach consists of three primary modules: data preprocessing, neural network models combining (GAN, VGG-16, and Resnet50), and ViTs classifiers, each serving crucial roles in the methodology. At the start of this section, we described the dataset used throughout this study (section 3.1). The MRI images undergo initial preprocessing steps before proceeding to the subsequent phase of the methodology (section 3.2). Next, a GAN model was used to generate additional realistic images, and CNNs were employed to extract features, accelerate model training with ViT, and prevent overfitting (section 3.3). Finally, the random forest regressor evaluated the images generated by the GAN and selected the most relevant ones for input into our ViT classifiers (section 3.4). Subsequent subsections provide a detailed discussion of the three essential elements.

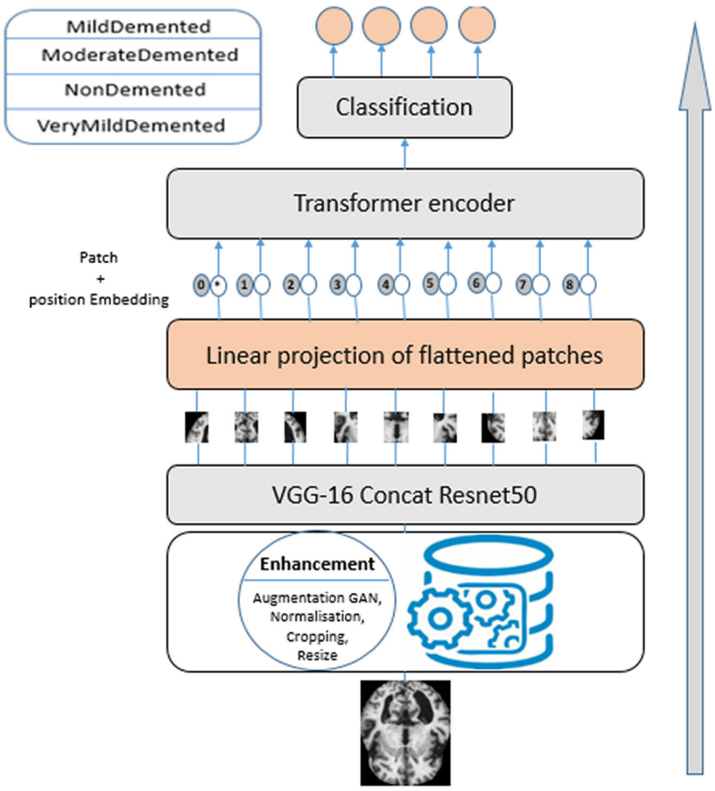

Fig. 1. Proposed architecture of GAN-CNN-Vit for AD classification

3.1 Data description

In this study, data were obtained from the freely available Kaggle Alzheimer's dataset (https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-4-class-of-images/data) [23]. This dataset comprises a total of 6400 2D slices, categorized as follows: 896 slices from mildly mentally impaired individuals, 64 slices from moderately mentally impaired individuals, 3200 slices from non-demented individuals, and 2240 slices from very mildly mentally impaired individuals. These MRI images serve as the primary data source for our analysis, enabling the investigation of various imaging features and patterns associated with different stages of mental impairment in AD. To enrich the analysis with practical examples, we implemented a GANs model. Subsequently, we extracted pertinent features from the generated data by CNN models. Finally, we employed a VIT model for classification purposes. This approach allowed us to leverage the strengths of both GANs and VIT models in the context of our research objectives, facilitating comprehensive analysis and classification of the data.

3.2 Data preprocessing

The application of techniques and methods of preprocessing the dataset in order to have more relevant information that will be exploited in the input of a classifier to predict results has become a trend in the fields of computer graphics and computer vision [24]. In this paper, we will use well-known image-processing methods; the key strategies employed in this section of the treatment are covered below.

- **Data crop:** Nearly all of the images in our brain MRI datasets have undesirable spaces, which results in subpar classification performance. Therefore, it is vital to crop the photographs in order to eliminate unnecessary portions and use only the pertinent information [25]. In this study, we employ the cropping approach, which computes extreme points and returns a geographic subset of an object as specified by an extent object. The application of this method consists of five steps: First, we load the original MR pictures. Secondly, we apply thresholding in order to create binary images; thirdly, we undertake dilation and erosion processes to reduce image noise; and fourthly, we use the threshold images' largest contour to determine the images' four extreme points (extreme top, extreme bottom, extreme right, and extreme left). Finally, we crop the image based on the contour and extreme point data. Bicubic interpolation is used to enlarge the cropped images. Figure 2 illustrates the steps of the cropping method applied to our dataset in this approach.

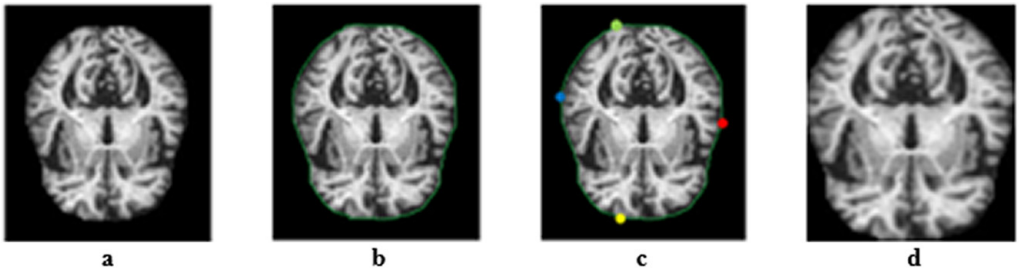

Fig. 2. Image representation of the different stages of cropping images from the dataset: (a) original image, (b) outer contour, (c) edge point; (d) image crop

- **Generative adversarial networks:** Deep learning networks, notably CNNs, demand substantial datasets to ensure effective training and mitigate overfitting issues. Traditional data augmentation techniques predominantly rely on geometric transformations of input images, encompassing rotations, zooming, resizing, noise addition, image translation, and flipping [26]. Nonetheless, these methods may not be optimal for medical image datasets due to their unique characteristics and requirements. Recent data increases are based on the generative adversarial neural network, which is an innovative technique for data augmentation. It is a class of deep learning techniques that creates a real image from noise data to enlarge the size of the dataset in order to ensure a generalizable deep learning model [27].

  A GAN comprises two distinct neural networks, the generator and the discriminator, engaged in an adversarial training process. The generator aims to fabricate synthetic data instances that closely resemble samples from a given dataset, while the discriminator endeavors to differentiate between genuine data instances and those produced by the generator [28]. This dynamic interplay between the generator and discriminator enhances the power of competitive learning, wherein the generator strives to generate increasingly realistic samples to deceive the discriminator, while the discriminator hones its ability to discern genuine from synthetic data [29].

  The innovative approach to data augmentation, termed GANS presents remarkable potential; however, it is accompanied by numerous drawbacks. A primary limitation of GANs lies in their convergence dynamics. The absence of a definitive criterion for terminating generator training renders the process unreliable [30]. Traditional metrics, including the loss function, fail to adequately capture GAN convergence, hindering the determination of when the generator achieves proficiency in generating high-quality synthetic images. Moreover, the GAN's objective function does not directly assess output image quality. To address these challenges, researchers have proposed various solutions, including refining the loss function formulation. Another approach involves replacing the Jensen-Shannon divergence, commonly used in traditional GANs, with the earth mover distance metric. These strategies aim to mitigate GAN limitations and enhance their effectiveness in generating high-quality synthetic data. In this study, we employed the conventional GAN framework over 10 epochs with the objective of achieving improved convergence in the assessment of key metrics, including generator loss and discriminator loss. The Figure 3 below illustrates the functioning of our GAN model.

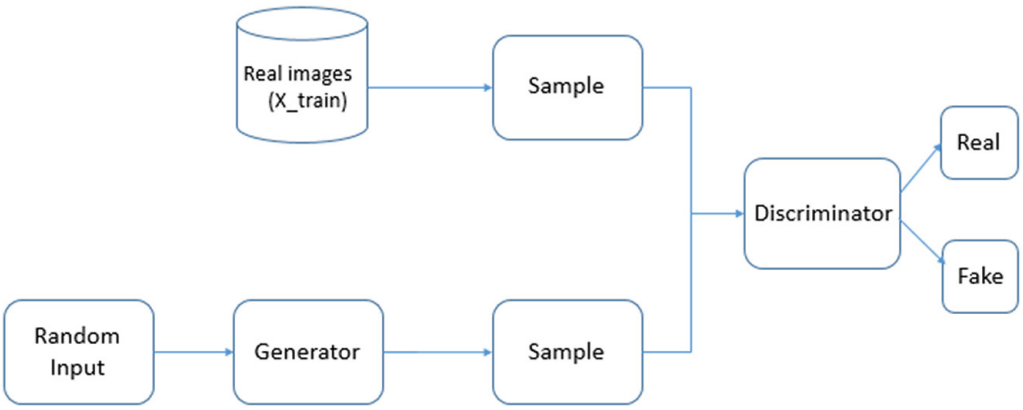

Fig. 3. The operational steps of our GAN model

3.3 Feature extraction

Convolutional neural networks are a class of deep learning models specifically designed for tasks involving image processing and pattern recognition. CNNs excel in feature extraction from images due to their unique architecture, which includes convolutional layers, pooling layers, and fully connected layers. These networks are adept at capturing hierarchical patterns and spatial relationships within images, making them ideal for tasks such as object recognition, image classification, and segmentation [31].

One of the key advantages of CNNs is their ability to automatically learn relevant features. This feature extraction capability accelerates model training ViT by reducing the dimensionality of the input space and focusing on the most discriminative features. Moreover, CNNs can avoid overfitting by incorporating techniques such as dropout and regularization, which help generalize learned patterns to unseen data [32].

In this study, to extract crucial features, accelerate model training with ViT, and prevent overfitting, we have employed VGG-16 and ResNet50, two well-known methods for extracting deep features. We utilized the convolutional, pooling, and normalization layers while excluding the fully connected classification layer. The classification phase will be conducted later in the ViT model. In the following two paragraphs, we define and present VGG-16 and ResNet50, two CNN models utilized in our approach.

Vgg-16 is a CNN model trained on millions of images from different categories. It deals with the classification tasks of image processing. It belongs to the family of models called VGG Net (Visual Geometry Group Network). Deep layers and uniform architecture characterize this model. The model has 16 trainable weight layers, hence the name "Vgg-16." Specifically, the architecture of VGG-16 is composed of 13 convolution layers followed by fully connected layers and output layers [33]. For our task, we utilized the convolutional, pooling, and normalization layers, while excluding the fully connected classification layer to extract deep features. Convolution layers use small filters (3x3) with a stride of 1 and pooling layers with a stride of 2. In the final layer, we aggregated the features extracted by VGG-16 into a vector of dimension (1.1.1000) to concatenate it with the other vector extracted by ResNet50, which has the same dimension.

ResNet50 is a deep learning model used for computer vision applications where weighting layers learn residual functions with reference to the input layer. ResNet architectures were developed to solve the problem of gradients disappearing or exploding when networks become very deep [34]. The ResNet50 architecture is made up of 50 layers composed of residual blocks; these layers are expressed by stages (input layer, initial convolutional block, four main stages, residual block, global average pooling (GAP), and fully connected (FC) Layer). For our task, we employed all these layers except for the fully connected classification layer to extract deep features. Moreover, in the final layer, we aggregated the features extracted by ResNet50 into a vector of dimension (1.1.1000) to concatenate it with the other vector extracted by VGG-16, which has the same dimension.

The vector concatenation technique utilized in this study enhances model performance by incorporating crucial information, facilitating a more comprehensive data representation. This can potentially enhance generalization and foster more dependable classification decisions. Furthermore, by reducing data dimensionality, concatenation aids in mitigating overfitting and enhancing model interpretability.

Hyperparameters: In this study, we utilized the PyTorch library for our experiments. Notably, the VGG-16 and ResNet50 architectures come with predefined kernel size, padding, and stride parameters, which are not explicitly specified during the training phase. Throughout training, the parameters, including weights and biases,

of these models are updated using the optimizer. We employed stochastic gradient descent (SGD) as our optimizer with a learning rate of 0.001 to facilitate model convergence and optimization. To measure the model's performance and guide the training process, we adopted cross-entropy loss as the loss function. Our training protocol involved 10 epochs, each comprising a batch size of 32 instances, ensuring thorough model optimization and robustness evaluation.

Discussion: In our approach, the use of CNNs, specifically VGG16 and ResNet50, offers several advantages in terms of performance and efficiency. VGG16, known for its simplicity and uniformity in layer design, enables the extraction of rich hierarchical features, making it well-suited for tasks requiring detailed feature representation, such as image classification. On the other hand, ResNet50 incorporates residual connections, which address the vanishing gradient problem, allowing for the training of deeper networks. This enhances its ability to model complex patterns while maintaining high accuracy. By combining these two architectures in our framework, we leverage the complementary strengths of VGG16's deep feature extraction and ResNet50's efficient gradient propagation, accelerate training model ViT, and prevent overfitting and improve the overall robustness and precision of our approach.

3.4 Vision transformer classifier

The ViTs is a neural network architecture that was developed for image processing [35]. In contrast to CNNs, which are based on convolution and pooling operations, ViT relies on an attention mechanism used in transformers, attention mechanism is inspired by the human cognitive process of selecting the relevant features of the image and ignoring the irrelevant features rather than concentrating on the whole image [36]. In a ViT, the input image is decomposed into patches, which are then flattened and treated as a sequence of tokens. These tokens are subsequently passed through a series of transformer layers, where they undergo attention operations and non-linear transformations [37]. Ultimately, the ViT generates a vector representation for each patch, which is then utilized for tasks such as image classification. The Figure 4 below illustrates the functioning of our ViT model.

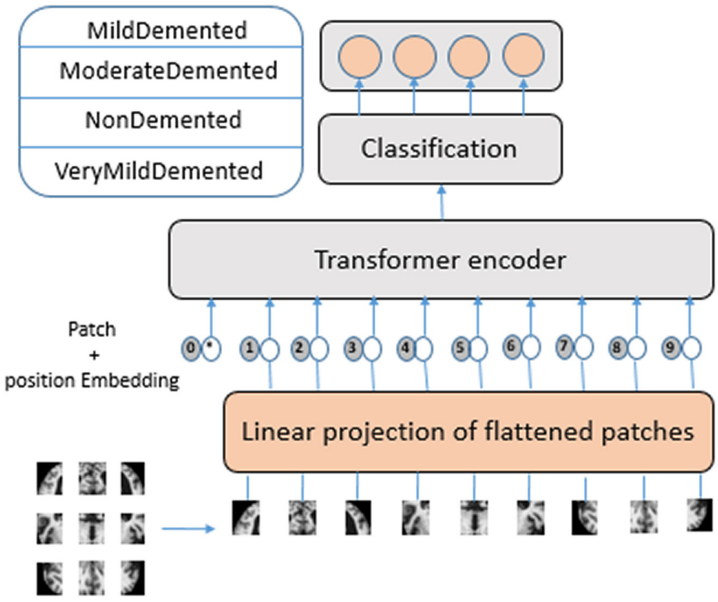

Fig. 4. Representation of the steps of ViT models

Discussion: In our approach, the integration of the ViT model offers several distinct advantages over traditional convolution-based architectures. ViT leverages self-attention mechanisms to capture long-range dependencies and global contextual information from input images, which are often overlooked by CNNs. This ability to model global relationships enables ViT to excel in tasks requiring a holistic understanding of the image. Furthermore, ViT's non-local operations allow it to better generalize across varying image scales and complex patterns without the inductive bias imposed by convolutions. By incorporating ViT into our framework, we significantly enhance the model's capacity for capturing intricate visual features and patterns, ultimately leading to improved performance and accuracy in our studied application.

3.5 Discussion

Our approach centers on a meticulously designed three-phase methodology aimed at maximizing the effectiveness of our scientific research. Initially, we have embarked on the pivotal phase of processing and augmenting our dataset, leveraging the capabilities of a GANs model. This foundational step not only refines the raw data but also imbues it with enhanced diversity realism, thereby fostering more robust and comprehensive feature representations. Following this, we have employed a novel strategy by concatenating the features extracted by the neural network models, VGG16 and ResNet50. This concatenation is designed to leverage the unique strengths of each model, accelerate model training for ViT, and prevent overfitting, ultimately demonstrating superior performance in capturing spatial dependencies in image data for classification purposes. By integrating the comprehensive feature representations generated by VGG-16 and ResNet50, we aim to achieve a more holistic understanding of the underlying patterns within the dataset. Lastly, we have culminated our approach with the utilization of a model, ViT, tailored specifically for classification tasks. This model operates on the premise of extracting spatial characteristics and capturing the intricate relationships inherent within the data, thereby enabling more nuanced and accurate classification outcomes. Through the synergistic integration of these three phases, our approach epitomizes a comprehensive and innovative framework poised to advance the frontiers of scientific inquiry and drive meaningful insights in our domain of study. Real-time performance plays a vital role in medical diagnosis, particularly in emergency scenarios where quick and accurate decisions are critical. In this study, we assessed the model's inference time and achieved classification results in less than one minute, effectively balancing speed and accuracy to ensure suitability for real-world applications.

4 EXPERIMENTS AND RESULTS

In this section, we present the experimental setup and results of our study on AD classification using neural network architectures enhanced by a GAN model. Initially, we provide an overview of our approach and the frameworks used to implement the code. Following this, we detail the evaluation metrics employed to assess the performance of our classification models. Subsequently, we conduct a comprehensive comparison of the different cases in our study. Finally, we discuss our findings and offer insights into future research directions.

4.1 Computational infrastructure

In this study, we utilized specific environments conducive to efficient model development and evaluation. The Google Colab environment served as the primary platform, offering approximately 12.67 GB of RAM and GPU accelerator capabilities. This environment facilitated the execution of computationally intensive tasks involved in the training, validation, and testing of the proposed model. The model development was implemented using TensorFlow 2.6.4 and Keras 2.6.0, which are widely recognized frameworks for deep learning tasks. Leveraging these environments ensured optimal utilization of computational resources and enabled seamless integration of GAN, various CNNs, and ViT models into our hybrid approach for AD classification.

4.2 Results

In evaluating the performance of our classification models, we employed several key metrics, including precision, recall, accuracy, F1-score, loss, and overall accuracy. In the following two paragraphs, we present the performance evaluation and the loss and accuracy curves.

- **Performances evaluation:** The process of evaluating the performance of an ML model on a task involves the use of various metrics and techniques to measure the effectiveness, accuracy, and generalization ability of a trained model. The goal of applying these metrics is to understand how well the model is likely to perform on new data and provide insights into potential issues such as overfitting or underfitting [38]. The effectiveness of our experiment was assessed using various kinds of performance metrics specific to the classification task, including precision, recall, accuracy, and F1-score.

  Precision: the percentage of results that are relevant and it is defined as:

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{1}$$

  Recall: the percentage of total relevant results correctly classified by the proposed algorithm which is defined as:

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{2}$$

  Accuracy: Formally, accuracy has the following definition:

$$Accuracy = \frac{True\ Positive + True\ Negative}{Total} \tag{3}$$

  F1-score: is a ML metric that can be used in classification models; f1-score has the following definition:

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{4}$$

  In this study, we used the Alzheimer's dataset from a Kaggle competition for an AD classification task to obtain the empirical results of this study. The main

objective of this experiment is to extract the deep features from two CNN models Vgg-16 and ResNet50, then consider these extracted characteristics as inputs in the ML classifier VIT and extract Feature spatial in order to obtain a high-performance classification. In the following paragraph, we present another two famous metrics (loss and accuracy) to evaluate the performance of our model.

- **Accuracy and loss: Accuracy:** Accuracy is a metric of a classification model's efficacy; it is the ratio of correctly predicted instances to the total number of instances. In another sense, precision refers to the model's accuracy rate for predictions. Figures 5 and 6 illustrate the prediction results.

  Loss: Loss is a measure of the performance of the model; it quantifies the difference between the predicted values and the actual values; the total of the errors produced for each example in the training or validation sets constitutes the loss [39]. Therefore, we presume that "the lower the loss, the better the model."

  The empirical findings were derived from the "Alzheimer's Dataset (4 classes of Images)" dataset sourced from the Kaggle competition focused on AD classification tasks. Following the initial pre-processing phase involving computer vision and computer graphics functions, the primary aim of this study was to extract deep features utilizing two distinct CNNs, namely VGG-16 and ResNet50. These extracted features served as inputs for a VIT classification model, which aims to address the limitations observed in CNN models, particularly their inability to capture long-range spatial relationships and data dependencies, thereby enhancing the effectiveness of classification. The subsequent two figures illustrate the performance metrics, comprising accuracy and loss curves, for each neural network CNN and VIT classifier model.

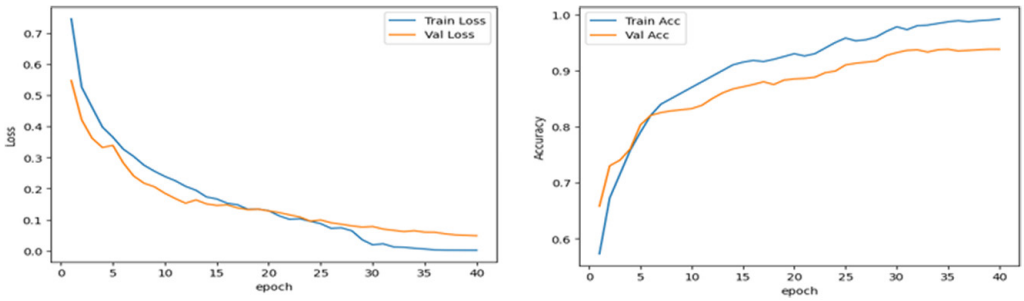

Fig. 5. Resnet50 + Vit architecture's loss and accuracy

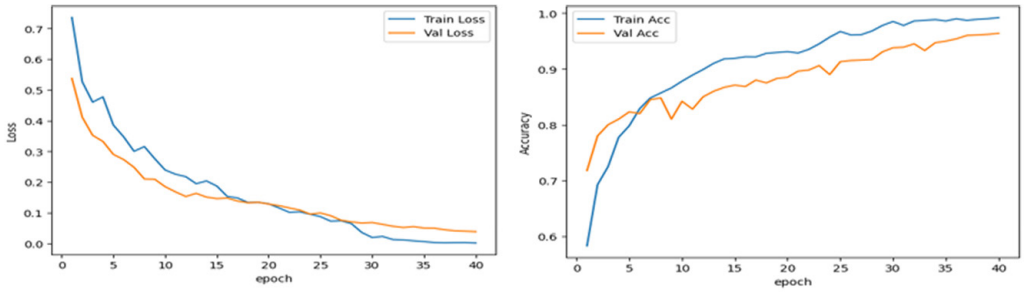

Fig. 6. VGG-16 + Vit architecture's loss and accuracy

In conclusion, our empirical findings demonstrate that employing a single CNN neural network model for feature extraction yields a commendable classification score when compared to prior studies within the same domain utilizing identical datasets and methodologies. To enhance the classification performance further,

we propose a strategy involving the integration of two neural network models, VGG-16 and ResNet50, to extract pertinent and relevant features. Leveraging vector combination as a technique, we effectively supply our classifier with crucial inputs, mitigating potential issues such as overfitting or underfitting. The subsequent figure illustrates the performance metrics, encompassing accuracy and loss curves, of the proposed approach.

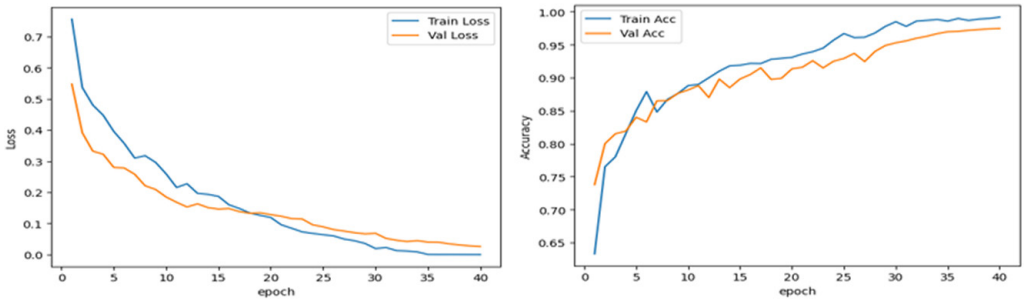

Fig. 7. Our approach architecture's loss and accuracy

Based on the comprehensive analysis of our results, we confidently assert that our approach yielded significant effectiveness, surpassing numerous state-of-the-art experiments. This success can be attributed to the synergistic integration of two neural network models, VGG-16 and ResNet50, complemented by the incorporation of the VIT model, which adeptly captures long-range dependencies and addresses spatial relationships.

4.3 Discussion

Our proposed approach incorporates pre-processing functions for computer vision and infographics, as well as data augmentation using GAN models to generate more realistic examples. By combining the key strengths of the ViT, VGG-16, and ResNet50 models, our method ensures the extraction of both global and local features, thereby enabling a robust feature representation that captures the primary long-range dependencies within the image. Table 1 shows the classification performance of the different steps of our approach.

Table 1. Comparative table

Feature Extraction Method	Classification Methods	Accuracy
Real Data	VGG16	73%
GAN	AlexNet	78%
GAN	DenseNet	87%
GAN	VIT	89%
GAN + ResNet50	VIT	92%
GAN + VGG16	VIT	95%
GAN + (vgg-16 Concat ResNet50) Proposed	VIT	96%

Our hybrid approach exhibits superior computational efficiency and offers distinct advantages compared to both CNN models and transformer methods by

effectively extracting diverse global and local features. Figure 7 illustrates the performance of our proposed model.

The robust performance of our method appears when it integrates a VIT model with the concatenation of two neural network models, VGG-16 and ResNet50t. These models of CNN exhibit notable advantages, including ease of optimization and efficient classification capabilities. However, CNN models lack the ability to encode and capture pixel-level relations within input, attend to relevant features within images, and do not encode the relative positions of different features. Thus, we can conclude that a primary constraint of CNNs lies in their inability to effectively encode long-range dependencies within input images and to assign relative importance to individual features within the image, because effective feature extraction is based on a relevant feature representation that tracks the long dependencies within the image [40]. ViT represents a specialized variant of transformer models tailored for computer vision tasks, integrating a self-attention mechanism to effectively weigh the importance of individual features within an image while maintaining computational efficiency. By dividing the input image into smaller patches, ViT facilitates the processing of both local and global features. Our results demonstrate the ViT's promising performance, accompanied by reduced architectural complexity, enhanced scalability, and efficient feature learning. Nonetheless, ViT's efficacy is contingent upon substantial training data, and optimization remains challenging.

Our approach integrates advanced feature extraction methods with the ViT prediction model to outperform existing techniques in AD classification. Furthermore, we assessed the model's real-time capabilities, achieving classification results within a timeframe of less than one minute and an accuracy of 96%. These results highlight the model's effectiveness and practicality for real-world applications.

5 CONCLUSION AND PERSPECTIVE

In this paper, we have proposed a hybrid model for AD classification based on neural network architectures. Our study focused on three main phases: the dataset processing phase, which contains usual computer vision and computer graphics functions such as normalization, resizing, cropping, and augmentation; the feature extraction phase, where we employed the VGG-16 and ResNet50 models; and the classification phase, where we applied the ViT model to predict the results.

The strength of our approach is illustrated by the intervention of two feature extractors: VGG-16 neural network, and the Resnet50 neural network models, as well as the concatenation between the vectors of the extracted characteristics, and finally the classification with the ViT model.

Finally, our experimental results are remarkable since they demonstrate the ability of the different methods of feature extraction used (VGG-16 and ResNet50) to extract the deep characteristics of MR images and the concatenation between them, as well as the efficiency of the ViT model to capture spatial dependencies in images. This study is aimed toward a promising computer-assisted diagnosis in digital pathology.

Several types of research in this domain of disease classification do not meet the expectations of experts in the medical field because they use methods that have poor performance, are data-hungry, or use deep learning models that have complex computing efficiency. Our future direction also includes the study of other brain disease datasets in order to be able to apply methods that deal with the above limitation.

6 REFERENCES

[1] K. K. R. Rajeswari, H. D. Maheshappa, and A. D. N. Initiative, "CBIR system using capsule networks and 3D CNN for Alzheimer's disease diagnosis," *Informatics in Medicine Unlocked*, vol. 14, pp. 59–68, 2019. https://doi.org/10.1016/j.imu.2018.12.001

[2] R. Farheen *et al.*, "A deep learning approach for automated diagnosis and multi-class classification of Alzheimer's disease stages using resting-state fMRI and residual neural networks," *Journal of Medical Systems*, vol. 44, 2020. https://doi.org/10.1007/s10916-019-1475-2

[3] S. Saman, D. Danielle, A. John, and T. Ghassem, "DeepAD: Alzheimer's disease classification via deep convolutional neural networks using MRI and fMRI," *BioRxiv*, p. 070441, 2016. https://doi.org/10.1101/070441

[4] L. Linfeng, L. Siyu, Z. Lu, T. Xuan Vinh, N. Fatima, and S. Shekhar Chandra, "Cascaded multi-modal mixing transformers for Alzheimer's disease classification with incomplete data," *NeuroImage*, vol. 277, p. 120267, 2023. https://doi.org/10.1016/j.neuroimage.2023.120267

[5] O. M. Al-hazaimeh, A. Abu-Ein, N. Tahat, M. Al-Smadi, and M. Al-Nawashi, "Combining artificial intelligence and image processing for diagnosing diabetic retinopathy in retinal fundus images," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 18, no. 13, pp. 131–151, 2022. https://doi.org/10.3991/ijoe.v18i13.33985

[6] M. K. Khazaaleh *et al.*, "Handling DNA malfunctions by unsupervised machine learning model," *Journal of Pathology Informatics*, vol. 14, p. 100340, 2023. https://doi.org/10.1016/j.jpi.2023.100340

[7] M. M. Al-Nawashi, O. M. Al-Hazaimeh, and M. K. Khazaaleh, "A new approach for breast cancer detection-based machine learning technique," *Applied Computer Science*, vol. 20, no. 1, pp. 1–16, 2024. https://doi.org/10.35784/acs-2024-01

[8] Z. Kanetaki, C. Stergiou, G. Bekas, C. Troussas, and C. Sgouropoulou, "A hybrid machine learning model for grade prediction in online engineering education," *International Journal of Engineering Pedagogy (iJEP)*, vol. 12, no. 3, pp. 4–24, 2022. https://doi.org/10.3991/ijep.v12i3.23873

[9] N. Gharaibeh, A. A. Abu-Ein, O. M. Al-hazaimeh, K. M. 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," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 19, no. 4, pp. 22–50, 2023. https://doi.org/10.3991/ijoe.v19i04.37677

[10] E. Zacharaki, S. Wang, Y. D. S. Chawla, R. Wolf, E. Melhem, and C. Davatzikos, "Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme," *Magnetic Resonance in Medicine*, vol. 62, no. 6, pp. 1609–1618, 2009. https://doi.org/10.1002/mrm.22147

[11] P. Afshar, K. Plataniotis, and A. Mohammadi, "Capsule networks for brain tumor classification based on MRI images and coarse tumor boundaries," in *2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 1368–1372. https://doi.org/10.1109/ICASSP.2019.8683759

[12] B. Ahmad, J. Sun, Q. You, V. Palade, and Z. Mao, "Brain tumor classification using a combination of variational autoencoders and generative adversarial networks," *Biomedicines*, vol. 10, no. 2, p. 223, 2022. https://doi.org/10.3390/biomedicines10020223

[13] I. Illán *et al.*, "Computer aided diagnosis of Alzheimer's disease using component based SVM. Applied soft computing," *Applied Soft Computing*, vol. 11, no. 2, pp. 2376–2382, 2011. https://doi.org/10.1016/j.asoc.2010.08.019

[14] W. A. Shehri, "Alzheimer's disease diagnosis and classification using deep learning techniques," *PeerJ Computer Science*, vol. 8, p. e1177, 2022. https://doi.org/10.7717/peerj-cs.1177

[15] Deepanshi, B. Ishan, and G. Deepak, "Alzheimer's disease classification using transfer learning," in *Advanced Computing, Communications in Computer and Information Science*, D. Garg, S. Jagannathan, A. Gupta, and L. Garg, Eds., Springer, Charm, vol. 1528, 2021, pp. 73–81. https://doi.org/10.1007/978-3-030-95502-1_6

[16] A. Tooba, A. Syed Muhammad, G. Nadia, M. Muhammad Nadeem, and M. Muhammad, "Multi-class Alzheimer's disease classification using image and clinical features," *Biomedical Signal Processing and Control*, vol. 43, pp. 64–74, 2018. https://doi.org/10.1016/j.bspc.2018.02.019

[17] M. Nuwan, C. Heung-Kook, S. Jae-Hong, and C. Boo-Kyeong, "Alzheimer's Disease classification based on multi-feature fusion," *Current Medical Imaging*, vol. 15, no. 2, pp. 161–169, 2019. https://doi.org/10.2174/1573405614666181012102626

[18] Z.-d. Iliass, R. Jamal, E. F. Khalid, M. Mohamed Adnane, and T. Hamid, "Brain tumor classification using machine and transfer learning," in *Proceedings of the 2nd International Conference on Big Data, Modelling and Machine Learning – BML*, SciTePress, 2022, pp. 566–571. https://doi.org/10.5220/0010762800003101

[19] F. J. Martinez-Murcia, A. Ortiz, J. M. Gorriz, J. Ramirez, and D. Castillo-Barnes, "Studying the manifold structure of Alzheimer's Disease: A deep learning approach using convolutional autoencoders," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 1, pp. 17–26, 2020. https://doi.org/10.1109/JBHI.2019.2914970

[20] Z.-d. Iliass *et al.*, "Brain tumor classification using feature extraction and ensemble learning," *Machine Graphics & Vision*, vol. 33, nos. 3/4, pp. 3–28, 2024. https://doi.org/10.22630/MGV.2024.33.3.1

[21] Z.-d. Iliass, R. Jamal, E. F. Khalid, M. Mohamed Adnane, and T. Hamid, "Alzheimer's disease classification using histogram of oriented gradient, transfer learning, and capsules network," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 4, pp. 5335–5350, 2024. https://ijisae.org/index.php/IJISAE/article/view/7325

[22] G. Li *et al.*, "Real-time classifcation of brain tumors in MRI images with a convolutional operator-based hidden markov model," *Journal of Real-Time Image Processing*, vol. 18, pp. 1207–1219, 2021. https://doi.org/10.1007/s11554-021-01072-4

[23] S. Dubey, "Alzheimer's dataset (4 class of images)," Kaggle, San Francisco, CA, USA, 2020. Accessed: Feb. 12, 2023. [Online]. https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-4-class-of-images

[24] N. Vinutha, S. Pattar, and S. Sharma, "A machine learning framework for assessment of cognitive and functional impairments in Alzheimer's Disease: Data Preprocessing and analysis," *The Journal of Prevention of Alzheimer's Disease*, vol. 7, no. 2, pp. 87–94, 2020. https://doi.org/10.14283/jpad.2020.7

[25] Y. Jianzhou, L. Stephen, K. Sing Bing, and T. Xiaoou, "Learning the change for automatic image cropping," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 971–978. https://doi.org/10.1109/CVPR.2013.130

[26] M. S. Hasan, M. S. Uddin, K. J. Hasan, M. Rahman, and M. Ali, "Deep learning for cultural heritage: A mobile app for monument recognition using convolutional neural networks," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 19, no. 3, pp. 22–40, 2025. https://doi.org/10.3991/ijim.v19i03.50935

[27] Y. Zhao, B. Ma, P. Jiang, D. Zeng, X. Wang, and S. Li, "Prediction of Alzheimer's Disease progression with multi-information generative adversarial network," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 3, pp. 711–719, 2021. https://doi.org/10.1109/JBHI.2020.3006925

[28] B. Christopher, G. Roger, H. Alexander, and R. Daniel, "Modelling the progression of Alzheimer's disease in MRI using generative adversarial networks," in *Medical Imaging 2018: Image Processing*, SPIE Medical Imaging, Houston, Texas, United States, 2018. https://doi.org/10.1117/12.2293256

[29] S. Hoo-Chang *et al.*, "GANDALF: Generative adversarial networks with discriminator-adaptive loss fine-tuning for Alzheimer's Disease diagnosis from MRI," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020*, A. L. Martel et al., Springer, Charm, vol. 12262, 2020, pp. 688–697. https://doi.org/10.1007/978-3-030-59713-9_66

[30] P. Jinhee, K. Hyerin, K. Jaekwang, and C. Mookyung, "A practical application of generative adversarial networks for RNA-seq analysis to predict the molecular progress of Alzheimer's disease," *PLoS Computational Biology*, vol. 16, no. 7, p. e1008099, 2020. https://doi.org/10.1371/journal.pcbi.1008099

[31] S. Deepak and P. M. Ameer, "Brain tumor classification using deep CNN features via transfer learning," *Computers in Biology and Medicine*, vol. 111, p. 103345, 2019. https://doi.org/10.1016/j.compbiomed.2019.103345

[32] H. Acharya, R. Mehta, and D. Kumar Singh, "Alzheimer Disease classification using transfer learning," in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, 2021, pp. 1503–1508. https://doi.org/10.1109/ICCMC51019.2021.9418294

[33] N. Deepa and S. P. Chokkalingam, "Optimization of VGG16 utilizing the arithmetic optimization algorithm for early detection of Alzheimer's disease," *Biomedical Signal Processing and Control*, vol. 74, p. 103455, 2022. https://doi.org/10.1016/j.bspc.2021.103455

[34] L. V. Fulton, D. Dolezel, J. Harrop, Y. Yan, and C. P. Fulton, "Classification of Alzheimer's Disease with and without imagery using gradient boosted machines and ResNet-50," *Brain Sciences*, vol. 9, no. 9, p. 212, 2019. https://doi.org/10.3390/brainsci9090212

[35] L. Yanjun, Y. Xiaowei, Z. Dajiang, and Z. Lu, "Classification of Alzheimer's Disease via vision transformer," in *Proceedings of the 15th International Conference on PErvasive Technologies Related to Assistive Environments*, 2022, pp. 463–468. https://doi.org/10.1145/3529190.3534754

[36] O. Modupe, M. Rytis, and D. Robertas, "Pixel-level fusion approach with vision transformer for early detection of Alzheimer's Disease," *Electronics*, vol. 12, no. 5, p. 1218, 2023. https://doi.org/10.3390/electronics12051218

[37] S. Hyunji, J. Soomin, S. Youngsoo, K. Sangjin, and K. Doyoung, "Vision transformer approach for classification of Alzheimer's disease using 18F-Florbetaben brain images," *Applied Sciences*, vol. 13, no. 6, p. 3453, 2023. https://doi.org/10.3390/app13063453

[38] C. Davide *et al.*, "Performance evaluation of an automated ELISA system for Alzheimer's disease detection in clinical routine," *Journal of Alzheimer's Disease*, vol. 54, no. 1, pp. 55–67, 2016. https://doi.org/10.3233/JAD-160298

[39] R. Prashanth, R. Sumantra Dutta, M. Pravat, and G. Shantanu, "High-accuracy detection of early parkinson's disease through multimodal features and machine learning," *International Journal of Medical Informatics*, vol. 90, pp. 13–21, 2016. https://doi.org/10.1016/j.ijmedinf.2016.03.001

[40] Z.-d. Iliass *et al.*, "A review: Machine learning techniques of Brain tumor classification and segmentation," *Machine Graphics & Vision*, Forthcoming.

7 AUTHORS

Iliass Zine-dine, PhD in Computer Science at the Laboratory of Informatics, Signals, Automation, and Cognitivism (LISAC), Faculty of Sciences Dhar El Mehraz (FSDM), Sidi Mohamed Ben Abdellah University (USMBA), Fes, Morocco (E-mail: iliass.zinedine@usmba.ac.ma).

Jamal Riffi, Professor in Computer Science at the Laboratory of Informatics, Signals, Automation, and Cognitivism (LISAC), Faculty of Science, Dhar El Mahraz (FSDM), Sidi Mohamed Ben Abdellah University (USMBA), Fes, Morocco.

Khalid El Fazazy, Professor in Computer Science at the Laboratory of Informatics, Signals, Automation, and Cognitivism (LISAC), Faculty of Science, Dhar El Mahraz (FSDM), Sidi Mohamed Ben Abdellah University (USMBA), Fes, Morocco.

Ismail El Batteoui, Professor in Computer Science at the Laboratory of Informatics, Signals, Automation, and Cognitivism (LISAC), Faculty of Science, Dhar El Mahraz (FSDM), Sidi Mohamed Ben Abdellah University (USMBA), Fes, Morocco.

Mohamed Adnane Mahraz, Professor in Computer Science at the Laboratory of Informatics, Signals, Automation, and Cognitivism (LISAC), Faculty of Science, Dhar El Mahraz (FSDM), Sidi Mohamed Ben Abdellah University (USMBA), Fes, Morocco.

Hamid Tairi, Professor in Computer Science at the Laboratory of Informatics, Signals, Automation, and Cognitivism (LISAC), Faculty of Science, Dhar El Mahraz (FSDM), Sidi Mohamed Ben Abdellah University (USMBA), Fes, Morocco.