



Investigating Deep Learning for Early Detection and Decision-Making in Alzheimer's Disease: A Comprehensive Review

Ghazala Hcini¹ · Imen Jdey¹ · Habib Dhahri²

Accepted: 17 March 2024 / Published online: 24 April 2024
© The Author(s) 2024

Abstract

Alzheimer's disease (AD) is a neurodegenerative disorder that affects millions of people worldwide, making early detection essential for effective intervention. This review paper provides a comprehensive analysis of the use of deep learning techniques, specifically convolutional neural networks (CNN) and vision transformers (ViT), for the classification of AD using brain imaging data. While previous reviews have covered similar topics, this paper offers a unique perspective by providing a detailed comparison of CNN and ViT for AD classification, highlighting the strengths and limitations of each approach. Additionally, the review presents an updated and thorough analysis of the most recent studies in the field, including the latest advancements in CNN and ViT architectures, training methods, and performance evaluation metrics. Furthermore, the paper discusses the ethical considerations and challenges associated with the use of deep learning models for AD classification, such as the need for interpretability and the potential for bias. By addressing these issues, this review aims to provide valuable insights for future research and clinical applications, ultimately advancing the field of AD classification using deep learning techniques.

Keywords Deep learning · Convolutional neural network · Vision transformer · Alzheimer's disease · Image classification

✉ Imen Jdey
imen.jdey@fstsbz.u-kairouan.tn

Ghazala Hcini
hcinighazala@fstsbz.u-kairouan.tn

Habib Dhahri
hdhahri@ksu.edu.sa

¹ ReGIM-Lab. Research Groups in Intelligent Machines (LR11ES48), University of Sfax, Sfax, Tunisia

² College of Applied Computer Sciences, King Saud University, Riyadh, Saudi Arabia

1 Introduction

Alzheimer's disease is a slow-developing neurodegenerative condition with symptoms including personality and behavior changes, memory problems, and dementia [1]. The disease's origins are unknown, and there is no reliable diagnosis or disease-modifying therapy. Treatment is expensive, burdening society and patients' families [2, 3].

The prevalence of AD and its impact on individuals and society are significant concerns. AD is the most common cause of dementia, affecting around 41 million people worldwide, and it is projected to increase to 152 million by 2050 due to the aging population and the lack of effective treatments [4]. AD can seriously impair a person's cognitive and functional abilities, leading to difficulties in performing daily activities and maintaining social relationships. The impact of AD on family caregivers, the dementia workforce, and society is also substantial. The disease takes a toll on caregivers, who often provide unpaid care for their loved ones with dementia and face physical, emotional, and financial burdens. The dementia workforce, including healthcare professionals, social workers, and other support providers, is also challenged by the increasing demand and complexity of Alzheimer's care. The costs associated with Alzheimer's care, including medical expenses and lost productivity, are significant and can have a ripple effect on society as a whole. According to the Alzheimer's Association, the total payments for healthcare, long-term care, and hospice for people with Alzheimer's and other dementias in the United States are estimated to be 355 billion in 2021, and this number could rise to 1.1 trillion by 2050 if medical breakthroughs are not developed to slow or cure the disease [5, 6].

Recent advances in AD research have focused on discovering new targets for drug candidates, identifying molecular subtypes of AD, and developing new biomarkers that may serve as the basis for precision medicine approaches to treatment development [7]. For example, researchers have identified ATP6VA1 and VGF as key regulatory genes in neural networks disrupted in AD and have shown that modulating their expression can improve neural function in cellular and animal models. Additionally, researchers used machine learning and network analysis to classify Alzheimer's patients into six molecular subtypes based on gene expression profiles, and these subtypes have distinct clinical and pathological features [8, 9]. Additionally, researchers have discovered new biomarkers that can detect early AD, including blood-based markers of amyloid-beta and tau proteins and retinal markers of neurodegeneration [10]. These discoveries have implications for improving the diagnosis, prognosis, and treatment of AD, as well as for designing more efficient clinical trials. However, many challenges and gaps still exist in AD research. For example, the lack of effective treatments to alter the course of AD, the limited understanding of environmental and lifestyle factors that influence the risk and progression of AD, and the need for a more diverse and representative patient population and a cohort of study participants.

Such machine learning (ML) algorithms need to be implemented with the correct architectural design and predefined preprocessing steps. Classification studies using ML typically involve four steps: feature extraction, feature selection, dimensionality reduction, and feature-based algorithm selection [11]. These procedures can be time-consuming and require specialized knowledge and multiple optimization stages. However, the reproducibility of these methods has been an issue. For example, in the feature selection procedure, AD-related features from different neuroimaging modalities are chosen to derive more informative combinatorial measures. These features may include mean subcortical volumes, gray matter densities, cortical thickness, brain glucose metabolism, and cerebral amyloid accumulation in regions of interest (ROIs), such as the hippocampus.

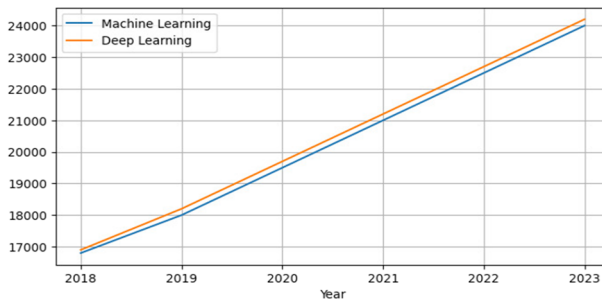


Fig. 1 The number of publications on AD diagnosis using ML and DL from 2018 to 2023. The figure shows the increasing interest and research activity in both fields, based on the results of a web search using the keywords “Alzheimer’s disease diagnosis”, “machine learning”, and “deep learning”. The figure also shows that DL publications are growing faster than ML publications, indicating that DL is a more novel and promising approach for AD diagnosis

The disease has significant impacts on public health, society, and the economy, including mortality, morbidity, and economic and social burdens [12]. Despite significant efforts, diagnosing AD remains a complex task, with a substantial percentage of undiagnosed cases globally. The current diagnostic challenges highlight the need for more accurate and early diagnosis methods. In this context, the role of DL has emerged as a promising approach for accurately and early diagnosing AD [13]. By leveraging deep neural networks, researchers have made strides in developing more effective diagnostic tools and categorizing the stages of the disease, offering hope for improved patient outcomes. Therefore, integrating DL techniques in Alzheimer’s research represents a significant step towards addressing the diagnostic challenges associated with the disease (Fig. 1) [14, 15].

This review comprehensively explores the use of deep learning models, particularly convolutional neural networks (CNN) and Vision Transformer (ViT), for classifying AD from medical images. It covers an introduction to the disease, the potential of deep learning in diagnosis, related works on CNN and ViT applications, and discusses combining both models’ strengths and limitations. Emphasis is placed on deep learning and image classification concepts relevant to AD diagnosis.

2 Background

Neurological diseases encompass a wide range of conditions that impact the nervous system, including the brain, spinal cord, and nerves, governing functions like movement, sensation, and cognition [16]. These diseases, such as Alzheimer’s disease, Parkinson’s disease, multiple sclerosis, epilepsy, and migraine, can arise from genetic, environmental, and lifestyle factors [17] (Fig. 2). Symptoms vary based on the specific condition and affected area of the nervous system, ranging from headaches to modifications in movement, sensation, and cognitive function. Diagnosis is intricate and may involve physical exams, medical history reviews, imaging tests, and specialized diagnostic procedures. Treatment options differ depending on the disease and can include medication, physical therapy, surgery, or other interventions [18].

Managing neurological diseases is crucial for improving quality of life; hence, early detection and treatment play a vital role in symptom management and better outcomes [19] (Fig. 3). Ongoing research and advancements in diagnostics and treatments are enhancing our

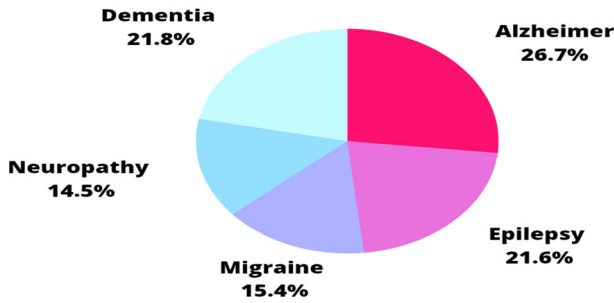


Fig. 2 Neurological disorders diagnosis using deep learning, Google scholar 2023

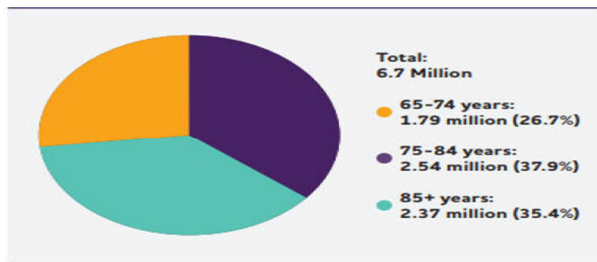


Fig. 3 AD prevalence and age distribution

understanding and management of these complex conditions, such as AD, which necessitate early and accurate diagnosis for optimal patient care [20].

2.1 Alzheimer's Disease Datasets

Several publicly available datasets of AD images exist, with the two most commonly used being the AD Neuroimaging Initiative (ADNI) and the Open Access Series of Imaging Studies (OASIS) (Tables 1, 2).

- Alzheimer's Disease Neuroimaging Initiative (ADNI)*: ADNI is a large-scale longitudinal multimodal study that aims to develop biomarkers to track the progression of Alzheimer's disease (AD) and evaluate potential therapeutics. The dataset includes imaging data from multiple sources, such as magnetic resonance imaging (MRI), positron emission tomography (PET), and other imaging modalities. ADNI offers both structural and functional MRI data, along with amyloid PET and fluoro-deoxyglucose PET (FDG-PET) scans. The dataset contains imaging data from both AD patients and healthy controls across multiple phases of the study, starting from ADNI1 in 2004 up through ADNI3. It is made freely available at <https://www.kaggle.com/datasets/madhucharan/alzheimersdisease5classdatasetadni>
- Open Access Series of Imaging Studies (OASIS)*: This dataset includes MRI scans of both healthy individuals and those with AD, as well as individuals with other neurological disorders. It is made freely available at <https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-4-class-of-images>

Table 1 Description about the Alzheimer's disease images: ADNI

Stage	Training images	Testing images
EMCI	204	36
LMCI	61	11
MCI	198	35
AD	145	26
CN	493	87

Table 2 Description about the Alzheimer's disease images: OASIS

Stage	Training images	Testing images
MildDemented	717	179
VeryMildDemented	1792	448
NonDemented	2560	640
ModerateDemeneted	52	12

2.1.1 Preprocessing Steps Applied to ADNI and OASIS Datasets

- *Intensity normalization* is a crucial step in image processing, especially for neuroimaging data and medical images with varying pixel intensities. It adjusts pixel intensity values to improve contrast and visual interpretation. Linear normalization transforms original grayscale images into desired intensities, ensuring consistency and enhancing image processing quality [21, 22].
- *Skull stripping* is a crucial process in Alzheimer's data images, eliminating non-brain signals for neuroimage processing tasks. Techniques include traditional and deep learning methods, with automated brain extraction methods used for consistency and precision. Skull stripping is essential for brain tissue segmentation and volumetric analysis [23].
- *Motion correction*: is a crucial stage in Alzheimer's data images, repositioning and orienting brain magnetic resonance images to account for motions during imaging. It is essential for neuroimage processing tasks like tissue segmentation and volumetric analysis. It removes motion, eddy-current, and echo-planar imaging artifacts from raw Diffusion-weighted imaging (DWI) images [24].
- *Spatial smoothing* in the context of Alzheimer's data images involves a filtering process based on segmented tissue images from previous steps. This technique often uses the Gaussian kernel function in normalized space to reduce noise and enhance the quality of the images. Spatial smoothing is commonly applied to various imaging data, including Amyloid PET scans, to decrease noise levels and improve the accuracy of clinical trial results.
- *Registration* is the process of aligning images into a common coordinate system, crucial for neuroimaging applications like AD diagnostic classification. It allows accurate analysis and interpretation of data, especially in studies using ML techniques to extract features for classification [25].

2.1.2 Image Classification

Medical images such as magnetic resonance imaging (MRI), computed tomography (CT), or positron emission tomography (PET) scans can be examined using image classification tasks

Fig. 4 MRI modality

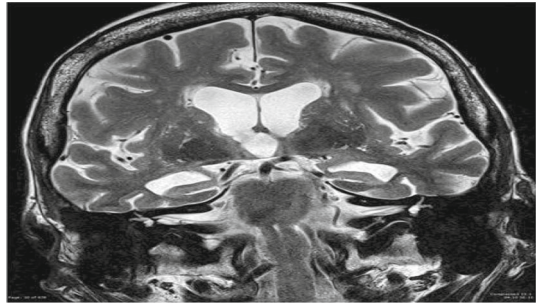
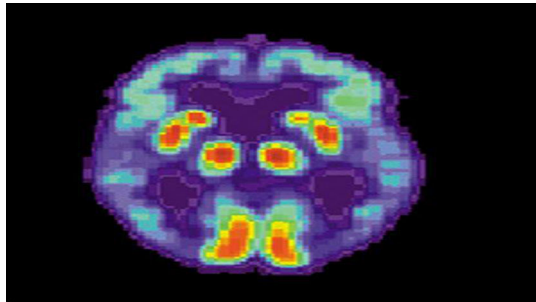


Fig. 5 PET modality



to assist with AD diagnosis. These images provide valuable information about the structure and functioning of the brain that can be used to identify AD-related alterations.

Magnetic Resonance Imaging: MRI modality

MRI is a medical imaging technique that uses magnetic fields and radio waves to produce detailed images of body structures, including the brain, spinal cord, joints, and abdomen [26]. It is widely used in AD research due to its excellent anatomical detail and high diagnostic accuracy, which are correlated with underlying pathology (Fig. 4) [27].

Positron Emission Tomography: PET modality

PET imaging uses radioactive material to create 3D body images, aiding in cancer diagnosis and treatment monitoring, as well as assessing brain function and cardiovascular health. Its advantages include detecting biochemical changes indicating diseases and monitoring treatment effectiveness such as chemotherapy (Fig. 5) [28, 29].

Image classification can provide insights into brain structure and function, revealing underlying pathology and disease progression. It aids early detection and diagnosis of AD, improving patient outcomes and enabling better treatments and disease management through the analysis of vast medical images.

2.2 Deep Learning

Deep learning is a branch of artificial intelligence that uses multi-layered neural networks to learn from large amounts of data and perform complex tasks [30–32]. CNN and ViT are two types of DL models [33] that can be used for image classification tasks [34, 35], such as Alzheimer's disease (AD) classification. CNN stands for Convolutional Neural Network, which is a neural network that uses convolutional layers to extract features from images. CNNs have been widely used for medical image analysis and have proven to be effective in AD classification tasks, as they can learn hierarchical patterns from brain imaging data, such as

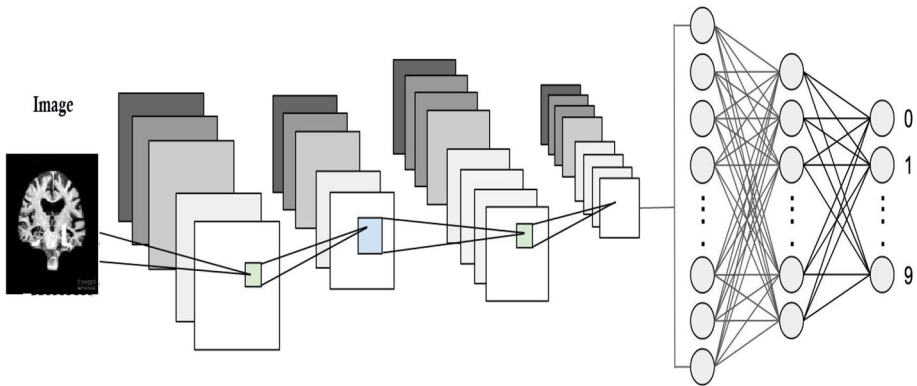


Fig. 6 Convolutional Neural Networks architecture

MRI and PET images. ViT stands for Vision Transformer, which is a neural network that uses transformer layers to encode patches of images as sequences of tokens. ViT models represent an input image as a series of image patches, similar to the series of word embeddings used in text, and directly predict class labels for the image. ViT models have shown exceptional performance in image recognition tasks when trained on enough data, and offer a promising alternative to traditional CNNs in handling the complexities of brain imaging data [36, 37]. Therefore, both CNN and ViT are well-suited for AD classification [38], depending on the specific problem and the available data [15].

2.2.1 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) is a type of DL algorithm that are commonly used for image classification, object detection, and image segmentation [39–41]. They are based on the concept of convolution, which involves sliding a small window, called a kernel or filter, over an input image to extract features [42].

Convolutional, pooling, and fully connected layers are the typical components of a CNN [43]. Convolutional layers train the filters to recognize particular aspects of the input image, such as edges or corners. The feature maps of the convolutional layers are downsampled by the pooling layers, resulting in smaller spatial datasets and a more computationally efficient network. The high-level features discovered by the convolutional layers are used to make predictions by the fully connected layers, which are used for classification (Fig. 6) [44, 45].

CNNs can have multiple convolutional, activation, and pooling layers to learn more complex features and improve classification accuracy. Additionally, we can pre-train CNNs on large datasets such as ImageNet to improve performance on smaller datasets with limited training data. The Table 3 below shows various pretrained models based on CNN architecture applied to the ImageNet dataset.

2.2.2 Vision Transformers

A recent development in DL architecture called Vision Transformers (ViT) has demonstrated outstanding performance in image classification challenges [46, 47]. The transformer model, initially created for Natural Language Processing (NLP) activities, serves as the foundation for the architecture of ViT [48]. A transformer model transforms the input sequence using a

Table 3 Pretrained CNN models applied on ImageNet validation dataset

Model	Size (MB)	Parameters	Depth	Time (ms) per step (CPU)	Time (ms) per step (GPU)
Xception	88	22.9M	81	109.4	8.1
VGG16	528	138.4M	16	69.5	4.2
VGG19	549	143.7M	19	84.8	4.4
ResNet50	98	25.6M	107	58.2	4.6
ResNet50V2	98	25.6M	103	45.6	4.4
ResNet101	171	44.7M	209	89.6	5.2
ResNet101V2	171	44.7M	205	72.7	5.4
ResNet152	232	60.4M	311	127.4	6.5
ResNet152V2	232	60.4M	307	107.5	6.6
InceptionV3	92	23.9M	189	42.2	6.9
InceptionResNetV2	215	55.9M	449	130.2	10.0
MobileNet	16	4.3M	55	22.6	3.4
MobileNetV2	14	3.5M	105	25.9	3.8
DenseNet121	33	8.1M	242	77.1	5.4
DenseNet169	57	14.3M	338	96.4	6.3
DenseNet201	80	20.2M	402	127.2	6.7
EfficientNetB0	29	5.3M	132	46.0	4.9
EfficientNetB1	31	7.9M	186	60.2	5.6
EfficientNetB2	36	9.2M	186	80.8	6.5
EfficientNetB3	48	12.3M	210	140.0	8.8
EfficientNetB4	75	19.5M	258	308.3	15.1
EfficientNetB5	118	30.6M	312	579.2	25.3
EfficientNetB6	166	43.3M	360	958.1	40.4
EfficientNetB7	256	66.7M	438	1578.9	61.6

Depth counts the number of layers with parameters from <https://keras.io/api/applications/>

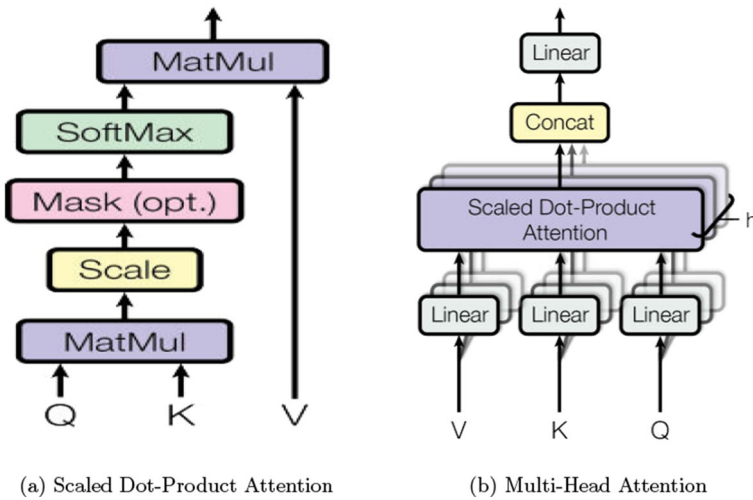


Fig. 7 Comparison of Scaled Dot-Product Attention and Multi-Head Attention

number of self-attention layers that encapsulate the interdependencies between the sequence’s various components. As a result, the model can learn representations of the input that are more suitable for tasks that come after them, like classification.

Attention in Transformer The Transformer model relies heavily on the attention mechanism, which comprises three distinct types of attention modules: self-attention, mask attention, and cross-sequence attention [49]. These modules are crucial for the ability of the Transformer to process and understand sequential data. Self-attention allows the model to focus on different parts of the input sequence, mask attention helps the model to ignore certain parts of the input sequence, and cross-sequence attention enables the model to establish relationships between different sequences [50].

Self-attention: In 2017, Ashish Vaswani et al. introduced the concept of the visual Transformer, which later became widely used in visual learning and understanding. The self-attention mechanism plays a crucial role in this approach, enabling the model to capture internal relationships among data or features without relying heavily on external information. By computing the interdependence between various patches of an image, self-attention is capable of resolving the issue of long-range dependencies.

To apply self-attention to an image X , we can use a particular model that involves transforming the input to obtain queries (Q), keys (K), and values (V). This transformation can be expressed mathematically using a standard formulation for Q , K , and V , as shown in Equation (1).

$$K = W^K X, Q = W^Q X, V = W^V X \tag{1}$$

Scaled Dot-Product Attention involves calculating the dot product of a query vector and a group of key vectors, scaling the output by the square root of the key vector dimension, and using a softmax function to compute a weighted sum of the value vectors. In contrast, Multi-Head Attention executes Scaled Dot-Product Attention several times using distinct learned parameters, enabling the model to focus on multiple aspects of the input concurrently (Fig. 7).

We may represent the scaled dot-product attention as Eq. (2):

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2}$$

where the $\left(\frac{QK^T}{\sqrt{d_k}}\right)$ is referred to as the attention matrix.

Multi-head attention: It can be expressed using equation 3.

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_h)W^O \tag{3}$$

where $head_i = Attention(W_i^Q X, W_i^K X, W_i^V X)$

The transformer design is modified for ViT to process and classify images. Smaller patches of the input image are first separated, and these patches are subsequently linearly projected into a series of embeddings. These embeddings are then sent via a number of transformer layers, allowing the model to pick up on the image’s general characteristics. The network is then completed by a straightforward classification head that takes the final transformer output and maps it to the output classes.

The popularity of transformers in NLP has spread to the CV research community, where there have been numerous attempts to modify transformers for use with vision-related problems. The most prominent of these models are detection transformer (DETR) [51], ViT [52], data-efficient image transformer (DeiT) [53], and Swin-Transformer [54]. Transformer-based models for vision have been developed at an unparalleled rate [55].

- *DETR*: proposed by Carion et al., is an end-to-end detection model that uses a trans-encoder to model image features, generate object queries, and assign labels to objects (Fig. 8).

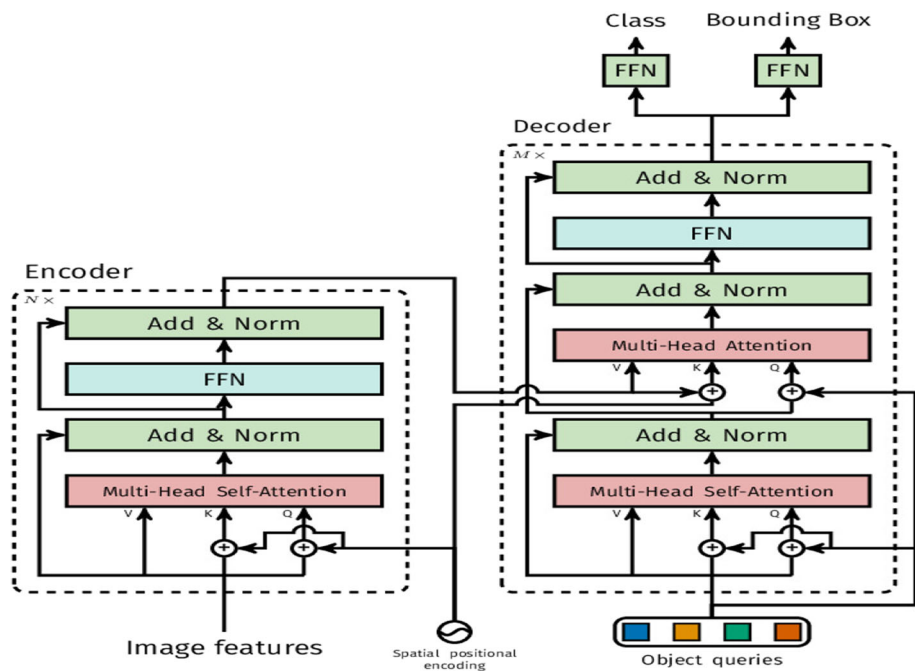


Fig. 8 DETR’s transformer Architecture [51]

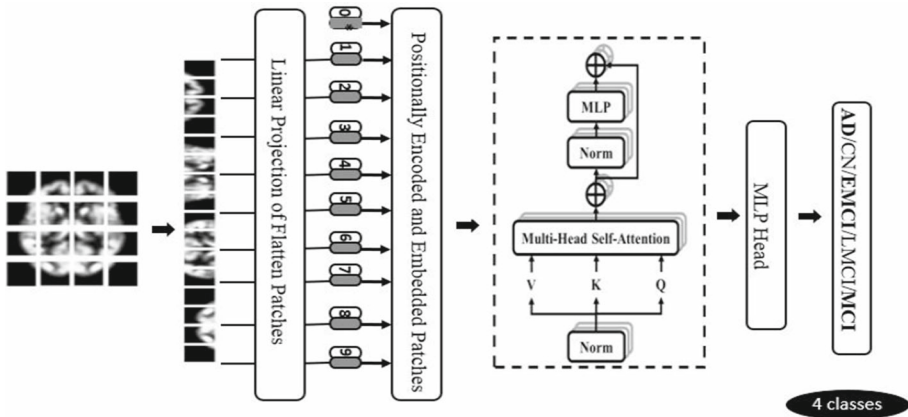


Fig. 9 ViT’s transformer Architecture

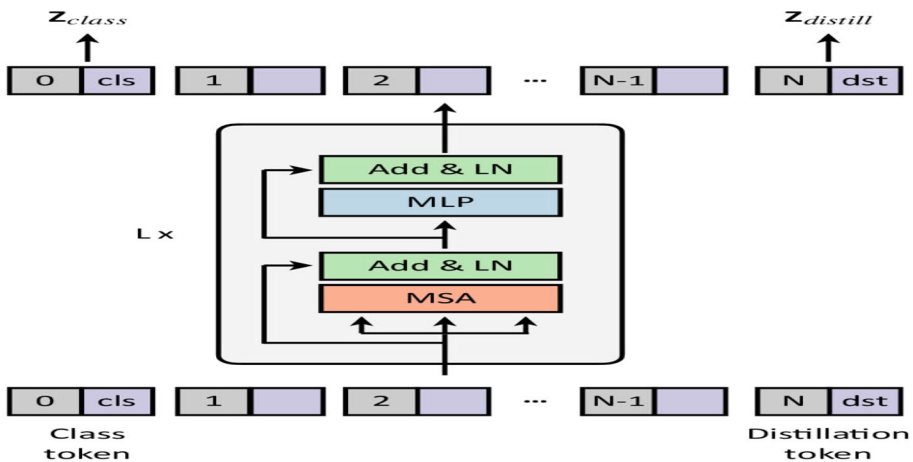


Fig. 10 DeiT’s transformer Architecture [56]

- *ViT*: Dosovitzky et al. proposed ViT, an image classification model that uses the Transformer architecture. It converts input images into patches, encodes spatial information, and outputs the learned patch embedding. The model classifies images using a multi-layer perceptron and can also incorporate feature maps from CNNs for relational mapping (Fig. 9).
- *DeiT*: Touvron et al. [53] proposed DeiT to address large training data requirements for ViT, using a knowledge distillation framework with Transformer terminology and CNNs for easier training and overcoming induced biases (Fig. 10).
- *Swin-Transformer*: To reduce the cost of computing attention for high-resolution images and to cope with varying patch sizes in scene understanding tasks (such as segmentation), Liu et al. proposed the Swin transformer. They introduced window self-attention to reduce computational complexity and used shifted window attention to model the relationship

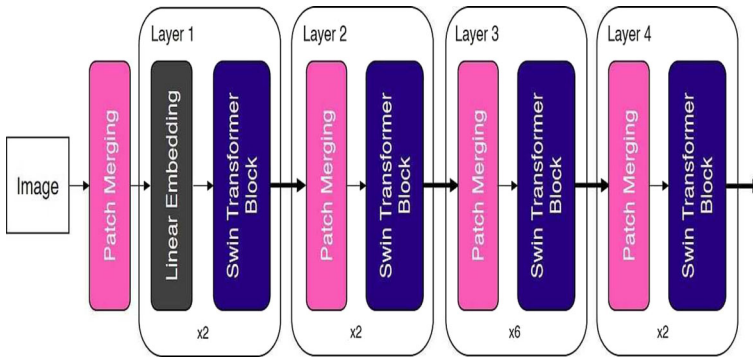


Fig. 11 Swin-transformer architecture

Table 4 CNN Vs Vision transformer

Aspect	CNNs	Vision transformers
Architecture	Consists of convolutional layers followed by fully connected layers	Uses self-attention mechanisms instead of convolutional layers
Feature Extraction	Extracts features from images using convolutional filters [58]	Extracts features via self-attention, attending to relevant parts of the input image
Application	Widely used for image classification, object detection, and segmentation	Initially developed for natural language processing but also performs well on image classification tasks
Performance	Demonstrates cutting-edge performance on benchmark datasets	Has shown encouraging results and outperformed CNNs on certain datasets [59]
Advantages	Efficient at capturing local patterns and edges	Can learn global context and relationships between distant image elements
Disadvantages	May struggle with capturing long-range dependencies and global context	Requires more computational resources due to self-attention mechanisms

between windows. In addition, they connected these attention blocks to patch merge blocks. This was done to merge adjacent patches and create a hierarchical representation to handle variations in visual unit scale (Fig. 11).

2.2.3 CNN Vs Vision Transformer

The Table 4 compares CNNs and VIT across key aspects such as architecture, feature extraction, application, performance, advantages, and disadvantages, highlighting their differences and respective strengths in handling image data. While CNNs excel at capturing local patterns, VIT leverage self-attention mechanisms to understand global context, offering unique advantages and trade-offs in image processing tasks [57].

3 Survey Methodology

The study investigated the detection of AD through extensive searches across renowned scientific databases, including Springer, IEEE Xplore, ScienceDirect, MDPI, Elsevier, Wiley, Taylor & Francis, and ACM libraries, using relevant keywords. Additionally, articles indexed

Table 6 Confusion matrix for image classification

	Actual positive	Actual negative
Predicted positive	True positive (TP)	False positive (FP)
Predicted negative	False negative (FN)	True negative (TN)

4.1 Metrics

Various evaluation metrics such as accuracy, precision, recall, F1 score, and AUC-ROC score are commonly employed to assess the performance of DL models in Alzheimer's image classification. These metrics provide insights into the model's ability to correctly identify AD from medical images, facilitating the optimization of algorithms for improved diagnostic accuracy. By using these evaluation metrics, researchers can ensure the reliability and effectiveness of their models in clinical settings (Tables 6, 7) [64].

4.2 Alzheimer's Disease Diagnosis Using CNN

Al Shehri [65] proposed a DL-based solution for the diagnosis and classification of AD using brain MRI images. The study used two CNN architectures, DenseNet-169 and ResNet-50, to classify the images into four stages of AD: Non-Dementia, Very Mild Dementia, Mild Dementia, and Moderate Dementia. The study reports that DenseNet-169 outperformed ResNet-50 in both training and testing accuracy, achieving 97.7% and 83.82% respectively. The study claims that the proposed model is usable for real-time analysis and classification of AD. The study contributes to the field of computer-aided diagnosis by applying DL techniques to a challenging and important problem.

Kaghan Oh et al. [66] employed a volumetric VCNN to extract features from the images and classify them into four categories: normal control (NC), progressive mild cognitive impairment (pMCI), stable mild cognitive impairment (sMCI), and Alzheimer's disease (AD). The study used a convolutional autoencoder (CAE)-based unsupervised learning for the AD vs. NC classification task, and supervised transfer learning for the other three tasks. The study also applied a gradient-based visualization method to identify the most important brain regions for the classification. The study reports that the proposed method achieved high accuracies of 86.60% and 73.95% for the AD and pMCI classification tasks respectively, outperforming other network models. The study also shows that the temporal and parietal lobes are the key regions for the diagnosis of AD and pMCI. The study contributes to the field of computer-aided diagnosis by providing a robust and interpretable deep learning solution for AD.

Rajendiran et al. [67] examined the effectiveness of different deep CNNs, including AlexNet, VGG-16 Net, ResNe-50, and Google Net models, for classifying the early stages of AD using the OASIS dataset. They compared their results with other existing methods, such as support vector machine (SVM), k-nearest neighbor (KNN), and random forest (RF). They reported that the Google Net model achieved the highest performance among all the methods with an accuracy value of 97.54%, a precision of 97.67%, a recall of 97.54% and a F1-score of 97.55%.

In terms of two fundamental phases, Sarang Sharma et al. [68] proposed a hybrid-based AI model that combines transfer learning (TL) and permutation-based ML voting classifiers. Two TL-based models, DenseNet-121 and DenseNet-201, are used in the first phase of

Table 7 Summary of evaluation metrics used in the relevant works in this review

Metric	Description	Formula
Accuracy	Is a fundamental metric used to evaluate classification models. It provides insight into how often a model's predictions are correct overall	Accuracy = $\frac{(TP+TN)}{(TP+TN+FP+FN)}$
Specificity	Is a statistical metric that measures the proportion of actual negative cases correctly identified by a ML model. In other words, it quantifies how well the model avoids false positives when classifying instances	Specificity = $\frac{(TN)}{(TN+FN)}$
Sensitivity /Recall	Measures how well a machine learning model can detect positive instances. It specifically focuses on the ability to identify true positive cases	Sensitivity = $\frac{TP}{(TP+FN)}$
Precision	Assesses the quality of positive predictions made by a model. It specifically focuses on the proportion of accurate positive forecasts among all instances predicted as positive	Precision = $\frac{TP}{(TP+FP)}$
F1-score	Combines precision and recall into a single score by calculating their harmonic mean	F1-score = $\frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$
Area Under the Curve (AUC)	It represents the entire two-dimensional area under the ROC curve, which captures the trade-off between true positive rate (sensitivity) and false positive rate (1-specificity) across different classification thresholds. A higher AUC value indicates better model performance	
Kappa	Kappa is an interesting metric for measuring classification performance	

implementation for feature extraction, and three different ML classifiers, SVM, Naive Bayes, and XGBoost, are used in the second phase for classification. The voting mechanism is permuted in order to evaluate the final classifier results. They performed the classification task of AD vs. NC and reported that their method achieved an accuracy of 91.75%, a specificity of 96.5%, and an F1-score of 90.25%.

Using 6400 labeled MRI scans from two classes, Tayyaba Shahwar et al. [69] proposed a hybrid classical quantum neural network to identify AD. In a quantum processor, traditional neural networks embed instructive feature vectors after extracting high-dimensional features. To produce a four-feature vector for precise decision boundaries, they fed a 512-feature vector into their quantum variational circuit (QVC) using ResNet34 to extract features from the image.

Using first- and second-order gradients, the Adam optimizer is used to take advantage of the adaptive learning rate corresponding to each parameter. Additionally, several quantum simulators (PennyLane, qiskit.aer, and qiskit.basicaer) are used for the detection of demented and non-demented images to validate the model. The learning rate is set at 10^4 for a six-layer optimized quantum depth.

Sheela et al. [70] created a model and used a Gaussian filter to remove unwanted noise, Otsu thresholding to segment the images, Prewitt edge detection to find the edges, GLCM to extract the features, FCM to cluster the images, and CNN to determine the final classification of the images. They reported that their model achieved an accuracy of 90.25% for the classification of AD vs. NC (normal control).

In order to classify the images into four categories-mild dementia (MD), moderate dementia (MOD), non-demented (ND), and very mild dementia (VMD)-Ghazal et al. [71] proposed a system for AD detection using transfer learning on multi-class classification using brain magnetic resonance imaging (MRI). They reported that their system achieved an accuracy of 91.70% for the classification of the four categories.

An artificial neural network with the combination of feed forward and back propagation was presented by Nancy et al. [72]. The output values are computed by the input layer and sent to the hidden layer. With the aid of its own weights and bias function, each hidden layer extracts signals from each input layer. The output layer receives these values in order to produce the output. The output is combined with the inputs and retrieved using an activation function. A threshold value is used to determine whether each input symbol is above or below the threshold value.

Using convolutional classifiers and Group Grey Wolf Optimization-based features, Shankar et al. [73] proposed a method for detecting AD.

Deep transfer learning models and MRI (Magnetic Resonance Imaging) images have been proposed by Sisodia et al. [74] to detect the various stages of AD, including "Very-Mild-Demented," "Mild-Demented," "Moderate-Demented," and "No-Demented." By applying data preprocessing and augmentation techniques, the model is able to identify the appropriate class of AD. The early stages of AD are then classified and predicted using additional deep transfer learning models (Resnet50, VGG19, Xception, DenseNet201, and EfficientNetB7).

Bhagat et al. [75] identified various stages of AD using pre-trained health data classification models like the MobileNet by using the transfer learning determinant.

A deep learning-based method that can predict MCI, early MCI (EMCI), late MCI (LMCI), and AD was proposed by Odusami et al. [76] For testing, the 138-subject Alzheimer's Disease Neuroimaging Initiative (ADNI) fMRI dataset was used.

Ebrahimi et al. [77] used a CNN called ResNet-18 that had been trained on an ImageNet dataset. The temporal convolutional network (TCN) and various recurrent neural network

types were the sequence-based models used. For AD detection, a number of deep sequence-based models and configurations were used.

In [78], they put forth AlzheimerNet, a refined convolutional neural network (CNN) classifier that can distinguish between the Normal Control (NC) category and all five stages of AD. The MRI scan dataset from the ADNI database is acquired for the proposed model's training and testing. They used the CLAHE image enhancement technique to get the raw data ready for analysis. The unbalanced nature of the dataset was corrected using data augmentation, and the resulting dataset contained 60,000 image data across the 6 classes.

In [79], they specifically used images that are segmented by the brain's gray matter (GM) to illustrate how transfer learning and convolutional neural network (CNN) customization can be used to segment and classify magnetic resonance imaging (MRI) of AD. They applied transfer learning after using a pre-trained DL model as their base model.

For the classification task, [80] used three pre-trained networks, GoogLeNet, AlexNet, and ResNet-18, which were trained and tested using 6000 images gathered from the ADNI database.

In [81], they employed a completely CNN architecture to detect AD. It uses brain MRI images that might be present in the ADNI dataset. The proposed process is primarily divided into two components. The first is feature extraction, the second is classification of Alzheimer disease into stages.

In [82], they suggested, using a dataset from magnetic resonance imaging (MRI) to diagnose neurodegenerative illnesses, an AlexNet-quantum transfer learning technique. By using a traditional pre-trained AlexNet model to extract a meaningful feature vector from high-dimensional data and then feeding this network to a quantum variational circuit (QVC), the hybrid model is created. The quantum computing phenomenon is used in quantum circuits. The 4096 features from the MRI dataset are extracted by the classical pre-trained model using AlexNet architecture, and this vector is then input to the quantum circuit. To solve the binary classification task for a brain illness, QVC generates a 4-dimensional vector and converts it into a 2-dimensional vector by connecting a fully connected layer at the end.

For the purpose of AD early detection, Balaji et al. [83] presented a hybrid Deep Learning Approach. Magnetic resonance imaging (MRI), positron emission tomography (PET), and conventional neuropsychological test results are combined in a strategy for early AD identification using multimodal imaging and convolutional neural network with the Long Short-Term Memory algorithm. The proposed methodology updates the learning weights, and accuracy is increased via Adam's optimization.

It was suggested to use a CNN to discriminate between patients with AD and those who had not yet received an AD diagnosis. Based on Matthew's Correlation Coefficient (MCC), the CNN's size, number of layers, and number of convolutional filters were adjusted and optimized [84].

The provided Fig. 13 presents a summary of the research conducted about AD diagnosis using CNN architectures, allowing for a comparison of the performance of various models.

4.3 Alzheimer's Disease Diagnosis Using ViT

In [85], they used resting-state functional magnetic resonance imaging (rs-fMRI) and structural magnetic resonance imaging (sMRI) data that had been aggressively preprocessed by their pipeline to separate healthy adults, those with mild cognitive impairment, and those with Alzheimer's brains within the same age group (> 75 years). They then introduced an optimized vision transformer architecture to predict the group membership.

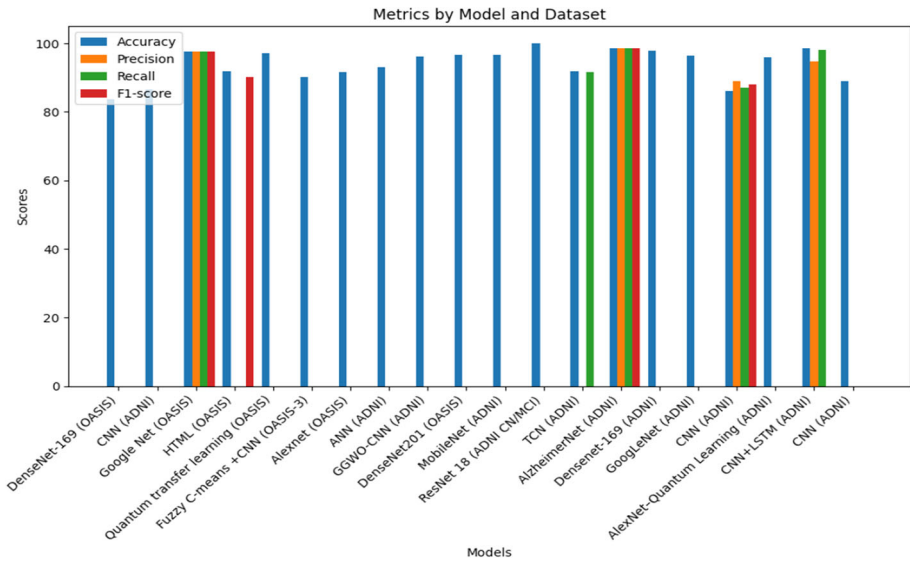


Fig. 13 Alzheimer’s disease diagnosis using CNN architectures

In [86], they set out to research an MCI-to-AD prediction approach that combines structural magnetic resonance imaging (sMRI) and Vision Transformers (ViT). It was possible to predict the development of MCI participants into AD using the 598 MCI subjects in the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database. Their study has three main goals: (i) to propose an MRI-based Vision Transformers approach for categorizing MCI-AD progression, (ii) to assess the effectiveness of various ViT architectures in order to select the best one, and (iii) to identify the brain region that has the greatest influence on the prediction of the deep learning approach to MCI progression.

The main contribution of the study lies in applying a ViT model to classify AD using 18F-Florbetaben (FBB) brain images [87]. The ViT model, originally proposed in the paper ‘Attention is All You Need,’ offers a novel approach to image classification by leveraging attention mechanisms. The architecture involves creating image patches, generating patch embeddings, adding class tokens, and positional embeddings, followed by linear projection for dimensionality reduction. By using ViT with FBB PET images, the study aims to enhance the accuracy of Alzheimer’s disease classification, potentially revolutionizing early detection and treatment strategies for this debilitating condition.

[88] proposed a ViT Approach for the detection of AD, which stands out for its remarkable efficacy in accurately diagnosing this mental disability. This innovative architecture involves converting input MRI images into a sequence of tokens suitable for the transformer model. The process includes tokenization, embedding each pixel into a token vector, incorporating positional encoding to capture spatial information, and integrating positional encoding with token embeddings to create the input sequence for the transformer. By leveraging this cutting-edge methodology, the study significantly advances the understanding and detection of AD through sophisticated artificial cognition techniques.

A summary Table 8 was created to compare the different approaches used on each dataset. Accuracy serves as the common performance metric across all models, facilitating a direct comparison of their effectiveness.

Table 8 Alzheimer's disease diagnosis using ViT

Contribution	Dataset	Performance [%]
OViTAD [85]	ADNI	F1-score for rs-fMRI modality=97 F1-score for rs-fMRI modality=99.55
Vision Transformers [86]	ADNI	1. ViT-S Accuracy=83.27 Sensitivity=85.07 Specificity=81.48 2. ViT-B Accuracy=80.67 Sensitivity=79.10 Specificity=82.22 3. ViT-L Accuracy=72.86 Sensitivity=74.63 Specificity=71.11
Vision Transformer [87]	ADNI	1. ViT with original data Accuracy=80 Recall=60 2. ViT with augmented data Accuracy=70 Recall=100
Vision Transformer[88]	OASIS	Accuracy=99.06 Precision=99.06 Recall=99.14 F1-score=99.1

4.4 Alzheimer's Disease Diagnosis Using Hybrid Models

The main contribution of the Dual-3DM-AD model is its innovative approach to the accurate and early diagnosis of AD by integrating MRI and PET image scans [89]. The model enhances image quality through preprocessing steps such as noise reduction and skull stripping, followed by 3D image conversion. It employs a Mixed-transformer with Furthered U-Net architecture for semantic segmentation, reducing complexity. Additionally, a multi-scale feature extraction module extracts relevant features from segmented images, which are aggregated using a densely connected feature aggregator module to leverage information from both modalities. A multi-head attention mechanism aids in feature dimensionality reduction, leading to multi-class Alzheimer's diagnosis through a softmax layer. These architectural components collectively enhance the accuracy and effectiveness of the model in AD diagnosis.

The significant contribution of Mahim et al.'s [90] research lies in the creation of a novel hybrid deep learning model. This model combines a Vision Transformer (ViT) with a Gated Recurrent Unit (GRU) to enhance the detection and classification of Alzheimer's Disease (AD) from brain MRI images. The ViT-GRU architecture effectively addresses challenges related to class imbalances and computational limitations in clinical settings. By capturing spatial and temporal information from diverse datasets, the model demonstrates improved

generalizability across different populations and imaging protocols, thereby expanding its utility in AD diagnosis. Additionally, the incorporation of eXplainable Artificial Intelligence (XAI) techniques, such as LIME, SHAP, and Attention maps, enhances the interpretability of the model's predictions, shedding light on the features influencing AD classification decisions.

On the basis of combining the Wide Residual Squeeze-and-Excitation Network with the Cross ViT, [91] they proposed a novel approach. The Alzheimer's Disease Neuroimaging Initiative (ADNI) and the Open Access Series of Imaging Studies (OASIS) provided them with MRI data. They also suggested a brand-new data augmentation method based on a self attention progressive generative adversarial neural network to get beyond the data's limitations.

A network model (WS-AMN) that combines weak supervision and an attention mechanism was put out by [92]. The basic model is the weakly supervised data augmentation network, the data augmentation is guided by the attention map produced by weakly supervised learning, and the residual network contains an attention module with a channel domain and a spatial domain to concentrate on the distinct channels and spaces of images, respectively. The influence of irrelevant features is reduced and the corresponding features of connected features are enhanced by the location information.

In [93], they proposed a 3D medical image classifier using multiplane and multislice transformer (M3T) networks to classify Alzheimer's disease (AD) in 3D MRI images. The proposed network synergistically combines 3D-CNN, 2D-CNN and transformers to achieve accurate AD classification. 3D-CNN is used to perform native 3D representation learning, and 2D-CNN is used to leverage pre-trained weights on large 2D databases and 2D representation learning. Using guided-biased CNN networks, it is possible to efficiently extract local information of AD-related abnormalities in the regional brain. Transformer networks are also used to obtain attentional relationships between multiplane (axial, coronal, sagittal) and multislice images according to CNN. Also, by using a transformer without inductive bias, we can learn abnormalities that spread over a wider area in the brain.

In [94] they proposed two improved ResNet algorithms that introduced the Contextual Transformer (CoT) module, group convolution, and channel shuffling mechanisms into the traditional ResNet residual block. They used the CoT module to replace the 3×3 convolution of the residue block and improve the feature extraction power of the residue block. Reorganize and improve feature maps for different groups of input layers using a channel shuffling mechanism. Communication between trait cards in different groups. Selected and edited images of his 503 subjects, including 116 healthy subjects (HC), 187 subjects with mild cognitive impairment (MCI), and 200 subjects with her AD from the ADNI database, preprocessed and sliced the data. 10,060 slices were then acquired and three groups, AD, MCI, and HC, were classified using an improved algorithm.

In [95] they used an established pipeline that included registration, slicing, and classification steps. Her research contributions include her three currently promising deep convolutional models (ResNet, DenseNet, and EfficientNet) and her two transform-based architectures (MAE and DeiT) for mapping input images to clinical diagnoses. to the best of her knowledge. Using multiple benchmarks obtained by varying the number of slices per subject extracted from the available 3D voxels, her two published data were used to allow fair comparison. Experiments were performed on the set (ADNI and OASIS). Experiments have shown that very deep ResNet and DenseNet models outperform shallow ResNet and VGG versions tested in the literature. We also found that the Transformer architecture, especially her DeiT, provided the best classification results and was more robust to noise added by increasing the number of slices.

In order to classify NDs, [96] created ViT models using the AD Neuroimaging Initiative (ADNI) dataset. Additionally, they used attention maps and the Automated Anatomical Labeling (AAL) brain atlas to create models that can be explained.

In [97] they proposed a new method for Alzheimer's disease image classification based on the mechanism of external attention. An external attention module is added after his fourth convolutional block in the fully convolutional network model. At the same time, a double normalization method of Softmax and L1 norm is introduced to improve the classification performance and yield richer feature information for disease probability maps. A softmax activation function can increase the fitness of a neural network to a training set. This transforms linearity into nonlinearity, making neural networks more flexible. The L1 norm can avoid the attention map from being affected by particularly large (and particularly small) eigenvalues. Experiments in their study used 550 3D MRI images and used 5-fold cross-validation.

In [98] they proposed a densely-connected CNN with per-connection attention mechanisms to learn multi-level features of brain MR images for AD classification. They used densely connected neural networks to extract multi-scale features from preprocessed images, applied a connection-oriented attention mechanism to combine connections between features in different layers, and they hierarchically transformed the MR images into more compact, high-level features. Furthermore, they extended the convolution operation to 3D to obtain spatial information for MRI. Features extracted from each 3D convolutional layer were merged with features from all previous layers with different levels of attention and finally used for classification. Their method was evaluated using baseline MRI of 968 subjects from the ADNI database, including (1) AD and healthy subjects, (2) MCI converters and healthy subjects, (3) MCI converters. and nonconverters.

A novel diagnostic model for AD based on multiview slice attention and 3D convolutional neural networks (3D-CNN) was presented [99]. The approach involved using multiple sub-networks to extract local slice-level features in various dimensions. A slice-level attention mechanism was proposed to highlight specific 2D slices and eliminate redundant features. To capture global structural changes at the subject level, 3D CNN was employed. All 2D and 3D features were then merged to create a more sophisticated representation. The model will be tested on 1,451 subjects from the ADNI-1 and ADNI-2 datasets.

In [100] they proposed a novel attention-based 3D multiscale CNN model (AMSNet) that can better capture and integrate multiple spatial features of AD with a concise structure. AMSNet is often generalized to other AD-related classification tasks such as the three-way classification (AD-MCI-CN). Their results demonstrate the feasibility and efficiency of the proposed multi-scale spatial feature integration and attention mechanism used in AMSNet for AD classification.

Wang et al. [101] first explore the usage of attention mechanisms and suggest a VGG-inspired network (VIN) as the backbone network. The convolutional block attention modules were integrated on a VIN backbone to form the Alzheimer's Disease VGG-Inspired Attention Network (ADVIAN) that was their suggested design. In order to prevent overfitting, 18-way data augmentation is also suggested. To report the unbiased performance, ten runs of 10-fold cross-validation are completed.

In their research [102], they investigate an attention-based convolutional neural network (CNN) model for Alzheimer's disease diagnosis and classification. The information of brain MRI images was fused from several levels to determine the correlation between the slices in their study, which introduced the attention mechanisms from the regional level and the feature level. The network model was then made more sensitive to the salient channel aspects

while reducing the impact of some noise features by building a spatio-temporal graph CNN with dual attention mechanisms.

An ADGNET-based weak supervision deep learning system was used in this study [103] to classify Alzheimer's disease. It is made up of a backbone network with an attention mechanism and a task network for concurrent image classification and image reconstruction in order to recognize and classify Alzheimer's disease with few annotated images. The ResNet-50 architecture, which has been modified to add an attention mechanism, serves as the foundation for the backbone network. One branch of the task network is for image classification, while the other is for image reconstruction.

Liu [104] developed of a new method called MPC-STANet for Alzheimer's Disease recognition. This method is based on multiple phantom convolution and spatial transformation attention mechanism. It is founded on a spatial transformation attention mechanism and multiple phantom convolution. MPC-STANet's network design is an improved version of ResNet50, a deep neural network architecture. The process involves employing several phantom convolutions to extract features from brain images, and then applying a spatial transformation attention mechanism to highlight the elements that are most important for identifying Alzheimer's disease. This method has demonstrated promising outcomes in precisely diagnosing Alzheimer's disease.

Wang et al. [105] created a deep learning method called IGnet to automatically classify AD cases using genetic sequencing and magnetic resonance imaging (MRI) data. With a deep three-dimensional convolutional network (3D CNN) being used to handle the three-dimensional MRI input and a Transformer encoder being used to handle the genetic sequence input, the suggested method combines computer vision (CV) and natural language processing (NLP) approaches. The Alzheimer's Disease Neuroimaging Initiative (ADNI) data set has been used to test the proposed approach (Table 9).

5 Discussion

This section focuses on the significance of early diagnosis in the treatment and management of Alzheimer's disease. Notably, image classification has seen remarkable advancements through deep learning algorithms, such as convolutional neural networks (CNNs) and Vision Transformers (ViTs). By combining these algorithms, we can improve the accuracy of classification and gain a deeper understanding of the visual data linked to Alzheimer's.

5.1 Strengths of CNNs for Alzheimer's Image Classification

CNNs have emerged as powerful tools for classifying AD using medical imaging data. They possess the ability to automatically extract relevant features from raw images, are robust to spatial transformations, and use parameters efficiently, making them well-suited for this task. Furthermore, by employing transfer learning with pretrained CNNs, performance can be improved even when working with smaller datasets. Deeper layers within CNNs capture hierarchical representations, aiding in the identification of complex brain abnormalities that are associated with AD.

Table 9 Alzheimer's disease diagnosis using hybrid models

Contribution	Dataset	Performance [%]
Dual-3DM-AD [89]	ADNI	Accuracy=98 Sensitivity=97.8 Specificity =97.5 f-measure=98.2
ViT-GRU [90]	ADNI	Accuracy 99.53 for 4-class 99.69 for binary classification 99.26 for 3-class
Cross ViT and Wide Residual Squeeze-and-Excitation Network [91]	ADNI-OASIS	Accuracy=99
WS-AMN [92]	OASIS	Accuracy=99.61 F1-score=99.63
M3T [93]	ADNI-OASIS	ADNI Accuracy=93.21 AUC=96.34 OASIS Accuracy=85.26 AUC=89.61
ViT with transfer learning [94]	ADNI	Accuracy=82
CCS-ResNet-50 [95]	ADNI	Accuracy=88.61
External-attention mechanism [96]	ADNI	Accuracy (MLP-A)=92.36 Accuracy(MLP-B)=98.55 Accuracy(MLP-C)= 98.73
1. DeiT (4 slices) 2. DeiT (8 slices) [97]	ADNI-OASIS	ADNI-2 1. Accuracy= 77 2. Accuracy=76 OASIS-1 1. Accuracy=74.4 2. Accuracy= 75.6
CAM-CNN [98]	ADNI	Accuracy=97.35
MSA3D [99]	ADNI	Accuracy=91.1 Sensitivity=88.8 Specificity=91.4 AUC=95
AMSNet [100]	–	Accuracy=91.3 Sensitivity=88.3 Specificity=94.2

Table 9 continued

Contribution	Dataset	Performance [%]
ADVIAN [101]	ADNI	Accuracy=97.76 ± 1.13 Sensitivity=97.65 ± 1.36 Specificity=97.86 ± 1.55 Precision=97.87 ± 1.53 F1-score= 97.75 ± 1.13 MCC=95.53 ± 2.27 FMI=97.76 ± 1.13 AUC= 0.9852
CNN with Attention Mechanisms [102]	ADNI	Accuracy AD/NC=95.26 AD/NC=87.29 NC/MCI=82.42 AD/NC/MCI=85.24
ADGNET [103]	KACD: 6400 2D MRI images ROAD: 416 3D MRI images	KACD – Kappa=99.22 – Sensitivity=99.69 – Specificity=99.53 – Precision=99.53 – Accuracy=99.61 – F1-score=99.61 ROAD – Kappa=97.36 – Sensitivity=98 – Specificity=99.24 – Precision=98.99 – Accuracy=98.71 – F1-score=98.49
MPC-STANet (1.69 M parameters) [104]	-	Accuracy=96.25 F1 score=95 mAP=93
IGnet [105]	ADNI	Accuracy=83.78 AUC-ROC=92.4

Bold indicate the highest accuracy achieved among the models presented

5.1.1 CNN Configuration Hyperparameters

When designing a CNN for Alzheimer's image classification, it's essential to consider several hyperparameters. These choices are influenced by the inherent strengths of CNNs, which we discussed earlier. Let's explore how these strengths guide our configuration decisions (Table 10):

The Fig. 14 presents a summary of hyperparameter configurations used in CNN architectures for Alzheimer's disease image classification. It includes information such as the number

Table 10 CNN configuration hyperparameters

Hyperparameter	Description
Number of layers	Affects network's depth; deeper networks can capture more complex features but may lead to overfitting
Size of the filter	Determines the receptive field; smaller filters capture more detail, larger filters capture more global features
Stride size	Determines how much the filter is shifted over the input image during convolution
Pooling method	Used to downsample feature maps; types include max pooling, average pooling, etc
Learning rate	Controls how fast the network updates its weights during training
Batch size	Determines number of samples used in each training iteration
Dropout rate	A regularization technique to prevent overfitting by randomly omitting nodes during training
Activation function	Determines the output of each neuron in the network; affects non-linear relationships between features
Weight initialization	Initial values assigned to the network's weights; affects convergence and accuracy
Data augmentation	Techniques used to artificially increase the size of the training data set and improve generalization

of layers, pooling size, learning rate, batch size, dropout rate, optimizer, weight decay, activation function, epoch, and data splitting. These hyperparameters play a crucial role in the performance and optimization of CNN models for AD classification.

5.2 Strengths of ViT for Alzheimer's Image Classification

ViTs are a revolutionary approach to global context modeling, enhancing feature representation and providing a holistic perspective on pathological changes. They significantly reduce GPU memory usage, making them ideal for Alzheimer's disease diagnosis, outperforming traditional CNNs.

5.2.1 ViT Configuration Hyperparameters

There are several parameters or variables that affect the performance of the ViT model for AD image classification (Table 11). Here are some of the most important:

The Fig. 15 illustrates a summary of hyperparameter configurations used in ViT architectures for Alzheimer's disease image classification. These hyperparameters are pivotal in optimizing the performance of ViT models for Alzheimer's disease classification tasks.

5.3 Benefits of CNNs and Vision Transformer Models fusion for Alzheimer's Image Classification

The fusion of CNNs and ViT models in image classification tasks, particularly in Alzheimer's image classification, holds significant promise. This fusion approach combines the strengths of both architectures, leveraging the feature extraction capabilities of CNNs and the self-

Table 11 ViT model parameters

Parameter	Description
Number of layers	It significantly impacts its performance, as deeper networks can capture more complex features but may also lead to overfitting
Patch size	It determines the granularity of image patches; larger sizes enhance computational efficiency but may compromise detail capture by the network
Number of patches	It dictates the quantity of tokens used as input; while adding patches enhances network detail capture, it can also escalate computational expenses
Embedding size	It determines the dimensionality of the feature vectors generated for each patch; larger sizes enable the network to capture more intricate functions, potentially demanding increased storage and processing power
Transformer architecture	It influences performance, with variations in attention mechanisms, normalization techniques, or feedforward layers impacting the network's capacity to capture complex feature relationships
Dropout rate	Dropout is a regularization technique that can be used to prevent overfitting. The optimal dropout rate will depend on the specific dataset and network architecture
Learning rate	It controls the pace at which the network adjusts its weights during training; high rates may lead to premature convergence and lower accuracy, while low rates can prolong training duration
Batch size	It determines the quantity of samples processed in each training iteration; while larger sizes may expedite convergence, they could also necessitate more memory and processing power
Weight initialization	The weights of the network's initial configuration can have an effect on how well it performs. An accurate weight initialization can hasten network convergence and improve accuracy
Data augmentation	By artificially expanding the size of the training dataset, data augmentation methods (including flipping, rotating, and scaling) can enhance the network's capacity to generalize to new images

5.4 DL Challenges in Alzheimer's Disease Imaging

When delving into DL challenges specific to Alzheimer's disease imaging, it is imperative to address the complexities of merging CNNs and ViT models. The fusion of these architectures presents significant hurdles, notably heightened computational complexity due to their distinct computational demands. While CNNs are renowned for their spatial hierarchy in image processing, ViT models excel in capturing long-range dependencies through self-attention mechanisms. This integration can lead to a more computationally intensive model, necessitating additional resources for training and inference.

Moreover, the intricate process of harmonizing features extracted by CNNs with tokenized embeddings of ViT models poses a notable challenge in achieving seamless information flow while maintaining model efficiency and interpretability. Striking a balance between the strengths of both architectures without introducing redundancies requires meticulous design and optimization.

Addressing these challenges effectively involves optimizing the fusion process, exploring efficient methods to combine CNN and ViT features, and potentially leveraging techniques

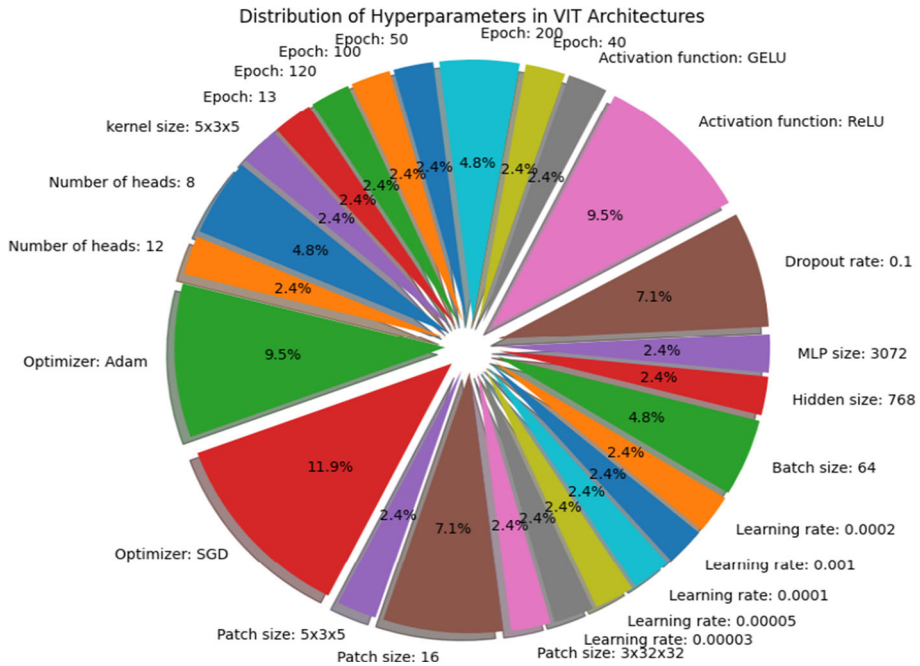


Fig. 15 Hyperparameter configurations in ViT architectures used for AD image classification

like knowledge distillation or architecture pruning to alleviate computational overhead. By acknowledging and actively working to overcome these limitations, researchers can advance the field of hybrid CNN-ViT models towards more practical and scalable solutions for complex image analysis tasks [107, 108]

1. *Limited data*: Many issues, such as the complexity and variability of AD, the difficulty and expense of obtaining and curating high-quality medical photographs, and strict privacy and confidentiality requirements, can be blamed for the lack of databases for images related to the disease [109]. ML models can be severely limited in their efficacy and application, which can result in overfitting and subpar performance on fresh data [110]. Collaboration between researchers, physicians, and data-sharing platforms is crucial to addressing these issues because it facilitates the acquisition, annotation, and exchange of high-quality imaging data related to Alzheimer’s disease. This may aid in the creation of stronger, more realistic models that more fully represent the intricacy of the illness. Potential solutions for addressing the challenges in Alzheimer’s disease imaging data include:

- *Transfer learning*: Employing well-trained models from similar tasks and fine-tuning the ending of two layers or even one layer based on the limited original data [111].
- *Data augmentation* is the process of adding more images to the training set by mirroring, rotating, and translating them. However, in some situations, such as with bioinformatics data, caution must be exercised while using this technique [112–114].
- *Simulated data*: If the problem is well understood, building simulators based on the physical process can help expand the amount of simulated data in the training set [112].

Table 12 Benefits of CNNs and vision transformer models fusion

Strengths	References
CrossViT combines local feature extraction (from CNNs) with global context modeling (from ViTs)	[91]
The Wide ResSE network enhances feature representation, prevents vanishing gradients, and dynamically weighs feature channels	
Dual-3DM3-AD combines MRI and PET scans, leveraging their combined information for accurate early Alzheimer's diagnosis	[89]
The innovative architecture includes mixed transformers, multi-scale feature extraction, and a densely connected feature aggregator module (DCFAM), achieving impressive performance in multi-class Alzheimer's diagnosis	
The study introduces a self-attention mechanism and partial convolution layer to extract global and local features from input information, enhancing feature extraction ability for Alzheimer's disease classification	[94]
The Contextual Transformer (CoT) module replaces the 3x3 convolution in residual blocks, improving feature extraction capability, while the Channel Shuffle mechanism enhances communication between feature maps	
The study introduces a self-attention mechanism and partial convolution layer to extract global and local features from brain MRI images, improving diagnosis accuracy	[102]
- Attention mechanisms are introduced at regional and feature levels, enhancing information fusion. - A spatio-temporal graph CNN with dual attention mechanisms is constructed, focusing on salient features while reducing noise impact	

2. *Data pre-processing*: is a crucial step in the development of deep learning models for accurate and efficient analysis of brain imaging data [115]. Proper architectural design and pre-processing steps are essential to ensure that the data is in a format that the model can accept and process effectively. Some of the key steps and considerations in data pre-processing for DL models include:
 - *Data cleaning*: This involves removing inconsistencies, missing values, and noise from the data to ensure data quality and consistency.
 - *Data transformation*: This step involves converting the data into a suitable format for the model, such as resizing images or normalizing data.
 - *Data reduction*: This step involves dimensionality reduction techniques to reduce the number of features while preserving the essential characteristics of the data, which can help improve the performance of deep learning models.
 - *Data normalization*: This step involves standardizing the data by subtracting the mean and dividing by the standard deviation, which can help improve the convergence of the model and reduce overfitting.
3. *Class imbalance*: Class imbalance in Alzheimer's disease datasets can pose a challenge for ML models, leading to difficulties in differentiating between classes and biased predictions [116]. To address this issue, various techniques can be employed to rebalance the

dataset and enhance the model's ability to learn from all classes. Some of these techniques include:

- *Oversampling and undersampling*: These methods can be applied to modify and balance the distribution of images among various classes. Undersampling entails choosing fewer examples of overrepresented classes, whereas oversampling entails making more copies of underrepresented classes.
 - *Weighted cross-entropy loss*: Use the weighted cross-entropy loss function to make sure the model works effectively with small class sizes. By giving the loss from underrepresented groups greater weights, it pushes the model to find more accurate representations for those classes [117].
 - *Building models for hierarchical levels*: Since a biological system often contains a hierarchical label space, this method can be used to build models for any hierarchical level. This may assist in resolving the problem of unbalanced data in DL models.
4. *Variability in images*: Quality, resolution, and image modality of images of Alzheimer's disease can vary (e.g., MRI, PET). It may be difficult for the model to learn traits that are resistant to these alterations as a result.
- [27]: This article addresses the difficulties in differentiating between asymptomatic and non-Alzheimer's moderate cognitive impairment (MCI) groups and emphasizes the use of brain imaging modalities in the diagnosis of Alzheimer's disease.
 - [118]: The necessity of medical image processing and analysis techniques to support Alzheimer's disease diagnosis is emphasized by this study. It also covers how to extract features from images using ML or DL approaches, which can enhance diagnosis accuracy.
 - [119]: The application of multimodal DL models for Alzheimer's disease early detection is the main emphasis of this work. It emphasizes how crucial it is to combine several imaging modalities-like MRI, PET, and genetic data-in order to provide a more thorough assessment of the illness.
5. *Interpretability*: DL models are frequently referred to as "black boxes," which makes it challenging to comprehend how the model generates its predictions. In medical applications where interpretability is crucial, this can be an issue. In the context of Alzheimer's disease, the interpretability of deep learning models poses significant challenges, particularly in understanding how these models generate predictions, which is crucial for medical applications. However, researchers have been exploring potential solutions to address this issue. Some relevant references include:
- [120]:The potential for interpretable neural network models to enhance patient outcomes is highlighted by this study, which introduces one that can assist in the noninvasive and cost-effective diagnosis of Alzheimer's disease.
 - [121]:The study addresses the current state of interpretable ML for dementia, including Alzheimer's disease, and highlights the necessity of involving doctors in future research to validate explanation techniques and draw firm conclusions regarding the pathology of dementia-related diseases.
 - [122]: The work highlights the construction of interpretable models suited to neuropathology, which is vital for understanding the disease. It reports on a DL pipeline that identifies unique neuropathologies in Alzheimer's disease.

6. *Ethical considerations*: The classification of Alzheimer's images raises ethical questions about patient privacy and potential technological abuse [123]. When creating and implementing these models, it is crucial to take these concerns into account [124, 125]. Researchers have been exploring potential solutions to address these ethical challenges. Some relevant references include:
- [126]: The ethical issues of openness, fairness, privacy, and data protection that arise from the use of AI in speech-based Alzheimer's detection are covered in this article. The writers also offer potential solutions for resolving these issues.
 - [127]: The ethical ramifications of using DL models to diagnose Alzheimer's disease in people who don't exhibit any symptoms are covered in this paper. The authors stress the significance of data protection, privacy, and informed consent in the creation and application of these models.
 - [128]: This essay addresses the moral dilemmas raised by the application of AI and digital biomarkers to Alzheimer's disease diagnosis and therapy. In the creation and application of these technologies, the authors stress the importance of openness, equity, and privacy.
7. *Clinical relevance*: In addition to accurately classifying Alzheimer's disease, it is crucial to think about how the model's predictions may be applied to enhance patient outcomes. In order to guarantee that the model is being used in a morally and meaningfully appropriate manner, collaboration between researchers, clinicians, and patients is required.

The benefits and drawbacks of using DL models for Alzheimer's disease diagnosis and classification have practical implications for their application in real-world scenarios. On the one hand, DL models can offer advantages such as high accuracy, automation, scalability, and generalization, which can facilitate the early detection and intervention of Alzheimer's disease, reduce the workload and bias of human experts, and enable the analysis of large and heterogeneous data sources. On the other hand, DL models also face challenges such as data quality and availability, interpretability and explainability, ethical and social issues, and technical and computational limitations, which can hinder their adoption and validation in clinical settings, affect their trustworthiness and reliability, and raise concerns about their impact on patients' privacy, autonomy, and well-being. Therefore, it is important to address these challenges and balance these trade-offs when developing and deploying DL models for AD, and to ensure that they are aligned with the best practices and standards of the medical domain.

- The accuracy and automation of DL models could have a practical effect of diminishing the need for costly and invasive diagnostic techniques, such as lumbar punctures or PET scans, which are presently used to validate the diagnosis of AD [129].
- Another practical implication of the scalability and generalization of DL models is that they could enable the integration and analysis of multimodal data, such as MRI, PET, and cognitive tests, which can provide a more comprehensive and personalized assessment of the disease progression and prognosis [130].
- Moreover, a practical challenge of the interpretability and explainability of DL models is that they often act as black boxes, meaning that their decisions and predictions are not transparent or understandable to the human experts or the patients, which can affect their trust and acceptance.

5.4.1 Illustrative Case Studies of Alzheimer's Disease

Case studies offer valuable insights into AD, providing examples that illustrate different aspects of the disease. For instance, one case study examines a man with a rare resilience to AD, while another focuses on a 63-year-old woman with early-onset AD. Additionally, a case report delves into a 37-year-old male patient also affected by early-onset AD. Lastly, a lesson plan incorporates fictitious case studies of AD. Overall, these case studies shed light on the diverse manifestations of AD, including rare resilience, early-onset cases, and the utilization of imaging techniques for diagnosis.

- *Rare Resilience to Alzheimer's Disease*: An NIH-funded case study examined a man who, although having a rare genetic form of early-onset Alzheimer's disease, shown little difficulty with verbal learning or language. In spite of his slim genetic chances, the study sought to explain how he avoided dementia for such a long time <https://www.nia.nih.gov/news/case-study-unlocks-clues-rare-resilience-alzheimers-disease>.
- *Early-Onset Alzheimer's Disease*: A case report included information on the clinical history, autopsy results, and brain histology of a 63-year-old woman with early-onset Alzheimer's disease. The study emphasized the substantial effects of Alzheimer's on patients' brains as well as the genetic abnormalities linked to early-onset Alzheimer's [131].
- *Distinguished Striatum Amyloid Retention*: A 37-year-old male patient with early-onset AD was the subject of another case study. The article covered his progressive cognitive impairment, the diagnosis made with beta-amyloid imaging, and the possible drawbacks of the case study, like the patient's absence of genetic testing [132].

These case studies offer valuable insights into the diverse manifestations of Alzheimer's disease, including rare resilience, early-onset cases, and the use of imaging techniques for diagnosis. By examining specific examples, researchers and healthcare professionals can gain a deeper understanding of the disease and explore potential opportunities for intervention and treatment.

6 Conclusion

In this review, we have provided a comprehensive overview of the use of deep learning models, especially Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs), for classifying Alzheimer's disease (AD) from medical images. The review highlights the potential of deep learning models to improve the accuracy of current diagnostic methods and showcases the promise of both CNN and ViT models in accurately classifying AD from medical images. We have discussed the challenges and considerations for AD classification using deep learning, compared the strengths and weaknesses of each approach, and emphasized their potential in improving AD diagnosis. The discussion section critically analyzes the advantages and limitations of combining CNN and ViT models, shedding light on their potential implications for AD diagnosis and treatment.

Our review aligns with the current trend of employing deep learning methods, which have shown superior performance in handling complex three-dimensional data and have achieved great success in AD diagnosis. The insights provided in this article contribute to the understanding of the current state and potential directions for future research in the field of deep learning-based AD classification from medical images. Throughout the review, the

focus is on deep learning and image classification concepts related to the specific problem of AD classification.

In conclusion, this review provides an insightful overview of the current state and potential future directions in using deep learning models for Alzheimer's disease (AD) classification. It carefully considers the challenges, including dataset biases, interpretability of models, scalability issues, and robustness to variations in imaging modalities, offering a balanced perspective. By acknowledging these challenges, the review aims to foster informed decision-making and guide future research effectively. Additionally, it underscores the importance of addressing ethical considerations and enhancing trust among stakeholders. Through recommendations for improving interpretability, robustness, and fairness of deep learning models for AD classification, the review aims to contribute to the advancement of AD research and clinical practice.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Raghavaiah P, Varadarajan S (2021) A CAD system design to diagnose Alzheimers disease from MRI brain images using optimal deep neural network. *Multimed Tools Appl* 80(17):26411–26428. <https://doi.org/10.1007/s11042-021-10928-7>
2. Mittal A et al (2023) A systematic review of updated mechanistic insights towards Alzheimer's disease. *CNS & Neurol Disorders-Drug Targets*. (Formerly *Current Drug Targets-CNS Neurol Disorders*). 22(8):1232–1242. <https://doi.org/10.2174/1871527321666220510144127>
3. Pelak Victoria S (2022) The clinical approach to the identification of higher-order visual dysfunction in neurodegenerative disease. *Curr Neurol Neurosci Rep* 22(4):229–242. <https://doi.org/10.1007/s11910-022-01186-7>
4. Marwa EL-G et al (2023) An MRI-based deep learning approach for accurate detection of Alzheimer's disease. *Alex Eng J* 63:211–221
5. Qiu S et al (2022) Multimodal deep learning for Alzheimer's disease dementia assessment. *Nat Commun* 13(1):3404. <https://doi.org/10.1038/s41467-022-31037-5>
6. Gaugler J et al (2022) 2022 Alzheimer's disease facts and figures. *Alzheimers Dement* 18(4):700–789
7. Kanthale SB et al (2024) Precision medicines for the treatment of Alzheimer's disease. In: *Alzheimer's disease and advanced drug delivery strategies*. Academic Press, pp 253–263. <https://doi.org/10.1016/B978-0-443-13205-6.00025-X>
8. Acharya M et al (2024) Vitamin D, calbindin, and calcium signaling: unraveling the Alzheimer's connection. *Cell Signal* 116:111043. <https://doi.org/10.1016/j.cellsig.2024.111043>
9. Bae JB et al (2020) Identification of Alzheimer's disease using a convolutional neural network model based on T1-weighted magnetic resonance imaging. *Sci Rep* 10(1):22252
10. Klyucherev TO et al (2022) Advances in the development of new biomarkers for Alzheimer's disease. *Transl Neurodegener* 11(1):1–24. <https://doi.org/10.1186/s40035-022-00296-z>
11. Al-juboori AM et al (2023) A hybrid cracked tiers detection system based on adaptive correlation features selection and deep belief neural networks. *Symmetry* 15(2):358. <https://doi.org/10.3390/sym15020358>
12. Liu S et al (2022) Generalizable deep learning model for early Alzheimer's disease detection from structural MRIs. *Sci Rep* 12(1):17106. <https://doi.org/10.1038/s41598-022-20674-x>
13. Dong S, Wang P, Abbas K (2021) A survey on deep learning and its applications. *Comput Sci Rev* 40:100379

14. Shen D, Wu G, Suk H-I (2017) Deep learning in medical image analysis. *Ann Rev Biomed Eng* 19:221–248
15. Jo T, Nho K, Saykin AJ (2019) Deep learning in Alzheimer's disease: diagnostic classification and prognostic prediction using neuroimaging data. *Front Aging Neurosci* 11:220
16. Lv H et al (2023) Application of curcumin nanoformulations in Alzheimer's disease: prevention, diagnosis and treatment. *Nutr Neurosci* 26(8):727–742
17. Parlak M et al (2022) Investigation of cognition, nutrition, independence and swallowing difficulty, relationship with quality of life, and effect levels in elderly people with Alzheimer's disease living with their families. *Neurol Asia* 27(3):701
18. Sharma L et al (2022) Promising protein biomarkers in the early diagnosis of Alzheimer's disease. *Metab Brain Dis* 37(6):1727–1744
19. Rajan KB et al (2023) Longitudinal changes in blood biomarkers of clinical Alzheimer disease in a biracial population sample. *Neurology* 100(8):e874–e883
20. Mirakhori F et al (2022) Diagnosis and treatment methods in Alzheimer's patients based on modern techniques: the original article. *J Pharm Negat Results* 2022:1889–1907
21. Sun X et al (2015) Histogram-based normalization technique on human brain magnetic resonance images from different acquisitions. *Biomed Eng Online* 14(1):1–17
22. Sethi M et al (2022) An exploration: Alzheimer's disease classification based on convolutional neural network. *BioMed Res Int*. <https://doi.org/10.1155/2022/8739960>
23. Rehman HZU, Hyunho H, Lee S (2020) Conventional and deep learning methods for skull stripping in brain MRI. *Appl Sci* 10(5):1773
24. Xia N et al (2021) Intravoxel incoherent motion diffusion-weighted imaging in the characterization of Alzheimer's disease. *Brain Imaging Behav* 2021:1–10
25. Shamonin DP et al (2014) Fast parallel image registration on CPU and GPU for diagnostic classification of Alzheimer's disease. *Front Neuroinform* 7:50
26. Song Juan et al (2021) An effective multimodal image fusion method using MRI and PET for Alzheimer's disease diagnosis. *Front Digit Health* 3:637386
27. Kim JH et al (2022) Neuroimaging modalities in Alzheimer's disease: diagnosis and clinical features. *Int J Mol Sci* 23(11):6079
28. Katal S et al (2022) Advantages and applications of total-body PET scanning. *Diagnostics* 12(2):426. <https://doi.org/10.3390/diagnostics12020426>
29. B-Ppet D (2009) Pittsburgh Compound. Application of PET imaging to diagnosis of Alzheimer's disease and mild cognitive impairment I. *Neurobiol Dement* 84:133
30. Jdey I, Hcini G, Ltifi H (2023) Deep learning and machine learning for malaria detection: overview, challenges and future directions. *Int J Inf Technol Decis Mak*. <https://doi.org/10.1142/S0219622023300045>
31. Brahmī W, Jdey I (2023) Automatic tooth instance segmentation and identification from panoramic X-Ray images using deep CNN. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-023-17568-z>
32. Brahmī W, Jdey I, Drira F (2024) Exploring the role of convolutional neural networks (CNN) in dental radiography segmentation: a comprehensive systematic literature review. *arXiv preprint arXiv:2401.09190*
33. Cuenat S, Couturier R (2020) Convolutional neural network (CNN) vs vision transformer (VIT) for digital holography. In: 2022 2nd international conference on computer, control and robotics (ICCCR). IEEE, 2022, Shanghai, China, pp 235–240
34. Lyu Y, Yu X, Zhu D et al (2022) Classification of Alzheimer's disease via vision transformer: classification of Alzheimer's disease via vision transformer. In: Proceedings of the 15th international conference on PErvasive technologies related to assistive environments. 2022. Corfu, Greece, pp 463–468
35. Tian J et al (2021) Modular machine learning for Alzheimer's disease classification from retinal vasculature. *Sci Rep* 11(1):238
36. Iqbal S et al (2023) On the analyses of medical images using traditional machine learning techniques and convolutional neural networks. *Arch Comput Methods Eng* 30(5):3173–3233. <https://doi.org/10.1007/s11831-023-09899-9>
37. Sharma N, Jain V, Mishra A (2018) An analysis of convolutional neural networks for image classification. *Procedia Comput Sci* 132:377–384
38. Heni A, Jdey I, Ltifi H, (2022) k-means and fuzzy c-means fusion for object clustering. In: (2022) 8th international conference on control, decision and information technologies (CoDIT), vol 1. IEEE, Istanbul, Turkey. <https://doi.org/10.1109/CoDIT55151.2022.9804078>
39. Chimakurthi VNSS (2020) Application of convolution neural network for digital image processing. *Eng Int* 8(2):149–158

40. Slimani N, Jdey I, Kherallah M (2023) Performance comparison of machine learning methods based on CNN for satellite imagery classification. In: 2023 9th international conference on control, decision and information technologies (CoDIT). IEEE, Rome, Italy
41. Jlassi S, Jdey I, Ltifi H (2021) Bayesian hyperparameter optimization of deep neural network algorithms based on ant colony optimization. In: Document analysis and recognition-ICDAR 2021: 16th international conference, Lausanne, Switzerland, September 5–10, Proceedings, Part III 16. Springer
42. Hcini G et al (2021) Hyperparameter optimization in customized convolutional neural network for blood cells classification. *J Theoret Appl Inf Technol* 99:5425–5435
43. Yamashita R et al (2018) Convolutional neural networks: an overview and application in radiology. *Insights Imaging* 9:611–629
44. Hcini G, Jdey I, Ltifi H (2022) Improving malaria detection using L1 regularization neural network. *JUCS: J Univers Comput Sci* 285(10):1087–1107. <https://doi.org/10.3897/jucs.81681>
45. Hcini G, Jdey I, Ltifi H (2023) HSV-Net: a custom cnn for malaria detection with enhanced color representation. In: 2023 International conference on cyberworlds (CW). IEEE (2023). Sousse, Tunisia, pp 337–340
46. Guo J, Han K, Wu H et al (2022) Cmt: convolutional neural networks meet vision transformers. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. New Orleans, LA, USA, pp 12175–12185
47. Bouzidi S, Jdey I, Alimi A A Vision transformer approach with L2 regularization for sustainable fashion classification. Available at SSRN 4686032
48. Rothman D, Gulli A (2022) Transformers for natural language processing: build, train, and fine-tune deep neural network architectures for NLP with Python, PyTorch, TensorFlow, BERT, and GPT-3. Packt Publishing Ltd
49. Yang Y et al (2022) Transformers meet visual learning understanding: a comprehensive review. *arXiv* 2022. *arXiv preprint* [arXiv:2203.12944](https://arxiv.org/abs/2203.12944)
50. Naseer MM et al (2021) Intriguing properties of vision transformers. *Adv Neural Inf Process Syst* 34:23296–23308
51. Carion N, Massa F, Synnaeve G et al (2020) End-to-end object detection with transformers. In: European conference on computer vision. Springer, Cham (2020). Glasgow, UK, pp 213–229
52. Dosovitskiy A et al (2020) An image is worth 16x16 words: transformers for image recognition at scale. *arXiv preprint* [arXiv:2010.11929](https://arxiv.org/abs/2010.11929)
53. Touvron H, Cord M, Douze M et al (2021) Training data-efficient image transformers & distillation through attention. In: International conference on machine learning. PMLR, pp 10347–10357
54. Liu Z, Lin Y, Cao Y et al (2021) Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF international conference on computer vision. Montreal, QC, pp 10012–10022
55. He K, Gan C, Li Z, Reik I, Yin Z, Ji W, Gao Y, Wang Q, Zhang J, Shen D (2022) Transformers in medical image analysis: a review. [ArXiv:2202.12165](https://arxiv.org/abs/2202.12165)
56. Courant R et al (2023) Transformers and visual transformers. *Mach Learn Brain Disord* 197:193–229
57. Deininger L et al (2022) A comparative study between vision transformers and CNNs in digital pathology. *arXiv preprint* [arXiv:2206.00389](https://arxiv.org/abs/2206.00389)
58. Maurício J, Domingues I, Bernardino J (2023) Comparing vision transformers and convolutional neural networks for image classification: a literature review. *Appl Sci* 13(9):5521
59. Matsoukas C et al (2021) Should we replace CNNs with transformers for medical images? International conference on learning representations (2021). Kigali, Rwanda, pp 1–15
60. Warren SL, Moustafa AA (2023) Functional magnetic resonance imaging, deep learning, and Alzheimer's disease: a systematic review. *J Neuroimaging* 33(1):5–18
61. Arya AD et al (2023) A systematic review on machine learning and deep learning techniques in the effective diagnosis of Alzheimer's disease. *Brain Inform* 10(1):17
62. Shukla A, Tiwari R, Tiwari S (2023) Review on Alzheimer disease detection methods: automatic pipelines and machine learning techniques. *Sci* 5(1):13. <https://doi.org/10.3390/sci5010013>
63. Dara OA et al (2023) Alzheimer's disease diagnosis using machine learning: a survey. *Appl Sci* 13(14):8298. <https://doi.org/10.3390/app13148298>
64. Gharaibeh N et al (2023) Swin transformer-based segmentation and multi-scale feature pyramid fusion module for Alzheimer's disease with machine learning. *Int J Online Biomed Eng* 19(4):22–50
65. Al Shehri W (2022) Alzheimer's disease diagnosis and classification using deep learning techniques. *PeerJ Comput Sci* 8:e1177. <https://doi.org/10.7717/peerj-cs.1177>
66. Oh K et al (2019) Classification and visualization of Alzheimer's disease using volumetric convolutional neural network and transfer learning. *Sci Rep* 9(1):18150

67. Rajendiran M et al (2022) Detection of Alzheimer's disease in MRI images using different transfer learning models and improving the classification accuracy. *Int J Health Sci* 6:11851–11869. <https://doi.org/10.53730/ijhs.v6nS3.8944>
68. Sharma S et al (2022) HTLML: hybrid ai based model for detection of Alzheimer's disease. *Diagnostics* 12(8):1833
69. Shahwar T et al (2022) Automated detection of Alzheimer's via hybrid classical quantum neural networks. *Electronics* 11(5):721
70. Kumari R, Nigam A, Pushkar S (2020) Machine learning technique for early detection of Alzheimer's disease. *Microsyst Technol* 26:3935–3944
71. Ghazal TM et al (2022) Alzheimer disease detection empowered with transfer learning. *Comput Mater Contin* 70(3):5005–5019
72. Nancy Noella RS, Priyadarshini J (2020) Diagnosis of Alzheimer's and Parkinson's disease using artificial neural network. *Int J Sci Technol Res* 9(3):3659–3664
73. Shankar K et al (2019) Alzheimer detection using Group Grey Wolf Optimization based features with convolutional classifier. *Comput Electr Eng* 77:230–243
74. Sisodia PS et al (2023) A review of deep transfer learning approaches for class-wise prediction of Alzheimer's disease using MRI images. *Arch Comput Methods Eng* 30(4):2409–2429
75. Mohi ud din dar G, et al (2023) A novel framework for classification of different Alzheimer's disease stages using CNN model. *Electronics* 12(2):469
76. Odusami M et al (2021) Analysis of features of Alzheimer's disease: Detection of early stage from functional brain changes in magnetic resonance images using a finetuned ResNet18 network. *Diagnostics* 11(6):1071. <https://doi.org/10.3390/diagnostics11061071>
77. Ebrahimi A et al (2021) Deep sequence modelling for Alzheimer's disease detection using MRI. *Comput Biol Med* 134:104537
78. Shamrat FMJM et al (2023) AlzheimerNet: an effective deep learning based proposition for Alzheimer's disease stages classification from functional brain changes in magnetic resonance images. *IEEE Access* 11:16376–16395
79. Raza N et al (2023) Alzheimer disease classification through transfer learning approach. *Diagnostics* 13(4):801. <https://doi.org/10.3390/diagnostics13040801>
80. Shanmugam JV et al (2022) Alzheimer's disease classification using pre-trained deep networks. *Biomed Signal Process Control* 71:103217
81. Ganesh D et al (2023) Implementation of convolutional neural networks for detection of Alzheimer's disease. *BioGecko, A J N Z Herpetol* 12(01)
82. Alsharabi N et al (2023) Implementing magnetic resonance imaging brain disorder classification via AlexNet-quantum learning. *Mathematics* 11(2):376. <https://doi.org/10.3390/math11020376>
83. Balaji P et al (2023) Hybridized deep learning approach for detecting Alzheimer's disease. *Biomedicines* 11(1):149. <https://doi.org/10.3390/biomedicines11010149>
84. de Silva K, Kunz H (2023) Prediction of Alzheimer's disease from magnetic resonance imaging using a convolutional neural network. *Intelligence-Based Med* 7:100091
85. Sarraf Saman et al (2023) OViTAD: optimized vision transformer to predict various stages of Alzheimer's disease using resting-state fMRI and structural MRI data. *Brain Sci* 13(2):260
86. Hoang GM, Kim U-H, Kim JG (2023) Vision transformers for the prediction of mild cognitive impairment to Alzheimer's disease progression using mid-sagittal sMRI. *Front Aging Neurosci* 15:1102869
87. Shin H et al (2023) Vision transformer approach for classification of Alzheimer's disease using 18F-Florbetaben brain images. *Appl Sci* 13(6):3453. <https://doi.org/10.3390/app13063453>
88. Almufareh MF et al (2023) Artificial cognition for detection of mental disability: a vision transformer approach for Alzheimer's disease. *Healthcare* 11(20):1–15. <https://doi.org/10.3390/healthcare11202763>
89. Ahmad Khan Arfat et al (2024) Dual-3DM 3-AD: mixed transformer based semantic segmentation and triplet pre-processing for early multi-class Alzheimer's diagnosis. *IEEE Trans Neural Syst Rehab Eng* 32:696–707
90. Mahim SM et al (2024) Unlocking the potential of XAI for improved alzheimer's disease detection and classification using a ViT-GRU model. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2024.3351809>
91. Kadri R et al (2021) CrossViT wide residual squeeze-and-excitation network for Alzheimer's disease classification with self attention ProGAN data augmentation. *Int J Hybrid Intell Syst* 17(3–4):163–177
92. Wu X et al (2022) Classification of Alzheimer's disease based on weakly supervised learning and attention mechanism. *Brain Sci* 12(12):1601. <https://doi.org/10.3390/brainsci12121601>
93. Jang J, Hwang D (2022) M3T: three-dimensional Medical image classifier using multi-plane and Multi-slice Transformer. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. New Orleans, LA

94. Li C et al (2022) An attention-based CoT-ResNet with channel shuffle mechanism for classification of Alzheimer's disease levels. *Front Aging Neurosci* 14:930584
95. Carcagni P et al (2023) Convolution neural networks and self-attention learners for Alzheimer dementia diagnosis from brain MRI. *Sensors* 23(3):1694. <https://doi.org/10.3390/s23031694>
96. Khorramyar P (2022) Visual transformers for 3D medical images classification: use-case neurodegenerative disorders
97. Jiang M et al (2022) Image classification of Alzheimer's disease based on external-attention mechanism and fully convolutional network. *Brain Sci* 12(3):319. <https://doi.org/10.3390/brainsci12030319>
98. Zhang J et al (2021) A 3D densely connected convolution neural network with connection-wise attention mechanism for Alzheimer's disease classification. *Magn Reson Imaging* 78:119–126
99. Chen L, Qiao H, Zhu F (2022) Alzheimer's disease diagnosis with brain structural MRI using multiview-slice attention and 3D convolution neural network. *Front Aging Neurosci* 14:871706
100. Wu Y et al (2022) An attention-based 3D CNN with multi-scale integration block for Alzheimer's disease classification. *IEEE J Biomed Health Inform* 26(11):5665–5673
101. Wang S-H et al (2021) ADVIAN: Alzheimer's disease VGG-inspired attention network based on convolutional block attention module and multiple way data augmentation. *Front Aging Neurosci* 13:687456
102. Chen Y (2021) An Alzheimer's disease identification and classification model based on the convolutional neural network with attention mechanisms. *Traitement du Signal*. <https://doi.org/10.18280/ts.380533>
103. Liang S, Gu Y (2020) Computer-aided diagnosis of Alzheimer's disease through weak supervision deep learning framework with attention mechanism. *Sensors* 21(1):220. <https://doi.org/10.3390/s21010220>
104. Liu Y et al (2022) MPC-STANet: Alzheimer's disease recognition method based on multiple phantom convolution and spatial transformation attention mechanism. *Front Aging Neurosci* 14:918462
105. Wang JX et al (2022) Alzheimer's disease classification through imaging genetic data with IGnet. *Front Neurosci* 16:846638
106. Hasan ME, Wagler A (2024) New convolutional neural network and graph convolutional network-based architecture for AI applications in Alzheimer's disease and dementia-stage classification. *AI* 5(1):342–363. <https://doi.org/10.3390/ai5010017>
107. Khan A et al (2023) A survey of the Vision Transformers and its CNN-transformer based Variants. arXiv preprint [arXiv:2305.09880](https://arxiv.org/abs/2305.09880)
108. Wang K et al (2023) Combining convolutional neural networks and self-attention for fundus diseases identification. *Sci Rep* 13(1):76
109. Alzubaidi L et al (2021) Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *J Big Data* 8:1–74. <https://doi.org/10.1186/s40537-021-00444-8>
110. Moayed H, Mansoori EG (2023) Improving regularization in deep neural networks by co-adaptation trace detection. *Neural Process Lett* 55:1–13
111. Tan C et al (2018) A survey on deep transfer learning. Artificial neural networks and machine learning-ICANN 2018: 27th international conference on artificial neural networks. Rhodes, Greece, October 4–7, Proceedings, Part III 27. Springer
112. Shorten C, Khoshgoftaar TM (2019) A survey on image data augmentation for deep learning. *J Big Data* 6(1):1–48
113. Saleh AM, Hamoud T (2021) Analysis and best parameters selection for person recognition based on gait model using CNN algorithm and image augmentation. *J Big Data* 8(1):1–20. <https://doi.org/10.1186/s40537-020-00387-6>
114. Alzubaidi L et al (2023) A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications. *J Big Data* 10(1):46
115. Baghdadi NA et al (2022) A3c-TL-GTO: Alzheimer automatic accurate classification using transfer learning and artificial gorilla troops optimizer. *Sensors* 22(11):4250
116. El-Assy AM et al (2024) A novel CNN architecture for accurate early detection and classification of Alzheimer's disease using MRI data. *Sci Rep* 14(1):3463
117. Li Y et al (2018) DEEPre: sequence-based enzyme EC number prediction by deep learning. *Bioinformatics* 34(5):760–769. <https://doi.org/10.1093/bioinformatics/btx680>
118. Prasath T, Sumathi V (2023) Identification of Alzheimer's disease by imaging: a comprehensive review. *Int J Environ Res Public Health* 20(2):1273. <https://doi.org/10.3390/ijerph20021273>
119. Venugopalan J et al (2021) Multimodal deep learning models for early detection of Alzheimer's disease stage. *Sci Rep* 11(1):3254
120. Zokaeinikoo M, Kazemian P, Mitra P (2023) Interpretable hierarchical deep learning model for noninvasive Alzheimer's disease diagnosis. *INFORMS J Data Sci* 2(2):183–196
121. Martin SA et al (2023) Interpretable machine learning for dementia: a systematic review. *Alzheimer's Dement*. <https://doi.org/10.1002/alz.12948>

122. Tang Z et al (2019) Interpretable classification of Alzheimer's disease pathologies with a convolutional neural network pipeline. *Nat Commun* 10(1):2173
123. Alzubaidi L et al (2023) Towards risk-free trustworthy artificial intelligence: significance and requirements. *Int J Intell Syst* 2023:1–41. <https://doi.org/10.1155/2023/4459198>
124. Moulahi W et al (2023) A blockchain-based federated learning mechanism for privacy preservation of healthcare IoT data. *Comput Biol Med* 167:107630. <https://doi.org/10.1016/j.combiomed.2023.107630>
125. Jdey I (2022) Trusted smart irrigation system based on fuzzy IoT and blockchain. In: International conference on service-oriented computing. Springer, Cham, pp 154–165, Sevilla, Spain
126. Petti U et al (2023) Ethical considerations in the early detection of Alzheimer's disease using speech and AI. In : Proceedings of the 2023 ACM conference on fairness, accountability, and transparency. Chicago, IL, pp 1062–1075
127. Ursin F, Timmermann C, Steger F (2021) Ethical implications of Alzheimer's disease prediction in asymptomatic individuals through artificial intelligence. *Diagnostics* 11(3):440. <https://doi.org/10.3390/diagnostics11030440>
128. Ford E, Milne R, Curlewis K (2023) Ethical issues when using digital biomarkers and artificial intelligence for the early detection of dementia. *Wiley Interdiscip Rev Data Min Knowl Discov* 13(3):e1492
129. Parmar H, Walden E (2022) Towards practical application of deep learning in diagnosis of Alzheimer's disease. arXiv preprint [arXiv:2212.04528](https://arxiv.org/abs/2212.04528)
130. Anand D et al (2023) Deep learning-based Alzheimer's disease classification using transfer learning and data augmentation. 2023 3rd international conference on intelligent technologies (CONIT). IEEE, Huzhou, China, pp 1–6
131. Wells Jennie L, Pasternak Stephen H (2019) Case report of a 63-year-old patient with Alzheimer disease and a novel presenilin 2 mutation. *Alzheimer Disease Assoc Disord* 33(2):166
132. Hyun Um Yoo et al (2017) A case report of a 37-year-old Alzheimer's disease patient with prominent striatum amyloid retention. *Psychiatry Investig* 14(4):521

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.