

Detection of Alzheimer's disease using deep learning models: A systematic literature review

Eqtidar M. Mohammed^a, Ahmed M. Fakhrudeen^b, Omar Younis Alani^{c,*}

^a Department of Computer Science, College of Computer Science and Information Technology, University of Kirkuk, Kirkuk, Iraq

^b Department of Software, College of Computer Science and Information Technology, University of Kirkuk, Kirkuk, Iraq

^c School of Science, Engineering & Environment University of Salford, Manchester, M5 4WT, UK

ARTICLE INFO

Keywords:

AD
Alzheimer's disease
Mild cognitive impairment
Datasets
Deep learning
Challenges

ABSTRACT

Alzheimer's disease (AD) is a progressive neurological disease considered the most common form of late-stage dementia. Usually, AD leads to a reduction in brain volume, impacting various functions. This article comprehensively analyzes the AD context in fivefold main topics. Firstly, it reviews the main imaging techniques used in diagnosing AD disease. Secondly, it explores the most proposed deep learning (DL) algorithms for detecting the disease. Thirdly, the article investigates the commonly used datasets to develop DL techniques. Fourthly, we conducted a systematic review and selected 45 papers published in highly ranked publishers (Science Direct, IEEE, Springer, and MDPI). We analyzed them thoroughly by delving into the stages of AD diagnosis and emphasizing the role of preprocessing techniques. Lastly, the paper addresses the remaining practical implications and challenges in the AD context. Building on the analysis, this survey contributes to covering several aspects related to AD disease that have not been studied thoroughly.

1. Introduction

1.1. Definition of Alzheimer's disease

Alzheimer's disease (AD) is a clinical syndrome characterized by a progressive decline in memory and cognitive ability. It is a common disorder among older people, with around 60–80 % prevalence. Despite its high prevalence, there is currently no cure for this condition. The timeline from the onset of AD to its ultimate diagnosis is lengthy [1]. Additionally, in elderly patients, its impact extends beyond oxidative burst, contributing to a progressive, irreversible, and incurable behavioral pattern. This deeper understanding of the disease's pathology [2].

Researchers focus their efforts on developing a straightforward and precise methodology for detecting AD in its early stages before symptoms become apparent. Thus, the early detection of AD has facilitated the identification of AD symptoms prior to the onset of high-risk stages [3]. Fig. 1 shows a comparison made between a healthy brain (left) and a brain suffering from AD (right).

Mild cognitive Impairment (MCI) refers to the initial phases of AD when mild symptoms are present [4]. Each year, 10 %–15 % of patients diagnosed with MCI transform AD [5,6]. Accordingly, accurate and

early diagnosis of AD is vital [7]. MCI stages are categorized into two main stages: Early MCI (EMCI) (also known as Stable MCI (sMCI)) and Late MCI (LMCI) (also known as Progressive MCI (pMCI)). While pMCI patients exhibit a gradual progression of AD symptoms, sMCI cases remain stable without significant deterioration [8].

For the elderly, AD is a growing concern in the United States. Currently, ranking as the sixth leading cause of mortality. However, new estimates suggest it could climb to 3rd, behind only cancer and heart diseases [9]. Fig. 2 depicts the correlation between age and the prevalence of AD in individuals aged 65 and older in the USA. The figure serves as a valuable visual aid to underscore the relationship between age demographics and the incidence of AD within the specified population [9]. Furthermore, Fig. 3 shows a significant increase in the population diagnosed with AD in the USA. By 2060, the number of individuals aged 65 and older with AD is expected to reach 13.8 million [10].

Therefore, Early detection and prevention of AD are critical goals. Diagnosis of AD requires an array of medical examinations, resulting in a large amount of multivariate heterogeneous data (complex and diverse data). Nevertheless, manual analysis and interpretation of the data is challenging and time-consuming due to its complexity. The

* Corresponding author.

E-mail addresses: eqtidar.m@uokirkuk.edu.iq (E.M. Mohammed), dr.ahmed.fakhrudeen@uokirkuk.edu.iq (A.M. Fakhrudeen), O.Y.K.Alani@salford.ac.uk (O.Y. Alani).

<https://doi.org/10.1016/j.imu.2024.101551>

Received 4 March 2024; Received in revised form 10 July 2024; Accepted 11 July 2024

Available online 14 July 2024

2352-9148/© 2024 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

classification of Magnetic Resonance Imaging (MRI) scans is a valuable tool for accurately predicting brain health. However, classifying them to predict AD remains a significant challenge. To address these challenges, scholars are developing novel algorithms that employ label propagation with Artificial Intelligence (AI) to achieve an efficient brain MRI classification [10].

1.2. Context of the AD

In 1910, Kraepelin published a detailed account of specific cases characterized by severe cell transformations. These changes included an excessive number of plaques and the loss of about a third of the brain's outer layer (cerebral cortex). The topic at hand is significant due to how it is presented when the medical description of AD lacks clarity, emphasizing the criticality of Kraepelin's contribution [11]. Dr. Alois Alzheimer is the first scholar that coined "Alzheimer's disease". In 1906, through the case of Auguste Deter, the initial disease detection was somewhat ambiguous. Yet, for more than one hundred years, credible descriptions of the clinical definition of AD have begun to emerge. In 1907, Dr. Alois Alzheimer described the disease, which was later supplemented by Proskin in 1909. These descriptions included the presence of senile plaques and neurofibrillary sections [12]. Although a clinical examination revealed no significant indications of arteriosclerosis, its potential contribution to the diagnosis couldn't be entirely ruled out. In 1998, scholars from the Max Planck Institute of Neurobiology in Martinsried and the prestigious University of Munich, Germany, investigated a study of the vulnerability of specific brain regions to the buildup of amyloid plaques and neurofibrillary tangles [13]. The study is considered to have re-examined the brain of the first identified patient of AD.

Histological analysis was performed in 1997 by Dr. Gerber and his colleagues from the Psychiatric Department of the Max Planck Institute of Neurobiology. The examination involved slices obtained from F. Johan, whose brain tissues had been meticulously maintained for over nine decades. This research was deemed as the second well-documented case of AD. Upon examining the said cuts, a multitude of amyloid plaques were discovered. Based on the study, a mutational analysis of well-preserved brain tissue was postulated to be possible. Dr. Alzheimer's paradigm-shifting findings were once more validated on the centenary occasion of its inception [14].

1.3. Research questions

The following research questions guide the scope of this survey.

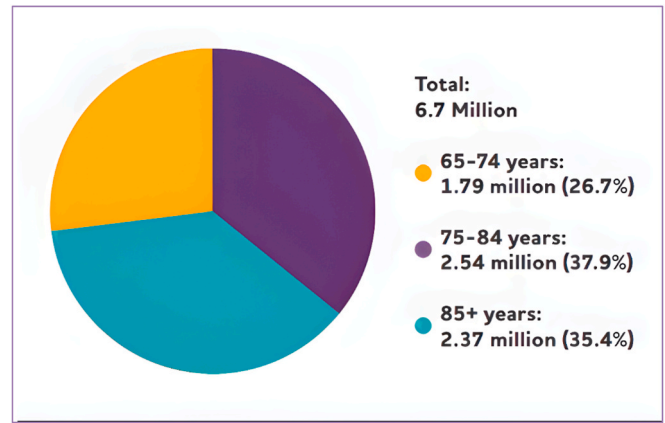


Fig. 2. Age and number of AD patients aged 65 and older in the USA [9].

- **RQ1:** Which specific Deep learning (DL) techniques have been shown to be effective for AD Diagnosis?
- **RQ2:** What imaging techniques could be made more specific to AD by mentioning modalities commonly implemented for AD diagnosis?
- **RQ3:** what are the common AD datasets used to identify more precise biomarkers to detect early-stage AD and track the disease's progression?
- **RQ4:** what is the Pre-processing of data for AD studies?
- **RQ5:** What metrics have been used to evaluate the AD diagnosis or classification models?
- **RQ6:** What are the challenges and open research directions in AD diagnosing?

The questions' order follows from general techniques (RQ1 and RQ2) to specific considerations for AD research (RQ3 and RQ4) and, finally, evaluation and future directions (RQ5 and RQ6).

While numerous surveys and reviews on dementia diagnosis appear in top-ranked journals (e.g., Science Direct, IEEE, and Springer), our survey offers a distinct perspective. Unlike previous works that often encompass various dementia diseases, our survey concentrates solely on AD, such as [15–17], and [18]. Additionally, other surveys, such as [19–27], did not cover the main points that we investigated in the current work. Accordingly, Table 1 summarizes the main focuses and major contributions of this work compared to previous comprehensive surveys.

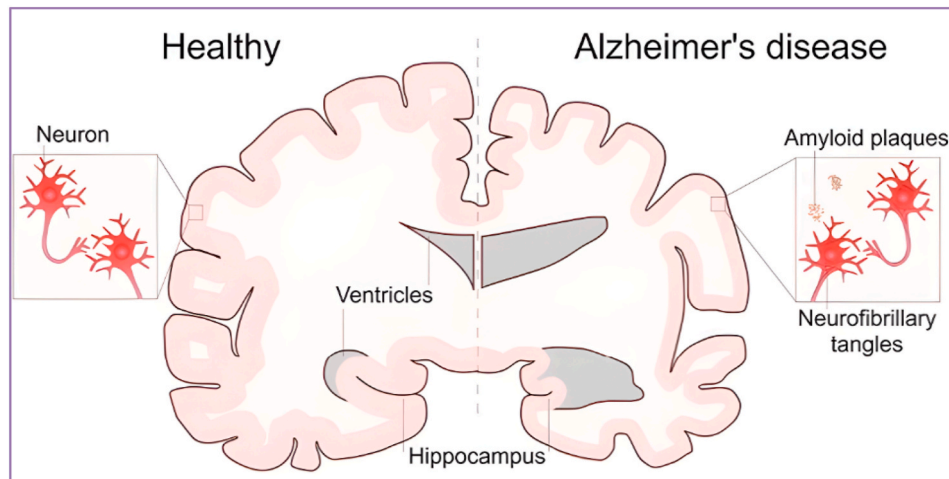


Fig. 1. Healthy brain vs. AD [1].

1.4. Paper organization

This survey is structured as follows. Section 2 focuses on DL techniques that have been utilized for AD diagnosis. Section 3 explains the commonly utilized brain imaging techniques in diagnosing AD. Section 4 investigates the most popular datasets utilized in the literature. Because of its importance, we dedicated Section 5 to explaining the pre-processing of data in AD studies (i.e., how to prepare data for analysis). Section 6 presents all utilized matrices in evaluating the performance of AD models. Section 7 thoroughly analyzes the performance of the selected models. Section 8 outlines the challenges and provides open research directions in AD diagnosis. Section 9 presents key findings and insights from this survey. Finally, Section 10 concludes this paper and explains our primary future focus. The road map of this article is shown in Fig. 4.

2. DL implementation in disease diagnosis

2.1. Preliminaries

This section answers RQ1 regarding the DL utilization in AD diagnosis. Artificial Intelligence (AI) techniques have been extensively utilized to advance therapeutic development. Including disease diagnosis, AI has emerged as a powerful tool in various medical applications. DL techniques have shown promising diagnostic performance, particularly when applied to medical imaging data. This applies to various medical imaging modalities [28], including Magnetic Resonance Imaging (MRI) in the Diagnosis of AD, brain tumors for example [29–31], Parkinson's disease for example [32–34], and chest X-rays such as COVID-19 for example [35–37]. Very briefly, Arthur Samuel introduced the Machine Learning (ML) concept in 1959, and it's about learning from data to make predictions or decisions for various tasks [38]. Supervised, Semi-supervised, and Unsupervised are the main categories of ML algorithms [39]. Fig. 5 shows the timeline of the development of ML algorithms.

However, extracting meaningful patterns from raw medical imaging presents significant challenges for ML models. Furthermore, feature extraction further complicates the process, demanding domain knowledge to select informative features. Incorrect handling of raw data or feature extraction can lead to suboptimal model performance [40]. Therefore, Deep Learning (DL) techniques have emerged as a powerful solution. The next subsection introduces the reader to the most popular utilized DL technique in AD detection.

2.2. DL techniques

DL refers to an area within the field of ML that utilizes artificial neural networks consisting of multiple layers to extract and capture complex features and patterns from diverse data sources, including images, numerical values, and categorical values [42]. Here, we explain the most popular utilized DL models in AD diagnosis.

2.2.1. Convolutional neural network (CNN)

CNN is a widely recognized DL technique considered one of the most effective image classification and recognition techniques in thin neural networks [43]. As depicted in Fig. 6, CNN is structured with multiple layers, including convolution layers, pooling layers, activation layers, fully connected layers, and a classifier layer.

The **convolution layer** is pivotal in feature extraction by passing the input image through learned filters or kernels of specific sizes [44]. Activation functions play a critical role in threshold-based classifiers. While activation functions don't directly. This implies that the decision of whether the linear transformation's value should activate the neuron is crucial. Subsequently, it can be stated that a neuron is activated when the input to the activation function exceeds a specific threshold value; otherwise, deactivation occurs. Such as sigmoid, Hyperbolic Tangent (Tanh), and Rectified Linear Unit (ReLU), various activation functions can be employed to generate feature maps [45].

While the convolution layers focus on feature extraction, **pooling layers** are responsible for dimensionality reduction. Techniques, such as max pooling or average pooling, summarize the information within a small region of the feature map while retaining essential features [46]. A **fully connected layer** establishes connections between every neuron from the preceding layer and all neurons in the current layer. Lastly, the **classifier layer** identifies the class or label with the highest probabilities [47]. A notable characteristic of CNNs is their ability to handle extensive datasets, resulting in high-performance outcomes for classification tasks [48].

In the domain of transfer learning [49], a powerful technique for leveraging knowledge from existing models, CNN pre-trained on massive datasets. The most popular pre-trained CNN architectures include Visual Geometry Group VGGNet based on [50], LeNet [51,52], GoogLeNet based on [53], Residual Network (ResNet) [54], and AlexNet based [55].

Leveraging a pre-trained CNN model allows developers to utilize the model's parameters (i.e., weights) and apply them to new tasks for improved performance [56,57]. Fig. 7 illustrates the history of deep CNN, offering valuable insights into its progressive development over time.

Selecting the optimal architecture for feature extraction and

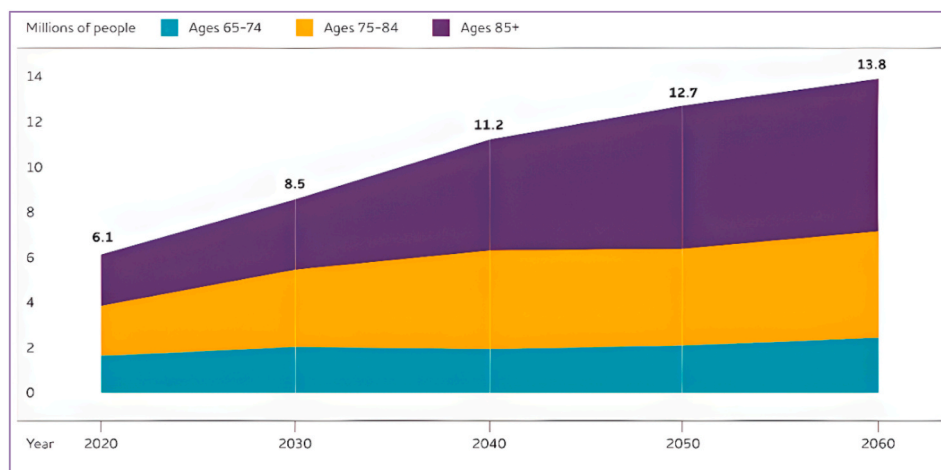


Fig. 3. Estimated population size of people 65 years and older with AD in the United States from 2020 to 2060 [9].

Table 1

The difference between the current study and previous surveys and reviews.

Ref.	Year	Publisher	Journal	Description of Imaging Techniques	Description of DL algorithms and models	Description of Performance Evaluation	AD Datasets	Systematic Review	Challenges
[19]	2020	IEEE	IEEE Access	3	CNN	×	5	×	×
[20]	2021	Science Direct	Materials Today: Processing	×	SVM	×	1	✓	×
[21]	2021	IEEE	4th Biennial International Conference on Nascent Technologies in Engineering (ICNTE)	×	CNN (3D DenseNet, GCNN) and LSTM	×	×	×	×
[15]	2021	IEEE	IEEE Access	×	×	×	8	✓	✓
[22]	2022	Science Direct	International Journal of Cognitive Computing in Engineering	×	×	9	3	×	✓
[23]	2022	Springer	The Egyptian Journal of Neurology, Psychiatry and Neurosurgery	×	CNN (VGG, DenseNet and 3DCNN)	×	×	×	×
[24]	2022	Springer	Multimedia Tools and Applications	4	DNN, CNN, RNN, AE and DBN	×	×	✓	✓
[25]	2023	Springer	Archives of Computational Methods in Engineering	×	CNN (Xception, EfficientB7, ResNet50, DenseNet201 and VGG19)	5	1	✓	×
[26]	2023	Springer	Brain Informatics	×	ANN, CNN (ResNet101, VGG, VoxCNN and DenseNet) and RNN	6	3	×	✓
[27]	2023	Science Direct	International Journal of Intelligent Networks	×	×	×	4	×	✓
Current Study				6	CNN (ResNet, VGG, AlexNet, EfficientNetB7 and GoogleLeNet), DNN, RNN, LSTM, GRU, AE and DBN	11	8	✓	✓

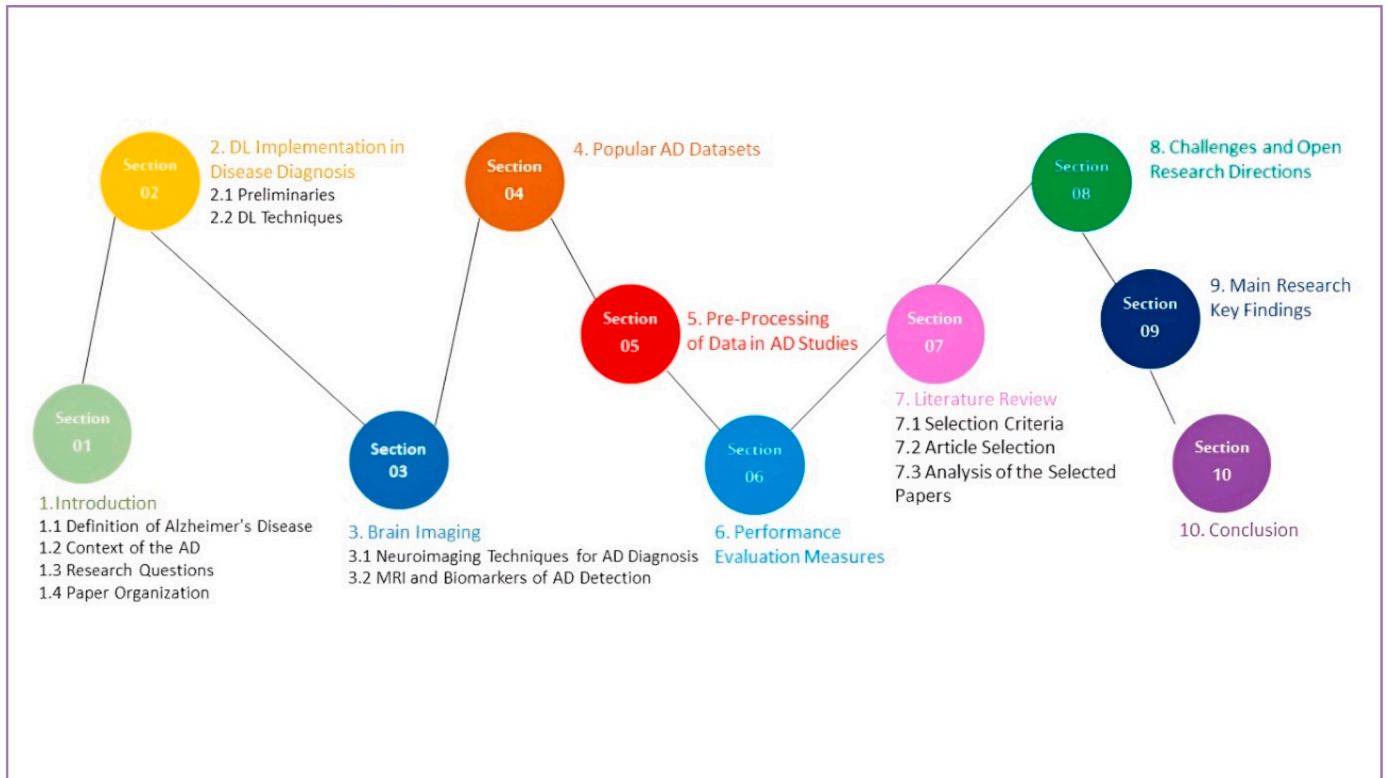


Fig. 4. The structure of the survey.

classification in CNNs can be time-consuming. Frequently necessitating a series of trial-and-error (hyperparameter tuning) experiments to assess different alternatives. Various CNN architectures are accessible, each distinguished by a unique arrangement of convolutional, pooling, ReLU,

fully connected, and custom layers (see an example in Fig. 8). The subsequent section delves into the most prevalent CNN architectures commonly utilized for medical image analysis.

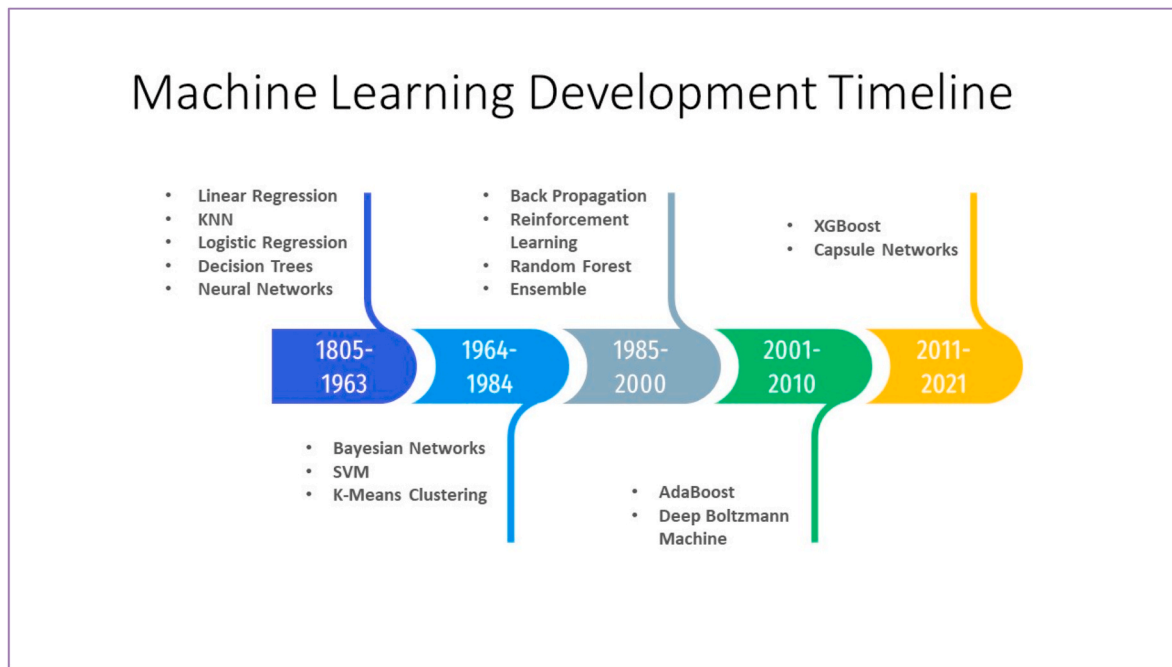


Fig. 5. ML algorithms development timeline (inspired from Ref. [41]).

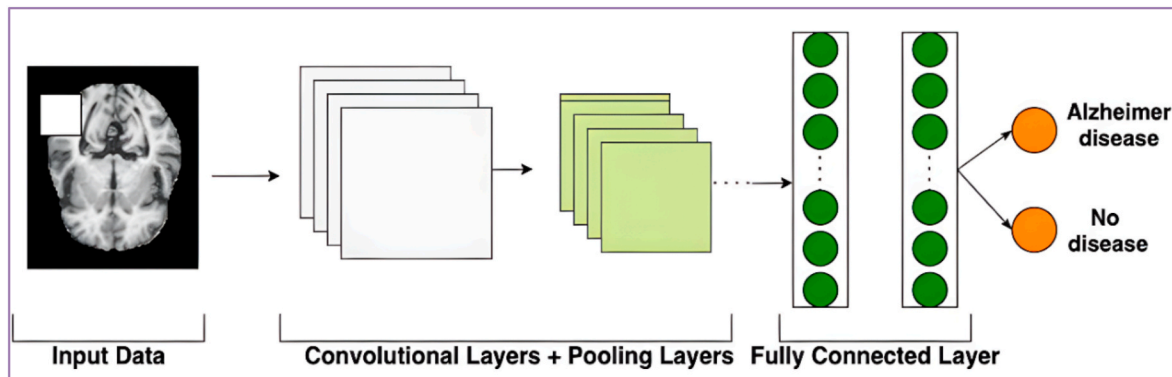


Fig. 6. CNN architecture [24].

2.2.1.1. ResNet. In 2015, Kaiming He and colleagues introduced the architecture of ResNet [54]. The developers tackled the issue of deeper networks through the incorporation of skip connections, enabling a pathway for the network to train its initial layer effectively. Fig. 9 depicts the identity blocks present in ResNet. The total number of layers in ResNet can reach up to 152 without experiencing any degradation. Max-pooling operations are carried out by the convolution layers with a stride value of 2 [62].

2.2.1.2. AlexNet. AlexNet was established by Krizhevsky et al. [55] in 2012. The model consists of approximately 61 million learnable parameters (also referred to as filters), establishing itself as the leading CNN architecture of its era. This model comprised five convolution layers and three fully connected layers. Pioneering the use of the ReLU activation function, a significant advancement that accelerated training speed, Krizhevsky's AlexNet architecture (see Fig. 10) paved the way for DL success in computer vision purposes [62].

2.2.1.3. Visual Geometry Group (VGG). The VGG model was developed by Karen Simonyan et al. [50] in 2014 and represents a more complex architecture compared to AlexNet. VGG-16 typically utilizes thirteen

convolution layers, followed by ReLU activation and max-pooling layers. For classification, these layers are followed by three fully connected layers. VGG-16 is equipped with 138 million learnable parameters, demonstrating its increased capacity compared to AlexNet by more than twofold. VGG-19, an even more intricate model within the VGG family, features a deeper architecture with additional convolutional layers [62]. A visual representation of the VGG-16 architecture can be found in Fig. 11.

2.2.1.4. EfficientNetB7. It is a result of iterative improvements implemented on its predecessors within the EfficientNet family. Foundational to the original EfficientNet design is the utilization of Mobile Inverted Bottleneck Convolution (MBConv). The range of EfficientNet structures consists of several pre-trained models, ranging from EfficientNetB0 to EfficientNetB7. They differ in the number of MBConv blocks, influencing their complexity and accuracy. Significantly, the EfficientNetB7 model emerges as the most efficient model within the EfficientNet series. The model outperforms its counterparts on ImageNet while also showcasing a reduction in size and an improvement in speed compared to contemporary CNN models. EfficientNetB7 is structured into seven sections based on filter dimensions, stride, and channel quantities [25,59]. A

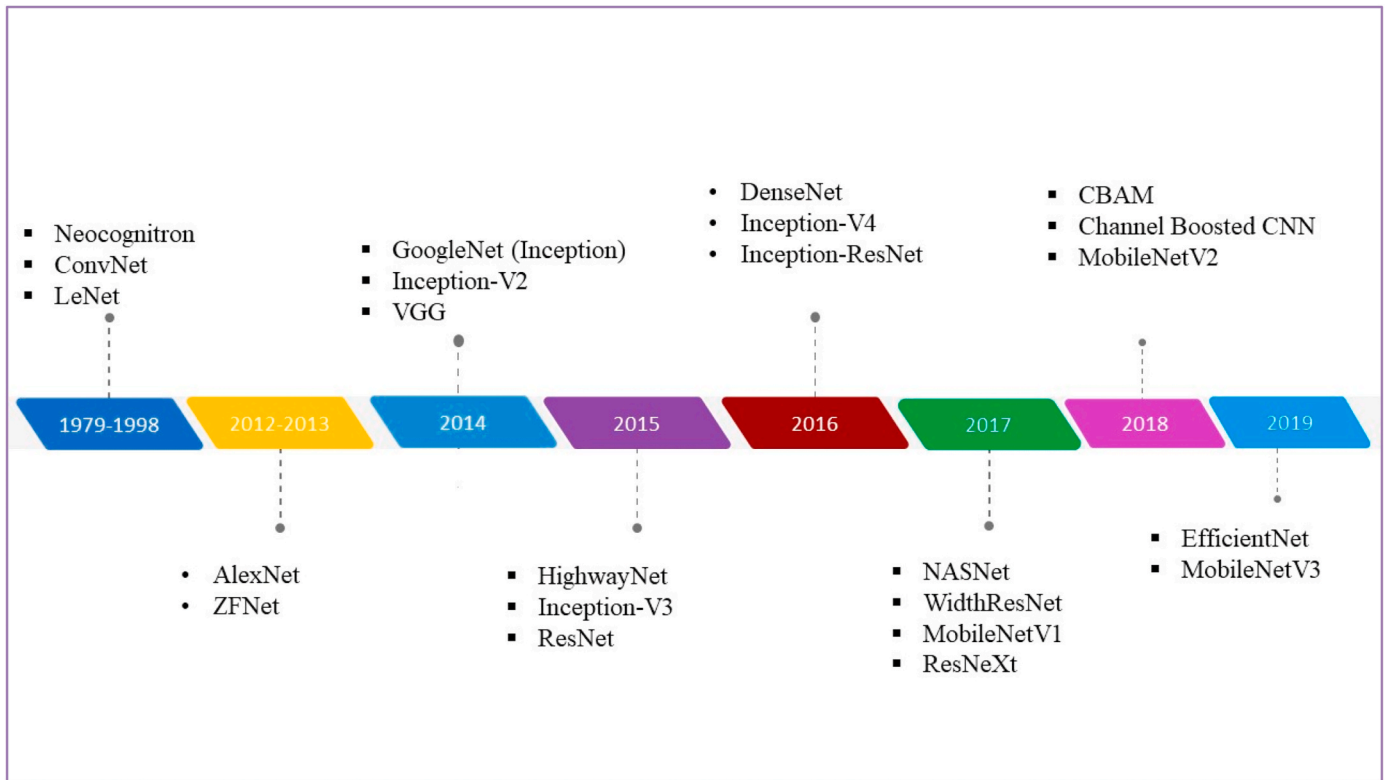


Fig. 7. Evolutionary history of deep CNN (inspired from Refs. [58–61]).

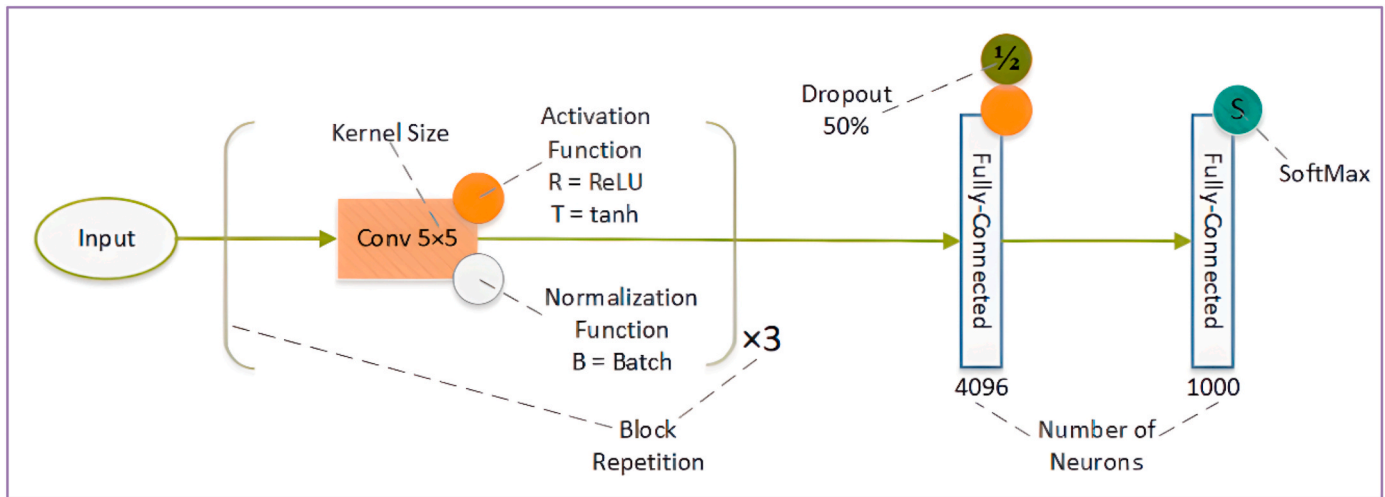


Fig. 8. Basic block diagram explanation [62].

visual representation of the EfcientNetB7 architecture is illustrated in Fig. 12.

2.2.1.5. GoogleNet. GoogleNet, proposed by Szegedy et al. [53], employs an Inception module to cut down on parameters by employing parallel convolution layers with varied settings merged through concatenation. GoogleNet, despite having 22 layers (see Fig. 13), remarkably reduces parameters to 5 million, a fraction of AlexNet's, while maintaining superior accuracy. It replaces a fully connected layer with a (7×7) average pooling layer for significant dimensionality reduction. Furthermore, the architecture utilizes parallel convolution lines with diverse filters (1×1) , (3×3) , and (5×5) , followed by concatenation, alongside (1×1) convolution layers within the Inception

module for feature map reduction and additional non-linearity [62].

2.2.2. Deep neural network (DNN)

Typically, DNN comprises an input layer, an output layer, and one or more hidden layers [64] (see Fig. 14 (a)). It is particularly adept at addressing intricate problems and discerning the connections between input and output data. DNNs excel at modeling complex non-linear relationships [44] and primarily employ supervised learning techniques. Unlike simpler models, DNNs can effectively capture the subtle changes in the brain structure that suffer from AD [65]. However, it's worth noting that DNNs demand substantial training data to extract features from labeled images effectively [66].

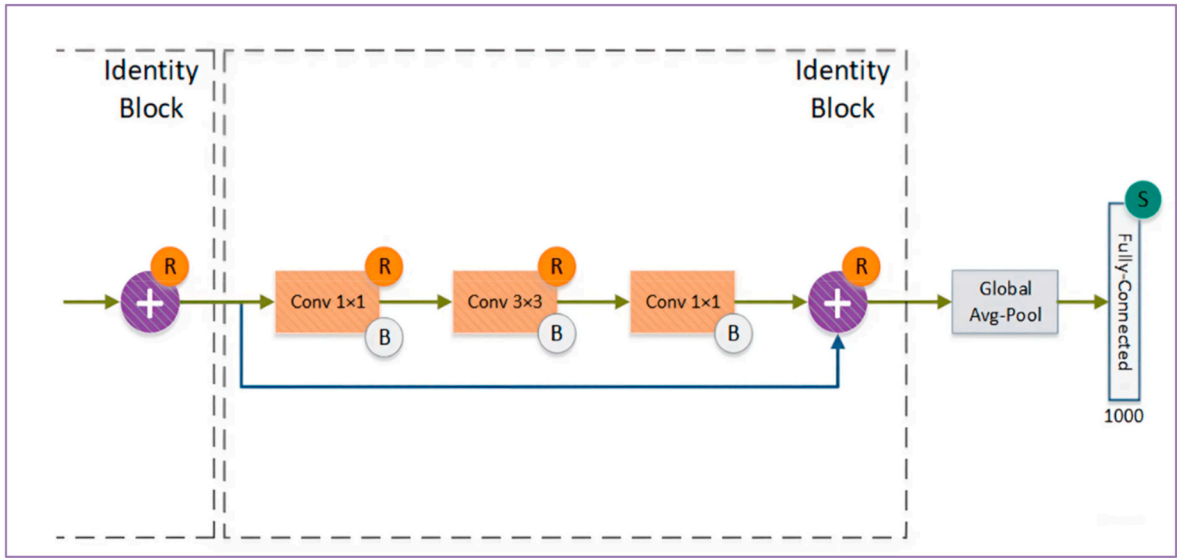


Fig. 9. The final Identity block within the architectural design of ResNet [62].

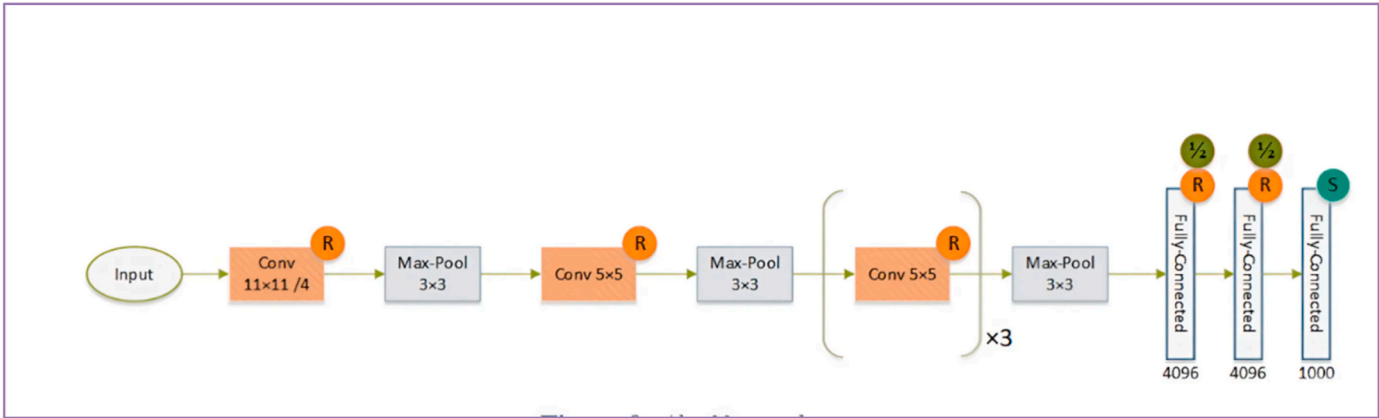


Fig. 10. Architecture of AlexNet [62].

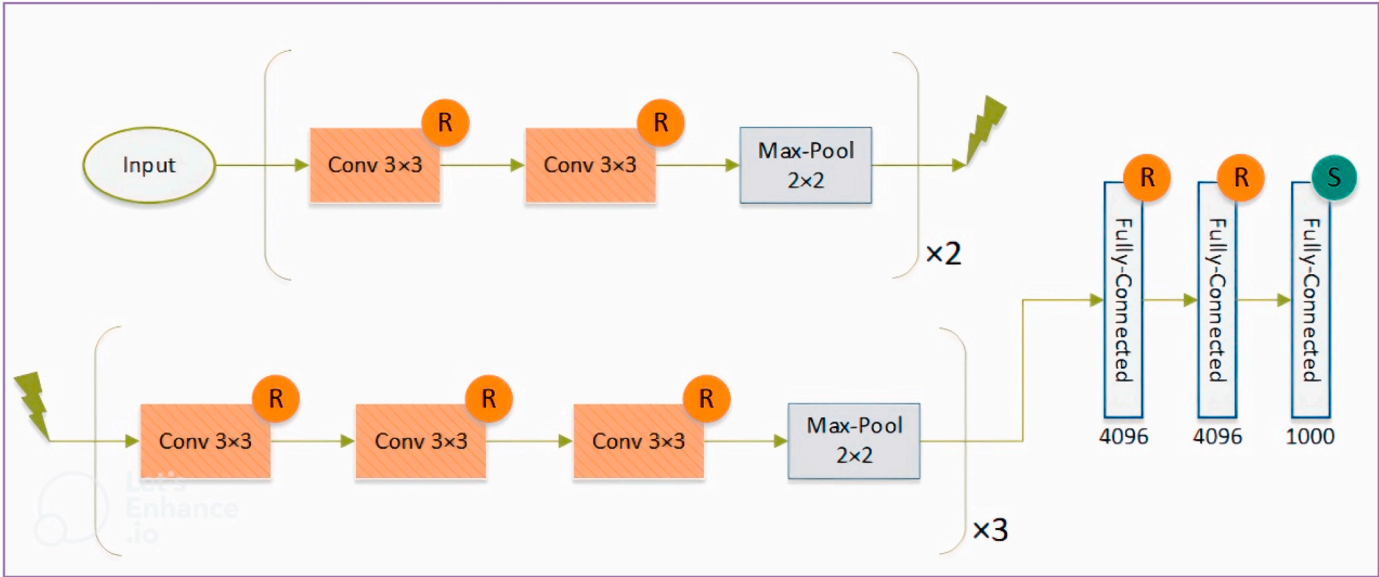


Fig. 11. Architecture of VGG [62].

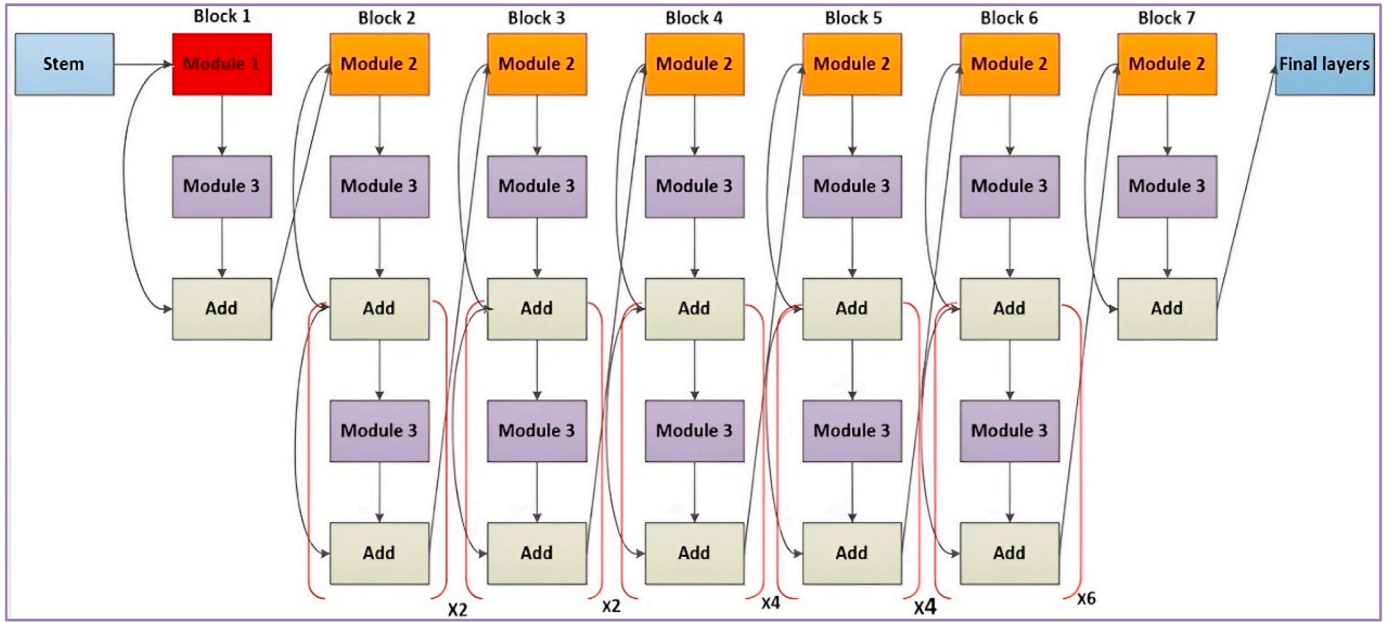


Fig. 12. Architecture of EfficientNetB7 [63].

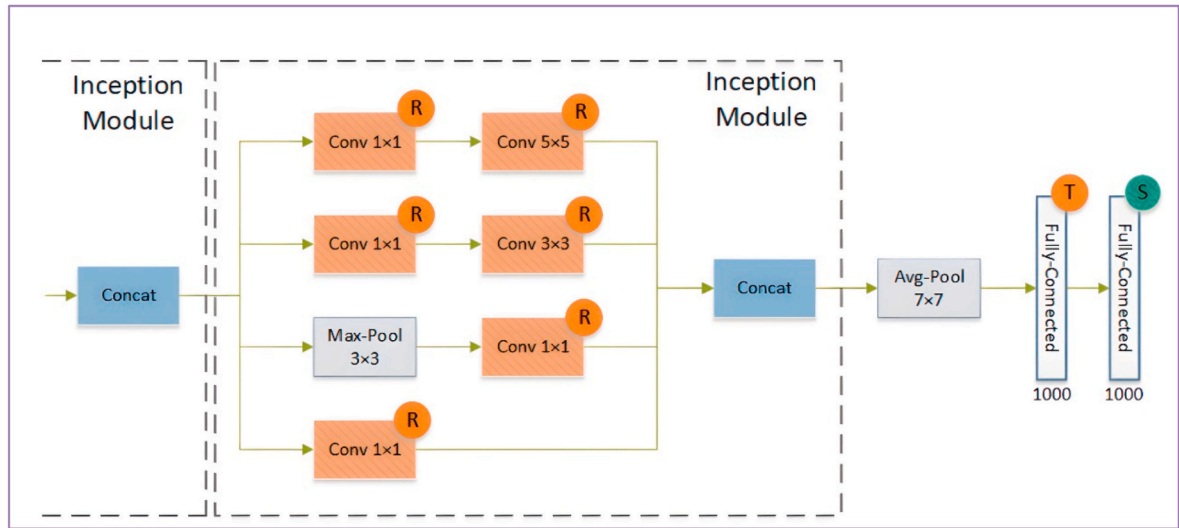


Fig. 13. GoogleNet module [62].

2.2.3. Recurrent neural network (RNN)

RNNs are primarily employed in solving sequence- or time-series-related problems [67]. One of their key advantages lies in their utilization of memory and hidden states. As shown in Fig. 14 (b), a typical RNN architecture consists of an input, hidden, and output layer. The hidden state is vital in retaining crucial information about the problem's sequential nature [70]. Unlike feedforward networks, RNNs share the same parameters within each network layer, unlike feedforward networks [71]. On the other hand, feedforward networks possess distinct parameters for each node, leading to a higher number of parameters overall [72]. RNNs, compared to CNNs and DNNs, do not comprise a large number of layers and lack significant depth [73]. Still, they pose challenges regarding training, often suffering from issues related to vanishing or exploding gradients. These limitations constrain their applicability in modeling long-term activity sequences and temporal dependencies in sensor data.

Long short-term memory (LSTM) and Gated recurrent units (GRUs) are common architectures of RNN [74]. Firstly, the primary function of

LSTM is to manage errors across various layers and time periods [75]. It incorporates cells in the hidden layer with three gates: input, output, and forget gates. These gates store information and regulate its flow to predict the network's output [72]. This single cell assists the model in determining what to store when to read, and when to update information through these gates [66]. Secondly, GRUs utilize a hidden state that consists of two gates: the reset gate and the update gate. These gates control the retention of information and the amount of information retained [72]. Finally, while GRUs offer simple alternatives to LSTM [74], LSTMs [76] often demonstrate superior performance in problems with long-term dependencies, integrating a generative adversarial imputation method, collaborative training, and adaptive model inputs to improve overall performance.

2.2.4. Autoencoders (AEs)

In the field of medical analysis for AD diagnosis, AEs are an unsupervised learning technique that transforms an unlabeled dataset into feature-encoded data. As shown in Fig. 14 (c), they are frequently used

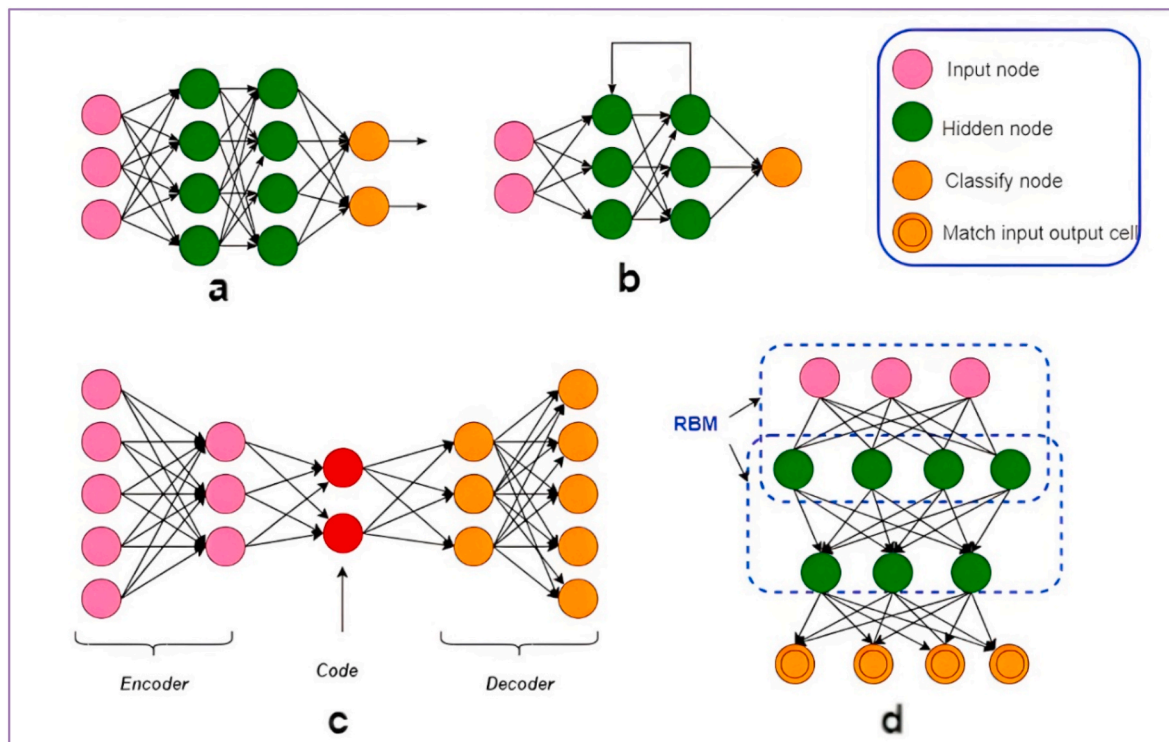


Fig. 14. a) DNN Architecture, b) RNN Architecture, c) Autoencoder, and d) Deep Belief Network (inspired from Refs. [64,67–69]).

for dimensionality reduction and consist of two primary components: an encoder and a decoder [66]. While the encoder converts input data into a code (essentially compressed data), the decoder reconstructs this code into an output that resembles the initial input [68]. Layers within the encoder can be either dense layers or convolution layers. The number of layers in the encoding section must match the number of layers in the decoding section. The encoder reduces the data's dimensions while the decoder increases them [77]. The central layer is called the bottleneck layer, responsible for compressing the representation of the input data [42].

AEs come in different types that enhance their performance: (1) denoising autoencoders, (2) sparse autoencoders, and (3) contractive autoencoders [78]. Autoencoders encounter challenges, such as merely copying input data to the hidden layer, leading to inefficient extraction of meaningful features, even though they can reproduce the input in the output layer [79]. Denoising autoencoders address this issue by corrupting the inputs that the autoencoder must subsequently reconstruct or denoise [80]. This approach assists the model in identifying features within noisy input, thus facilitating classification. With denoising autoencoders, the model doesn't simply copy input to output but rather learns meaningful data features [24]. For instance, in Ref. [81], AE is employed to compress ECG heartbeat signals and extract features to mitigate the imbalanced dataset.

2.2.5. Deep Belief Networks (DBNs)

In the context of medical image analysis for AD prediction, DBNs offer an interesting technique for feature learning. DBN is a supervised learning technique that can establish connections between unsupervised features extracted from stacked layers [69]. It operates as a generative graphical model constructed by a stack of Restricted Boltzmann Machines (RBM) that is capable of feature extraction and input reconstruction. In DBN, an undirected connection exists between the top two layers, as illustrated in Fig. 14 (d). Using RBMs, DBN offers two crucial benefits: 1) *Feature learning* (DBNs excel at learning hierarchical features from data) and 2) *Reduced overfitting* (The pre-training of RBMs helps initialize the weights in the network) [69]. DBN was designed to analyze

the apparent distribution between the input and hidden layers so that lower-layer nodes have direct connections while upper-layer nodes are indirectly connected [82]. Due to their ability to learn hierarchical features and model complex relationships, the model is particularly well-suited for tasks requiring feature extraction in different domains. This includes applications in biological data and situations where classes are not linearly separable [83].

3. Brain imaging

This section answers **RQ2**, which concerns brain imaging methods that offer non-invasive means to visualize the brain's structure, function, and pharmacological aspects [84]. The imaging techniques are typically categorized into two main groups: structural imaging and functional imaging [85]. Structural imaging furnishes insights into the brain's physical composition, encompassing elements like neurons, synapses, and glial cells. In contrast, functional imaging offers data on the brain's activities and functions [86].

3.1. Neuroimaging techniques for AD diagnosis

Neuroimaging techniques are valuable tools for diagnosing AD. Here, we explain the commonly used techniques.

3.1.1. Magnetic Resonance Imaging (MRI)

This imaging technique employs radio waves and magnetic fields to produce high-quality, high-resolution 2D and 3D images of brain structures. It does not generate harmful radiation, such as X-rays or radioactive tracers. The structural MRI (sMRI) is the most frequently utilized MRI for AD cases (see Fig. 15). It measures brain volumes in vivo, enabling the detection of brain degeneration, which involves the loss of tissue, cells, neurons, and other components—an inevitable progressive aspect of AD [87,88].

Normally, MRI employs various sequences to capture images of different body tissues, with T1 and T2 being the predominant ones [89]. Tesla (1.5T) and 3 T (3T) are MRI scanners with different field strengths.

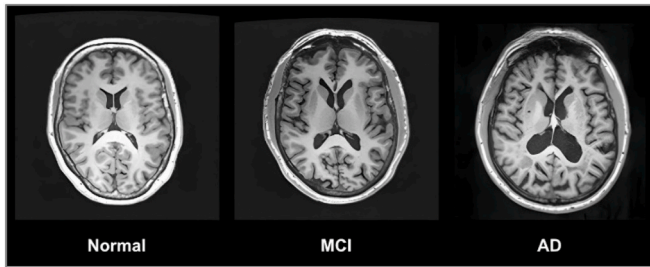


Fig. 15. Real sMRI images of persons with normal MIC and AD [92].

The major difference between them is the toughness of the magnetic field. A 3T scanner boasts a stronger magnetic field compared to 1.5T. This translates to several pros. The advantages include a high signal-to-noise ratio, improved spatial resolution, and enhanced temporal resolution [90,91].

While MRI excels at measuring brain volume, functional MRI (fMRI) measures brain activity. More specifically, fMRI has been suggested to enable the access of the human primary visual cortex and mapping of the brain topography [93]. It offers valuable data and insights into the functioning and activities of the human brain. A few illustrations of fMRI methods encompass brain imaging employing spin-labeling and utilizing Blood Oxygenation Level Dependent (BOLD) signals (see Fig. 16). These techniques exhibit a high sensitivity to the metabolic rate of the brain, encompassing factors such as cerebral blood flow and oxygen consumption [38].

3.1.2. Positron emission tomography (PET)

PET employs radiotracers, such as fluorodeoxyglucose, to access brain glucose metabolism and amyloid tracers to identify any amyloid plaques. It can specifically assess functions such as working, remembering, thinking, listening, and visual observation [95,96] (See Fig. 17). In other words, PETs provide insights into how well these functions are supported by underlying brain activity.

3.1.3. Computed tomography (CT)

CT scans are a medical imaging technique that uses X-rays to produce detailed images of the body's internal structures [97] (see Fig. 18). It helps rule out other brain conditions but is less sensitive than other imaging methods for early AD diagnosis [98].

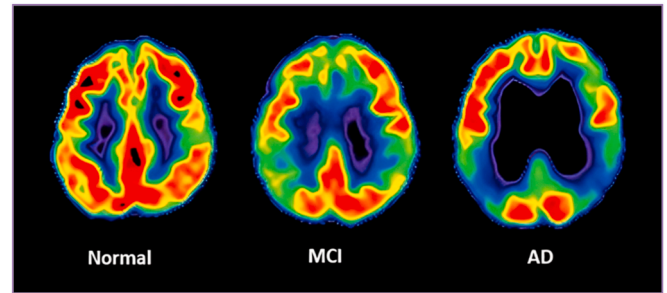


Fig. 17. Real PET images of persons with normal, MCI and AD [26].

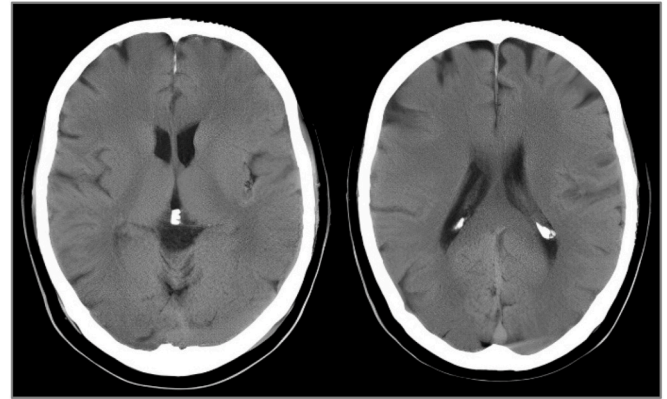


Fig. 18. Real CT images of persons with AD [99].

3.1.4. Diffusion tensor imaging (DTI)

As shown in Fig. 19, DTI is a specialized MRI technique that plays a valuable role in identifying the abnormal diffusion patterns associated with AD. DTI analyzes the movement of water at the microstructural level within the brain to examine white matter integrity and connectivity disruptions in AD [100]. Thus, DTI can potentially aid in diagnosing AD by detecting the disruptions through abnormal water diffusion patterns.

3.1.5. Single photon emission computed tomography (SPECT)

SPECT utilizes a gamma-ray-emitting radioactive tracer introduced into the bloodstream. It gauges the dispersion of this tracer within the

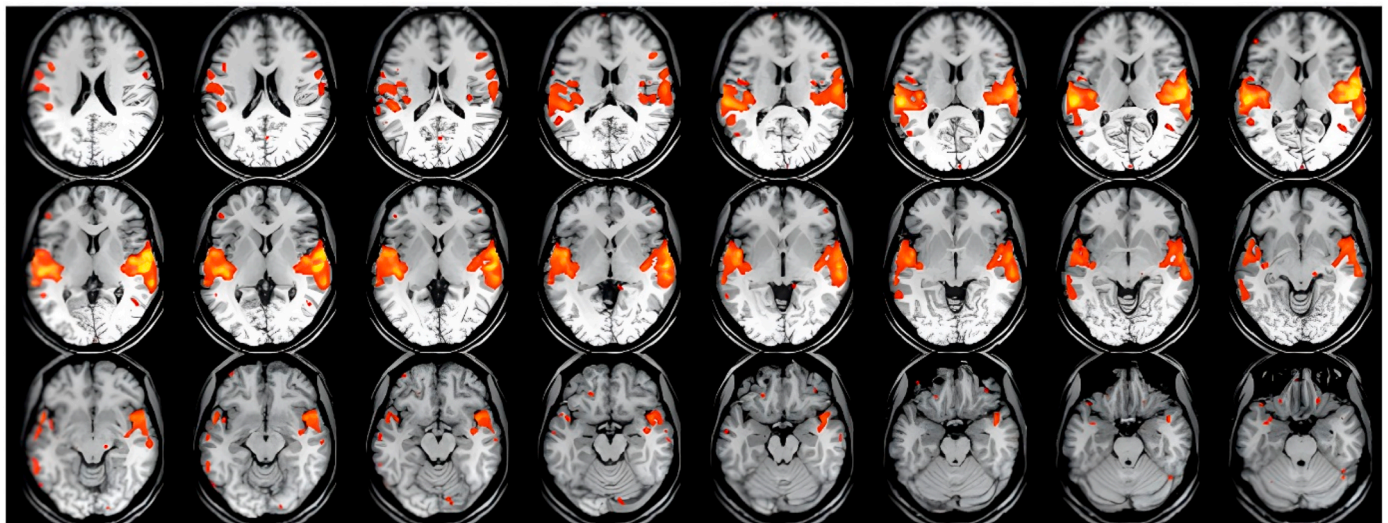


Fig. 16. Example of fMRI [94].

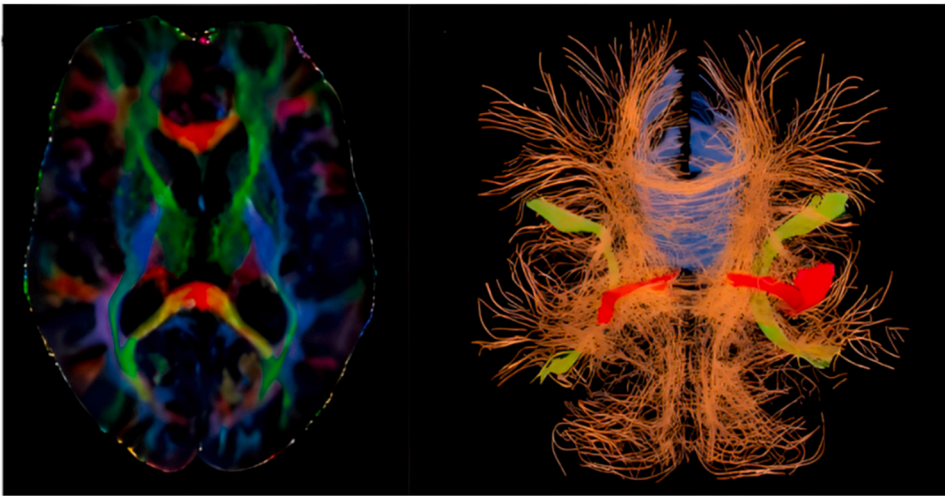


Fig. 19. Real DTI images of persons with and without AD (the right one is normal) [101].

brain, offering details about regional cerebral blood flow and metabolic activity. Reflecting its potential for earlier diagnosis, DTI has been included as a potential biomarker in recently updated clinical criteria for AD. DTI has a role in not only identifying AD in its early stages but also predicting its progression in individuals diagnosed with MCI [102] (see Fig. 20).

3.2. MRI and biomarkers of AD detection

Biomarkers are unequivocally regarded as medical indicators, representing the perceptible outward indications of patients' medical circumstances or states. These biomarkers can be quantified with a notable level of exactitude, enabling meticulous evaluation and surveillance of

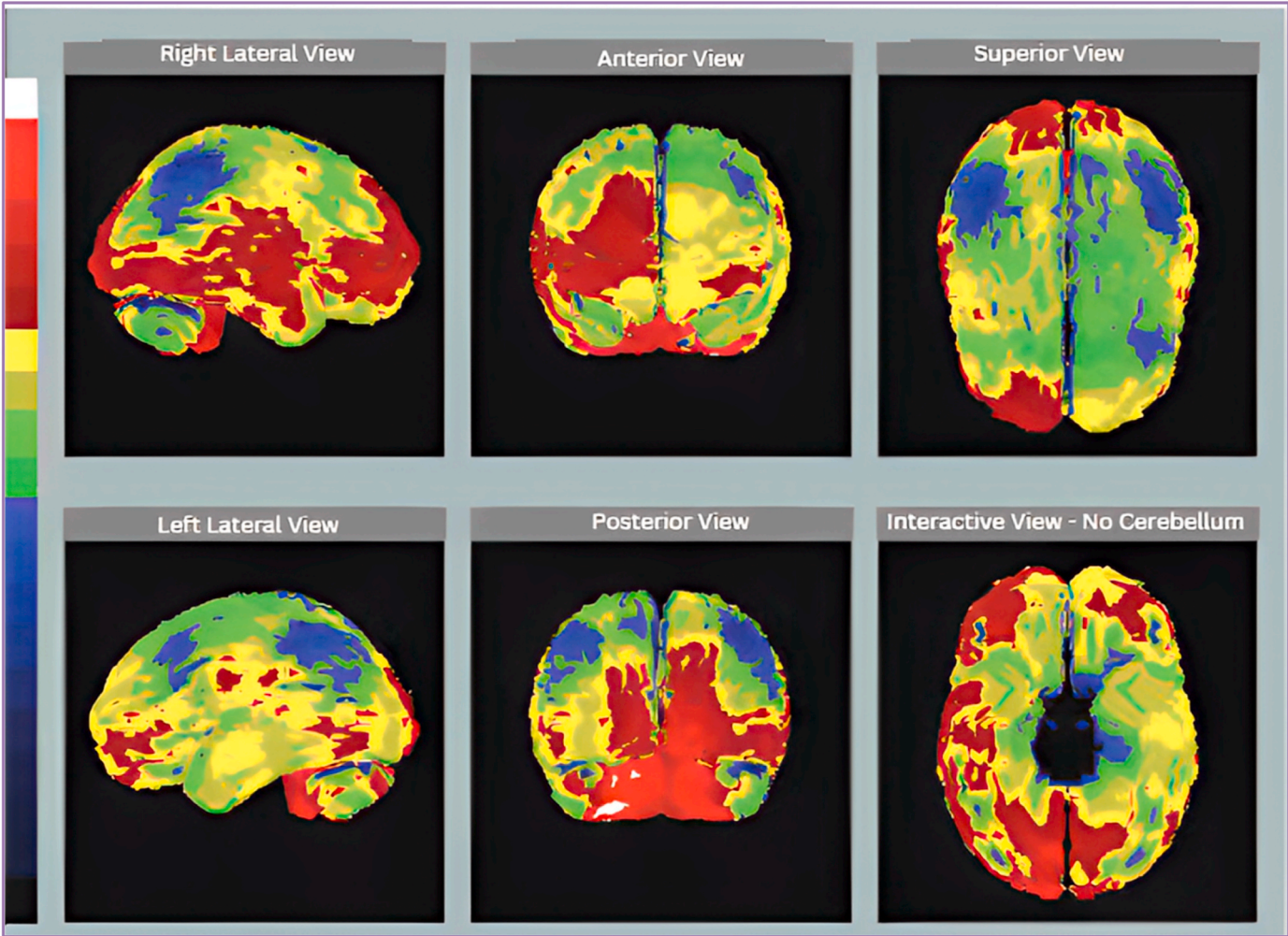


Fig. 20. Real SPECT images of persons with AD [102].

diverse medical conditions [103]. Biomarker characterization can vary depending on the circumstances. For example, the International Program on Chemical Safety defines biomarkers as measurable entities, configurations, or methodologies within the organism that indicate the presence of AD [104]. In the context of AD, biomarkers possess specific properties [19].

1. They can identify fundamental characteristics of AD's neuropathology.
2. They can confirm cases of AD based on neuropathological evidence.
3. They are efficient, meaning they can identify the early stages of AD and differentiate AD from other types of dementia.
4. They are reliable, non-invasive, easy to implement, and cost-effective.

Three categories of biomarkers could potentially provide further insight into AD: Genetic biomarkers, biochemical biomarkers, and neuroimaging biomarkers [105].

4. Popular AD datasets

To address **RQ3**, which concerns the importance of AD datasets, AD research relies on comprehensive encompassing diverse data types. The datasets, collected longitudinally, are essential resources for advancing our understanding of AD progression across various stages [106]. Collaborative efforts from various sectors contribute to a data-driven quest for understanding and combatting AD. This section explains the most popular utilized AD datasets.

4.1. Alzheimer's Disease Neuroimaging Initiative (ADNI)

ADNI is a longitudinal multicenter study. Its primary goal is to develop clinical, imaging, genetic, and biochemical biomarkers for the early detection and continuous monitoring of AD. Launched over a decade ago, this significant public-private collaboration has played a pivotal role in advancing AD research. It has facilitated the global sharing of data among researchers, making substantial contributions to understanding AD [107,108]. Here, we explain the development phases of ADNI datasets (see Fig. 21).

4.1.1. ADNI-1

ADNI-1 commenced in 2004 and lasted for five years. Its primary goal was to identify more precise biomarkers to detect early-stage AD and track the disease's progression. The study included a cohort of 200 elderly controls, 400 individuals with MCI, and 200 subjects with AD. ADNI-1 collected extensive data, including brain scans, genetic profiles, and blood and cerebrospinal fluid (CSF) biomarkers, for analysis. Imaging modalities, such as structural MRI and PET scans, were employed [107].

4.1.2. ADNI-grand opportunities (ADNI-GO)

ADNI-GO was initiated in 2009 and spanned two years. Its primary focus was on the early detection of AD. In conjunction with ADNI-1, ADNI-GO introduced 200 new subjects with EMCI. This phase meticulously examined biomarkers at pre-AD stages and adjusted MRI protocols accordingly.

4.1.3. ADNI-2

ADNI-2 was established in 2011 and ran for five years. Its primary aim was to discover biomarkers capable of predicting and assessing cognitive impairment. Building on the foundation of ADNI-1 and ADNI-GO, this phase included 150 elderly controls, 100 individuals with EMCI, 150 with LMCI, and 150 subjects with AD. To better distinguish between healthy controls and those with MCI, a new cohort known as Significant Memory Concern (SMC) was introduced (107 SMC participants have been added). Furthermore, ADNI-2 significantly contributed by incorporating amyloid PET with Florbetapir at all ADNI-2 sites and for all ADNI-2 and ADNI-GO participants [107].

4.1.4. ADNI-3

ADNI-3 launched in 2016, focusing on exploring the interrelationships among clinical, cognitive, imaging, genetic, and biochemical biomarker characteristics of AD. ADNI-3 introduced the identification of tau protein tangles (tau PET) in brain scans. Building on the prior phases, ADNI-3 continued to refine, validate, and authenticate test measures and biomarkers used in AD diagnosis. ADNI-3 expanded its reach by involving 59 research centers and recruited 133 healthy controls, 151 individuals with MCI, and 87 participants with AD [107].

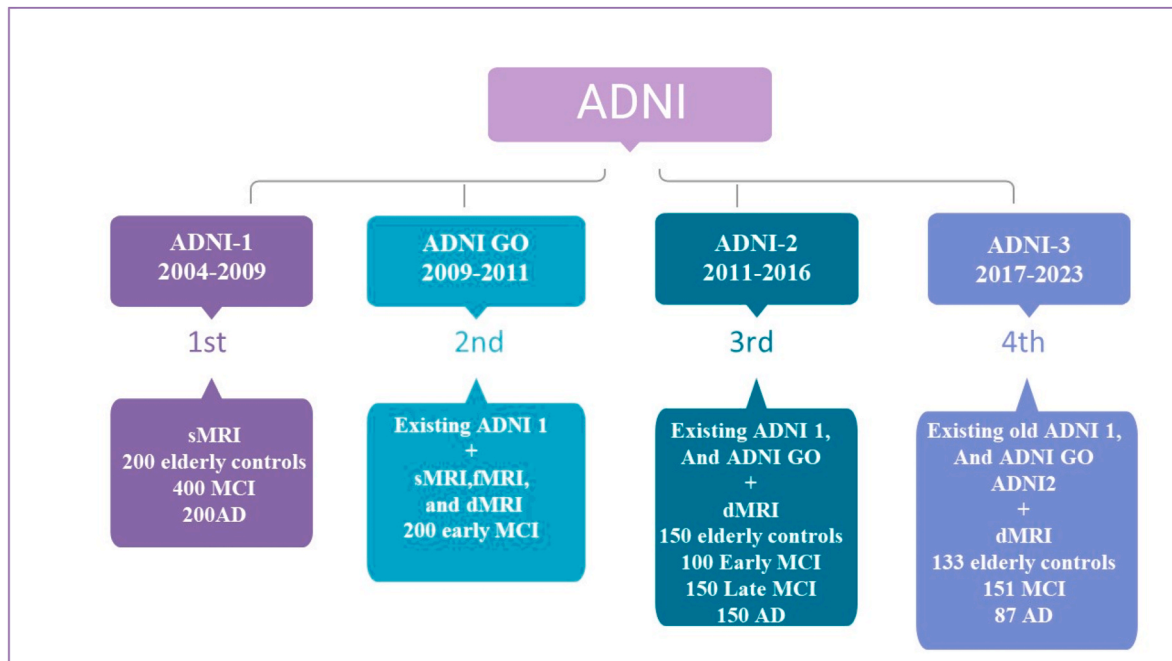


Fig. 21. Summary of ADNI dataset classifications (inspired from Ref. [107]).

4.2. Open Access Series of Imaging Studies (OASIS)

OASIS is a valuable initiative that promotes open science in neuroscience research. OASIS collects and freely shares brain MRI datasets with the scientific community, fostering advancements in both basic and clinical neuroscience research areas [109]. The OASIS dataset is categorized into five stages [110].

1. OASIS 1.
2. OASIS 2.
3. OASIS-3.
4. OASIS-3 TAU.
5. OASIS-4.

Fig. 22 summarizes the classification attributes of the participants within the OASIS datasets. It's important to note that certain attributes may have different settings in various OASIS phases. For instance, in OASIS 1, the education level (EDUC) is categorized on a scale of 1–5, where 1 indicates below high school, and 5 represents post-graduate education. However, subsequent OASIS (OASIS 2 onwards) represent education level in years of education [111].

4.3. MILAN

In 1984, the MILAN dataset represents a revised version of the criteria initially established by the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association. This updated version aims to be adaptable for use by both general healthcare providers and researchers. It provides criteria for diagnosing all-cause dementia (such as AD). The fundamental framework for diagnosing probable AD remains largely unchanged. Nevertheless, building on the knowledge accumulated over the past twenty-seven years, several modifications have been introduced to the clinical criteria for diagnosis. The term "possible AD dementia" is retained but is now defined in a more focused manner than before. Additionally, biomarker evidence, primarily used in research settings, has been incorporated into the diagnostic criteria for probable and possible AD. While the core clinical criteria for AD continue to serve as the foundation for diagnosis in clinical practice, including biomarker evidence enhances the pathophysiological specificity of AD diagnosis [112].

4.4. The national Alzheimer's coordinating center (NACC)

NACC, established in 1999, originally offered standardized data to

approximately 32 Alzheimer's Disease Centers (ADCs). Over time, NACC has evolved to provide longitudinal and standardized data related to neuroimaging and biofluids. This dataset encompasses various imaging modalities, including MRI, PET scans, and CSF measurements [113]. NACC boasts one of the world's largest AD datasets. The dataset encompasses data from over 47,000 participants. It includes the following data: 1) Over 16,000 currently participating persons, 2) More than 174,000 clinical assessments, and 3) Over 7340 neuropathology exams from 58 % of deceased participants. Over the past decades, the dataset has been built in collaboration with more than 42 ADCs throughout the US and is freely available to researchers [114].

4.5. The australian imaging biomarkers and lifestyle study of ageing (AIBL)

The AIBL dataset is a subset of the ADNI dataset and contains data from subjects with PET and MRI. Its data are not freely available; instead, they can be used by authorized users only. The process to obtain access is integrated with procedures used to access ADNI data and uses the Laboratory of Neuroimaging (LONI) infrastructure [115]. Participants experience follow-up assessments every 18 months, including an extensive evaluation of cognitive functions, neuroimaging modalities like MRI and PET, analysis of biofluid biomarkers, and evaluations of lifestyle factors [116].

4.6. The gwangju Alzheimer's and related dementias (GRAD)

GARD Cohort is a group of individuals living in a senior citizen's community in Gwangju, South Korea. The dataset was collected from the National Research Center for Dementia (NRCD), South Korea, from January 2014 to March 2018. The dataset consists of 326 MRI scans with four classes: (1) 171 individuals with NC, (2) 81 persons with AD, (3) 35 participants with asymptotic Alzheimer's disease (aAD), and (4) 39 cases with mild Alzheimer's disease (mAD). The subjects' average age range is 70.018 ± 6.074 years old [117]. While some have also undergone PET-amyloid imaging, many participants in this cohort have undergone assessments using the Seoul Neuropsychological Screening Battery (SNSB) and received brain MRI scans. Previous research involving this cohort includes studies on the genetics of AD [118] and brain MRI traits [119].

4.7. Minimal interval resonance imaging in Alzheimer's disease

The dataset contains longitudinal volumetric T1 MRI scans from 46 mild-moderate AD subjects and 23 controls. It includes 708 baseline

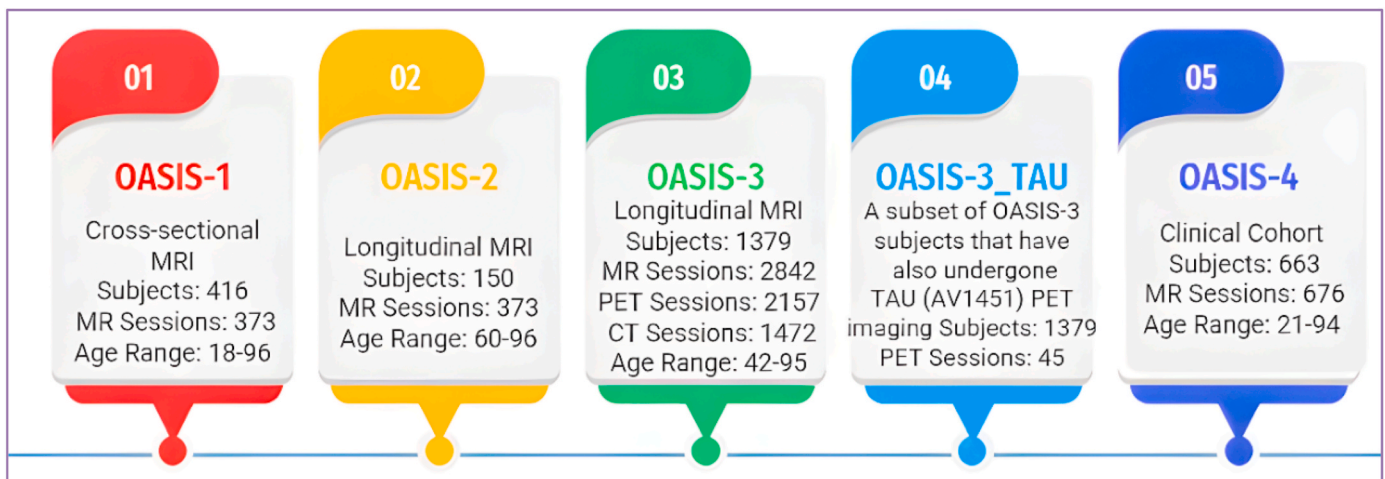


Fig. 22. Summary of OASIS classification (inspired from Ref. [111]).

scans and additional information, such as gender, age, and MMSE score. These details and results are publicly available, serving as a resource for researchers to develop, validate, and compare techniques, especially for measuring longitudinal volume changes in MRI [120].

4.8. Kaggle repository

Kaggle is a popular platform for data science and machine learning, containing over 50,000 publicly available datasets. It serves as a valuable resource for scholars investigating dementia and AD. Examples of Kaggle AD-related datasets are as follows [15].

1. ADNI contains MRI data from initial screening or baseline, including ADNI 1,2/GO and 3.
2. Alzheimer MRI Preprocessed Dataset, which consists of preprocessed MRI images resized to 128 x 128 pixels.
3. AD Dataset (4-class Images): Contains four classes of images related to AD.

5. Pre-processing of data in AD studies

This Section addresses RQ4 regarding pre-processing techniques that are employed to clean and reliable AD data. The techniques can enhance the quality and accuracy of the analysis results. Here, we briefly explain the most frequently employed pre-processing techniques are.

5.1. Data cleaning

Missing data points (arise due to technical issues or participant dropout) require appropriate methods. Therefore, the cleaning process entails removing or correcting incomplete, inconsistent, or erroneous data within the dataset [121]. Typically, it involves identifying missing values, outliers, and inconsistencies and then either removing them or replacing them with accurate data [122].

5.2. Feature selection

The feature selection process involves selecting the most pertinent features from the dataset, likely influencing the analysis results significantly. Various statistical techniques can accomplish this task, such as correlation analysis, principal component analysis, and mutual information [123].

5.3. Data transformation

This process aims to convert the data into a format that facilitates the analysis. The transformation can be achieved through techniques like normalization, standardization, and logarithmic transformation [124].

5.4. Imputation

This process includes filling in missing data by estimating values. In the literature, various imputation techniques were proposed, such as mean imputation, regression imputation, and multiple imputations [26].

5.5. Brain Extraction and skull stripping

This preprocessing technique can eliminate non-brain tissues, including elements like eyes, necks, and skulls. As shown in Fig. 23, the structures are segmented based on the dark space between the skull and brain, which is occupied by the CSF [24].

5.6. Image pre-processing methods

The image dataset for AD typically comprises MRI and PET. Various pre-processing techniques were applied to these images in the research articles reviewed for this study. Commonly employed pre-processing methods in AD research articles include the following.

- 1) **Image Normalization:** It is the procedure of aligning the intensity values of images to a uniform scale. The adjustment can be achieved through methods like histogram equalization, contrast stretching, and normalization to enhance the visual quality of the images [126].
- 2) **Image Cropping and Resizing:** This process includes resizing images to a consistent size and removing any extraneous parts of the image that do not contain pertinent information. The action can assist in lowering the computational complexity of the analysis and enhance the algorithms' efficiency [24].
- 3) **Image Augmentation:** It aims to create additional images from the original dataset by implementing various transformations, including flipping, rotating, and scaling. This approach can be beneficial in augmenting the dataset's size and enhancing the robustness of the analysis [127].
- 4) **Image Segmentation:** It is a fundamental process in medical imaging, dividing images into distinct regions based on visual characteristics. This technique is essential for various medical image

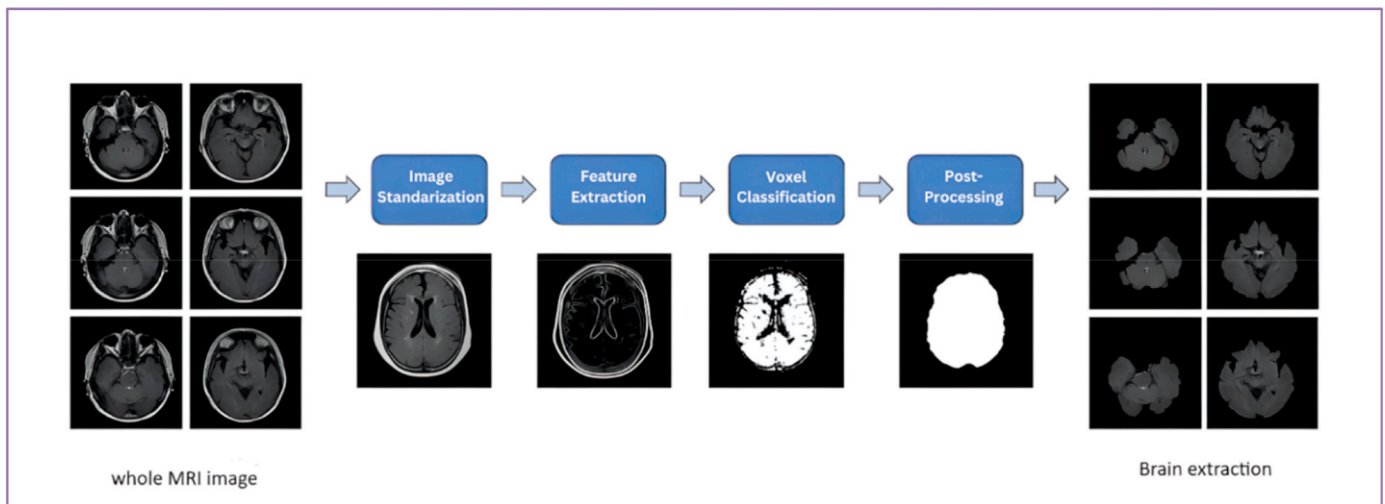


Fig. 23. Brain extraction with a binary segmentation mask [125].

analysis tasks, enabling the identification of structures or anomalies. Common segmentation methods include thresholding, edge detection, and region-based approaches, which are vital in medical diagnosis and research [128].

6. Performance evaluation measures

In this Section, we address **RQ5** on how models are evaluated. The developers assess the performance of their model using a confusion matrix. The matrix is an $N \times N$ grid, where N denotes the number of classes present in the dataset [111]. In binary classification ($N = 2$), the outcomes of the confusion matrix represent positive and negative outcomes. The evaluation metrics are derived from the confusion matrix. The parameters of the matrix are as follows.

1. True positive (T_P).
2. True negative (T_N).
3. False positive (F_P).
4. False negative (F_N).

While T_P and T_N refer to the person being diagnosed correctly, F_P and F_N mean misdiagnosed. Accordingly, the confusion matrix is formed (see Fig. 24). As shown in Fig. 25, there are several evaluation metrics. Here, we introduce the reader to the most popular metrics implemented in the AD context [26].

6.1. Accuracy

Accuracy evaluates the correctness level of the model's predictions, particularly in AD diagnosis. It calculates the number of MRIs accurately classified among the total MRI inputs [111].

$$Accuracy = \frac{T_P + T_N}{T_P + F_P + T_N + F_N} \quad (1)$$

6.2. Sensitivity

Sensitivity (also called Recall) is a metric used in binary classification to assess how well a model identifies T_P (correctly labeled positive AD records concerning all individuals who are suffering from AD) [26].

$$Sensitivity = \frac{T_P}{T_P + F_N} \quad (2)$$

6.3. Specificity

Specificity corresponds to the proportion of correctly labeled negative AD records in compare to all individuals who are genuinely CN [129].

$$Specificity = \frac{T_N}{T_N + F_P} \quad (3)$$

6.4. Precision

Precision is defined as the ratio of correctly labeled positive AD records by the model to the entirety of positive records labeled by the model [123].

$$Precision = \frac{T_P}{T_P + F_P} * 100 \quad (4)$$

6.5. Negative predictive value (NPV)

It is the opposite of Precision; it quantifies the proportion of T_N (negatives) among all the cases classified as negative, whether correctly or incorrectly. It addresses the question: How many of those labeled as not having AD truly do not have AD [26]?

$$NPV = \frac{T_N}{T_N + F_N} \quad (5)$$

6.6. F-measure (also known as F-score or F1-score)

The F-measure score is the harmonic mean of accuracy and recall. When precision and recall scores are perfect, it results in the highest F-score, which equals 1 [105].

$$F - measure = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (6)$$

6.7. The receiver operating characteristic (ROC) curve

Assessing a classifier's ability to discriminate within a sample is straightforward at a specific threshold [26]. The ROC curve effectively illustrates the relationship between the specificity and sensitivity of a particular analytical method. It comprehensively portrays the diagnostic method's accuracy [24]. The ROC curve assesses the overall effectiveness of a test across various threshold values. A larger area under the

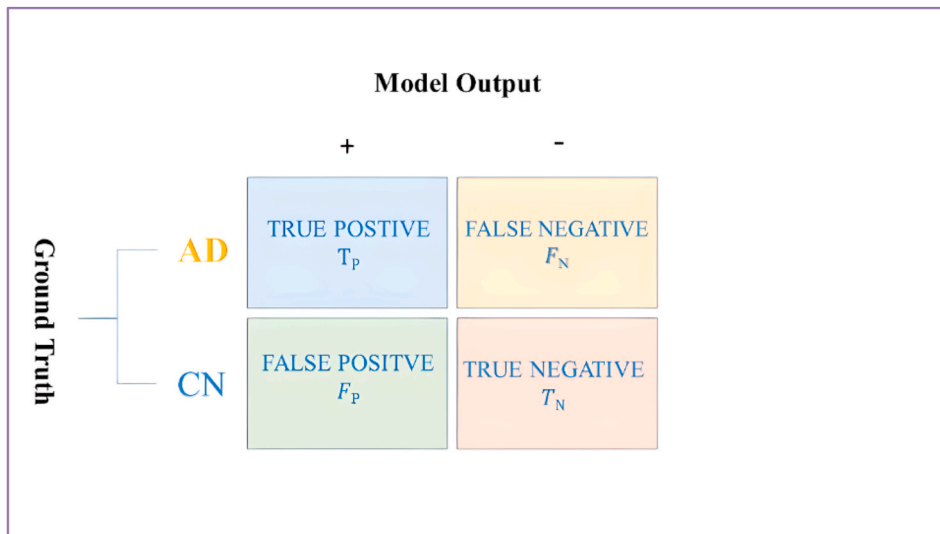


Fig. 24. Confusion matrix for AD prediction (inspired from Ref. [26]).



Fig. 25. Evaluation metrics.

curve indicates a more valuable test, and its values range from 0 % to 100 % [123].

6.8. Area under the curve (AUC)

The AUC quantifies the area under the T_p versus F_p rates. It assesses the model's capacity to differentiate between classes and serves as a concise summary of the receiver characteristics curve [111].

6.9. Macro-average

Macro-average is calculated separately for each performance measure [130].

A. The Macro-average of Precision is computed by averaging the precision values calculated for each class.

$$\text{Macro - average of Precision} = \frac{\sum_{i=1}^n p_i}{n} \quad (7)$$

In this formula, "n" represents the count of classes in the dataset, and " $p_1, p_2, \dots, p_n$ " represents the Precision values calculated for each class.

B. The Macro-average of Recall is calculated by taking the average of all Recall values calculated for each class.

$$\text{Macro - average of Recall} = \frac{\sum_{i=1}^n R_i}{n} \quad (8)$$

where n is the number of classes in the dataset and " $R_1, R_2, \dots, R_n$ " are Recalls calculated for each class.

C. The Macro-average F1 Score is calculated as the harmonic mean of the Macro-average Recall and Precision.

6.10. Micro-average

Micro-average is computed individually for each performance measure as follows [130].

A. The micro-average of Precision is calculated by dividing the sum of T_p by the sum of both T_p and F_p as follows:

$$\text{Micro - average of Precision} = \frac{\sum_{i=1}^n TP_i}{\sum_{i=1}^n TP_i + \sum_{i=1}^n FP_i} \quad (9)$$

where n is the count of classes in the dataset.

B. The micro-average of Recall is computed by dividing the sum of T_p by the sum of T_p and F_n as follows:

$$\text{Micro - average of Recall} = \frac{\sum_{i=1}^n TP_i}{\sum_{i=1}^n TP_i + \sum_{i=1}^n FN_i} \quad (10)$$

C. The Micro-average of F1 Score is calculated as the harmonic mean of the Micro-average of Recall and Precision.

To recap, unlike existing surveys, the current survey introduced all performance evaluation metrics employed in AD literature (see Table 2).

7. Literature review

7.1. Selection criteria

Criteria for exclusion and inclusion ensure that data analysis comprises entirely appropriate research. In this survey, a three-stage procedure has been employed for paper selection. The procedure comprises the following.

1. Paper selection is determined by the title.
2. Reading the abstracts constitutes the following step.
3. The final phase consists of thoroughly reading the documents.

7.2. Article selection

The search was carried out as follows: 'Alzheimer's Disease' AND ('AD stages' OR 'early detection' OR 'Diagnoses' OR 'Classification' OR 'Classifier' OR using 'Deep learning' OR networks OR 'CNN' OR 'pre-trained models' OR 'transfer learning' OR 'Dataset' OR 'ADNI' OR 'OASIS'). To ensure a focused analysis, we carefully selected Thirty-nine papers published between 2017 and 2023 for in-depth review. Table 3 and Fig. 26 summarize the scientific papers considered to be within the scope of our section.

7.3. Analysis of the selected papers

A literature review critically analyzes existing papers. There are various approaches to conducting a literature review, including chronological, thematic, and methodological. This study adopted the

Table 2

Current study vs. the previous surveys in describing evaluation metrics.

Review paper	Publisher	Journal	No. of metrics	Metrics
[19]	IEEE	IEEE Access	0	–
[20]	Science Direct	Materials Today: processing	0	–
[21]	IEEE	4th Biennial International Conference on Nascent Technologies in Engineering (ICNTE)	0	–
[15]	IEEE	IEEE Access	0	–
[22]	Science Direct	International Journal of Cognitive Computing in Engineering	9	Sensitivity, Specificity, Precision, Accuracy, MCC, F1-score, FMI, ROC, and AUC.
[23]	Springer	The Egyptian Journal of Neurology, Psychiatry and Neurosurgery	0	–
[24]	Springer	Multimedia Tools and Applications	0	–
[25]	Springer	Archives of Computational Methods in Engineering	3	Accuracy, ROC, and AUC.
[26]	Springer	Brain Informatics	6	Accuracy, Sensitivity, Specificity, Precision, AUC, and ROC.
Current Study			11	Accuracy, Sensitivity, Specificity, Precision, F1-score, ROC, AUC, NPV, Macro-average, Micro-average, and Weighted-average.

chronological approach, focusing on studies published between 2017 and 2023 (mentioned earlier) to gain insights into the evolving trends of the AD diagnosis models.

In [131], the authors employed ResNet models and GoogleLeNet using the ADNI dataset. They implemented a four-way classifier to categorize AD, MCI, LMCI, and CN. The results demonstrated that GoogLeNet outperformed ResNet and achieved a prediction accuracy of 98.8 %. However, the model limitation resided in the absence of the incorporation of clinical data and multimodal data management, which restricted the system's robustness. In 2018, the authors in Ref. [132] introduced cascaded CNNs for MRI and PET brain image analysis without requiring specialized expertise or image segmentation. The author utilized the ADNI dataset covering 100 NC, 93 AD, and 204 MCI cases. While the model achieved 95.68 % accuracy in multi-modality analysis, it required a substantial training dataset, posing challenges for practical implementation. Additionally, the absence of clinical data integration and multimodal data handling limited the system's robustness.

Using an ensemble Deep CNN (DCNN) system, five approaches were introduced with MRI scans from OASIS [133]. The article reported that strong gradient flow in the training data improved performance when using inception-v4 and ResNet. Conversely, the E1 (M1 + M2 + M3 + M4) and E2 (M1 + M2 + M3 + M4) approaches yielded poor performance. The ensemble of M1, M2, and M3 (the proposed model) achieved an accuracy of 93.18 % with only 416 sMRI data points. Nevertheless, the depth of these neural networks required a substantial dataset for proper training. Based on a solitary cross-sectional brain sMRI scan, a deep learning algorithm was developed in Ref. [134]. The purpose of

Table 3

The sources of the considered papers.

Ref	Year	Publication	Publisher
[131]	2017	2017 IEEE International Conference on Imaging Systems and Techniques (IST)	IEEE
[132]	2018	Neuroinformatics	Springer
[133]	2018	Brain Informatics	Springer
[134]	2019	NeuroImage: Clinical	Science Direct
[135]	2019	IEEE Access	IEEE
[136]	2019	Brain Science	MDPI
[137]	2019	NeuroImage	Science Direct
[138]	2019	IEEE Access	IEEE
[139]	2020	Brain Science	MDPI
[140]	2020	Advances in Electrical and Computer Engineering	Ștefan cel Mare University of Suceava
[141]	2020	Journal of Neuroscience Methods	Science Direct
[142]	2020	IEEE Access	IEEE
[143]	2020	NeuroImage	Science Direct
[144]	2021	Proceedings of the Fourth International Conference on Microelectronics, Computing and Communication Systems	Springer
[145]	2021	Multimedia Tools and Applications	Springer
[146]	2021	IEEE Access	IEEE
[147]	2021	Diagnostics	MDPI
[148]	2021	Neuroscience	Science Direct
[149]	2021	IEEE Access	IEEE
[150]	2021	IEEE Access	IEEE
[151]	2021	Neural Computing and Applications	Springer
[106]	2021	Sensors	MDPI
[152]	2021	Journal of Physics: Conference Series	IOPScience
[153]	2021	Cognitive Computation	Springer
[154]	2022	Measurement: Sensors	Science Direct
[155]	2022	Sensors	MDPI
[156]	2022	Neuroscience Informatics	Science Direct
[157]	2022	Arabian Journal for Science and Engineering	Springer
[158]	2022	Journal of Big Data	Springer Open
[62]	2022	Arxiv preprint	Cornell university
[159]	2022	Computers Materials & Continua	Tech Science Press
[160]	2022	IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS)	IEEE
[161]	2022	International Conference on Emerging Smart Computing and Informatics (ESCI)	IEEE
[162]	2023	IEEE Access	IEEE
[163]	2023	Alexandria Engineering Journal	Science Direct
[164]	2023	Electronics	MDPI
[165]	2023	Electronics	MDPI
[166]	2023	Mathematics	MDPI
[167]	2023	Journal of Medical and Biological Engineering	Springer
[168]	2023	Diagnostics	MDPI
[169]	2023	International Conference on Innovative Data Communication Technologies and Application (ICIDCA)	IEEE
[170]	2023	Multimedia Tools and Applications	Springer
[171]	2023	International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)	IEEE
[172]	2023	International Conference on Electrical, Computer and Communication Engineering (ECCE)	IEEE
[173]	2023	Soft Computing	Springer

forecasting the individual diagnosis of AD and MCI individuals who are likely to progress to AD (c-MCI). The algorithm attained a precision level of 99 % (ADNI) and 98 % (ADNI + Institute of Experimental Neurology (Italy)). Nevertheless, the algorithm is unable to rule out the potential development of c-MCI in individuals with s-MCI.

In [135], Lulu Yue and her team proposed a Voxel-based Hierarchical Function Extraction (VHFE) method for early AD diagnosis through a hierarchical feature extraction process, aiding in diagnosing the disease. They segmented the entire brain into 90 Regions of Interest (ROIs) using the Anatomical Automatic Labeling (AAL) template. In these ROIs, they identified informative voxels, considering their values, and organized

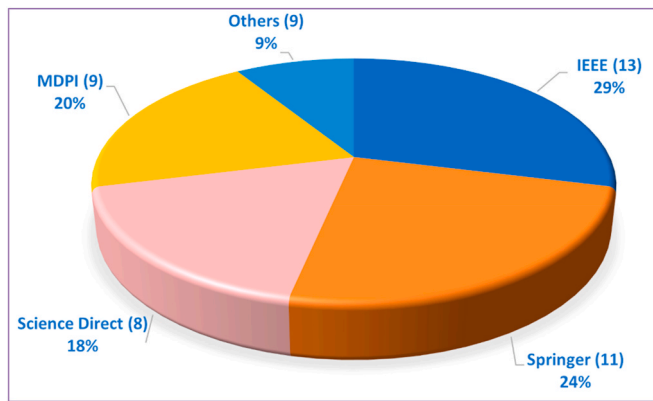


Fig. 26. Distribution of the selected papers.

them into a vector format. To refine the features, the authors applied a correlation-based approach among voxels in different groups to select the most relevant first-stage characteristics. These first-stage features were then used in a CNN to extract hidden characteristics from brain feature maps obtained from each subject's voxel data. However, the universality of this approach was not confirmed.

The researchers in the study [136] introduced a CNN-based model with a classification accuracy of 94.54 % for EMCI and LMCI. The work aligned with previous works that applied DL techniques to medical image and text classification tasks. CNNs have demonstrated their effectiveness by automatically learning relevant features specific to the given task, leading to improved results in various medical applications. The study focused on healthy individuals, EMCI, and LMCI classification. The study's limitations included using a relatively small sample size (600 individuals, 200 in each group) and limited result robustness.

An approach in Ref. [137] was suggested to identify individuals with MCI who are at a higher risk of progressing to AD. The proposal encompassed classifying MCI to AD conversion and distinguishing between AD and NC. Notably, the approach was adaptable to various image modalities (PET scans). Utilizing a convolutional framework allows the system to handle different 3D image datasets. Also, the study incorporated a multi-modal feature extractor and employed a 10-fold cross-validation approach for testing and evaluation. The paper acknowledged the problem of data overfitting in DL methods and the data-scarce nature of medical datasets. In Ref. [138], the VGG architecture was optimized for the multi-classification of AD. The authors employed fine-tuned pre-trained CNNs for feature extraction, resulting in classification accuracy across the whole-brain model. The model achieved an impressive accuracy of 99.36 % in the context of multi-classification. However, the model limitations included training from scratch with 64 images per subject, which did not improve accuracy, possibly due to redundant or noisy information when adding more slices. Therefore, results were reported only for up to 32 images per subject.

The authors in Ref. [139] developed a Siamese CNN (SCNN) model, drawing inspiration from the VGG-16 architecture (two modified VGG16 parallel layers, "Siamese"), for the classification of dementia stages. To address the challenges posed by insufficient and imbalanced data, dataset (OASIS dataset) diversity (has been enhanced by integrating augmentation techniques. Employing the proposed approach, the authors achieved a test accuracy rate of 99.05 % in the classification of dementia stages. Limited annotated data and complex preprocessing steps with convoluted parameters impacted accuracy and constrained the research. In Ref. [140], DCNN and a CNN model inspired by VGG-16 (VCNN) were developed for the classification of various stages of AD using MRI images. Experiments were conducted on the ADNI dataset, and the results showed that the proposed models achieved excellent accuracies, with 93.44 % for CN vs. EMCI in the case of DCNN and 96.19

% for VCNN, respectively. The primary emphasis was placed on the sagittal perspective of T1-weighted MRI scans with a restricted scope. Model enhancement required fine-tuning parameters, such as layer count, filter number, size, pooling, and dropout, while ensuring the model size was kept to a minimum.

The study [141] investigated the suitability of ResNets for predicting the progression from MCI to AD. It focused on identifying MCI individuals who will progress to AD within three years of pMCI versus those who remain sMCI. While a test classification accuracy achieved 83.01 %, challenges included limited training data and the computational cost of training DCNNs. In Ref. [142], researchers proposed an Improved Deep Learning Algorithm (IDLA) by utilizing Resting-state functional data (R-fMRI). IDLA aimed to establish brain function and measure connectivity among brain regions. The authors employed a specialized network of autoencoders for early diagnosis, distinguishing between natural aging and disorder progression. Although the model achieved an accuracy of 94.6 %, the investigation into brain region connections using R-fMRI data was restricted at that time.

A multi-task deep CNN and deep 3D DenseNet model were introduced in Ref. [143]. The models involved automatic hippocampal segmentation for AD classification. Although the authors achieved an accuracy of 88.9 % in classifying AD vs. NC, there were limitations in the detailed medical interpretation of learned features relevant to diseases like AD or MCI. Therefore, to enhance interpretation ability, A visualization technique based on class saliency maps was employed (highlighting important brain areas for disease diagnosis). However, the proposed method failed to provide a detailed medical interpretation of the learned features relevant to diseases like AD or MCI.

The authors in Ref. [144] proposed the DCNN model to classify AD, MCI, and NC. The model has been trained and tested based on ADNI datasets and achieved an accuracy of 98.57 %. However, the absence of fine-tuning of hyperparameters may have conceivably impacted the accuracy of predictions. A computer-aided diagnosis (CAD) system has been proposed in Ref. [145]. The authors introduced a DNN classifier for image classification CN, AD, and MCI using ADNI datasets. To find the best weight parameters, the model was optimized with the Enhanced Squirrel Search Algorithm (ESSA). The accuracy of the proposed method for AD/CN, AD/MCI and MCI/CN reached 96.43 %, 94.64 % and 91.07 %, respectively. Nevertheless, the segmentation and classification processes had several challenges in the CAD-based AD diagnosis field. Hence, the development of accurate approaches is needed for better diagnosis.

Modifying the LeNet model proposed in Ref. [146], which incorporated both MinPooling and MaxPooling layers. The model outperformed the original LeNet mode and achieved an average accuracy rate of 96.64 % in classifying AD. This modification allowed for better utilization of low-intensity valued pixels in brain images, enhancing the model's performance compared to the original LeNet model. However, for predicting AD stages (NC, EMCI, LMCI, and AD) using fMRI data from the ADNI dataset, the modified LeNet model has increased memory requirements. Consequently, the model was restricted to investigation into different stages of AD. To achieve an accurate classification among (AD, MCI, EMCI and LMCI), the authors in Ref. [147] presented a fine-tuned ResNet18 model. The model achieved outstanding accuracy rates: 99.99 % for EMCI vs. AD, 99.95 % for LMCI vs. AD, and 99.95 % for MCI vs. EMCI classification. However, the paper did not investigate the model's performance in a multiclass classification scenario.

In [148], the authors introduced a layer-wise transfer learning approach utilizing VGG 19. They conducted experiments on 300 ADNI subjects, split into six binary groups. Although the proposed model achieved an overall accuracy of 83.72 %, it came with a high computational complexity with a small dataset. The authors in Ref. [149] constructed multilayer perceptron representations for individual AD risk and high-resolution disease probability maps. The assessment was based on regional brain activity. In the paper, the class imbalance was addressed using the Synthetic Minority Oversampling Technique

(SMOTE) technique. The ADNI dataset cases comprise five subtypes: AD, MCI, EMCI, LMCI, and NC. The 1296 MRIs from the ADNI dataset were resized to 176×176 to suit the DEMentia NETwork (DEMNET). The model achieved an accuracy of 84.83 %. The study included pre-processing steps, including skull stripping and intensity normalization, which may affect result comparisons. However, the authors did not explore the potential benefits of fine-tuning pre-trained convolutional layers, which could have enhanced the base model's performance. In Ref. [150], the authors introduced a combined approach that employed Hough CNN, CNN, and DNN models for diagnosing AD. They used volumetric features extracted from sMRI data. Using the GARD dataset, the accuracy rates reached 94.82 % (94 %) in the left (right) hippocampus. However, validating the algorithm with larger open-source datasets and exploring its applicability to other modalities' data was essential.

A framework based on DBN was established in Ref. [151] for AD/MCI diagnosis. The authors employed dropout techniques and zero-masking technology to enhance the DBN model's generalization capability. Using the ADNI dataset, the model achieved better classification accuracy rates: 98.62 % for HC vs. AD and 99.62 % for sMCI vs. AD. However, the study mentioned that it did not incorporate additional neuroimaging modalities, such as PET alongside MRI, limiting its ability to comprehensively explore the synergies between different imaging modalities for more reliable biomedical support.

3D DenseNet architecture proposed in Ref. [106] for AD detection. ADNI dataset was used to develop the model. Although the accuracy reached 86 %, the work was conducted within the constraints of freely available tools like Google Collaboratory with limited computational resources and session time. The dataset exhibited a case of class imbalance, with a specific focus on the low number of instances representing individuals with LMCI. This imbalance, along with the overall instance count, had a notable impact on the accuracy of the dataset's predictions. Moreover, the existence of multiple occurrences erroneously classified as AD led to decreased specificity. The study [152] employed several DL-based approaches to classify AD, MCI, and CN cases using brain MRI data. The dataset consisted of 111 cases from ADNI. Multiple ML algorithms, including SVM, were tested for AD classification. Additionally, commonly used DNN models, including AlexNet, VGG-16, VGG-19, and GoogleNet were implemented. After performance comparison, it was evident that DNN models consistently outperformed other methods, achieving 80–90 % accuracy rates. Notably, VGG-19 emerged as the top-performing model, achieving an accuracy of approximately 90 %. Nevertheless, the clinical trials of these computational techniques remain limited.

The authors in Ref. [153] employed two key approaches. Firstly, they utilized standard CNN methods with 2D and 3D structural brain images from the ADNI dataset. The model showed an excellent performance, achieving an accuracy of 97 %. Secondly, they applied transfer learning techniques, resulting in 95.17 % and 93.61 % accuracy rates for 3D and 2D multiclass AD and MCI classifications. However, applying additional pre-trained models and using data augmentation techniques were not evaluated in the study, leaving their effectiveness and limitations unexplored. The article [154] employed a neural network classifier featuring a VGG16 feature extractor for early AD diagnosis. The authors developed the model using two MRI datasets comprising 6400 (dataset 1) and 6330 (dataset 2) images. The model achieved showed better performance using dataset 1, where the accuracy rates achieved 90.4 % for dataset 1 and 71.1 % for dataset 2. Nevertheless, the paper lacked details regarding the diversity and representativeness of the MRI datasets, potentially affecting the model's generalizability.

For multiclass classification of AD, randomized weights of ResNet18 and DenseNet201 were employed in Ref. [155]. The authors used gradient class activation maps to identify the discriminating regions in the model's predictions. Although the results showed that the modified model achieved an accuracy of 98.89 %, the study had limitations. These include the need for multimodal input to distinguish EMCI-LMCI

differences. Furthermore, the suitability of the base domain model from brain images for the target domain. In addition, the requirement for a large dataset reduces overfitting. For brain MRI analysis and AD detection, VGG16 inspired the underlying network architecture utilized in Ref. [156]. The network incorporated a conditional loss function to address the limited sample issue and enhance model accuracy. The authors utilized the OASIS dataset and achieved an accuracy of 99.41 %. Nevertheless, the dataset's limited number of image samples could lead to overfitting and negatively impact the performance of DL models.

CNN architecture in Ref. [157] was employed to classify 2182 image objects from the ADNI. The evaluation involved 29 different pre-trained models, where EfficientNetB0 achieved the highest test accuracy at 92.98 %. The study excluded certain models from further evaluation due to their low performance, resulting in 18 models being considered for accuracy and confusion matrix evaluations. However, certain model groups, including ResNet V2, DenseNet, NASNet, Zeiler and Fergus Network (ZFNet), and Inception, did not achieve efficient results. Furthermore, the study highlighted challenges related to hardware requirements in DL research and the limitations of available online ML/DL tools, leading to the need for repeated tests. The authors in Ref. [158] developed a practical AD diagnostic classifier using DL (the Inception-ResNet-V2 CNN) and transfer learning. They utilized a vast and diverse brain MRI dataset from over 217 sites/scanners, totaling 85,721 scans from 50,876 participants. The model was tested using four datasets and achieved 90.9 % accuracy on the ADNI dataset and 94.5 %, 93.6 %, and 91.1 % on three separate independent datasets, AIBL, MIRIAD, and OASIS, respectively. However, the study exclusively used structural MRI due to concerns about functional MRI reproducibility, although functional measures could potentially improve brain disorder classification.

To assess the impact of image segmentation on classification, the work in Ref. [62] investigated various DL architectures applied to brain MRI and segmented images. The authors implemented pre-processing techniques to enhance MRI scans and post-processing for brain tissue segmentation. Results showed that processed images achieved better accuracy in AD vs. CN binary classification across architectures. The model achieved an accuracy of 90.83 % for original images and 93.50 % for processed ones. Although the model successfully enhanced brain 3D MR images using segmented tissues, it is limited in its focus on tissue-labeled images. Additionally, it lacks an empirical assessment of potential overfitting or network instability.

For detecting AD using transfer learning, the authors in Ref. [159] introduced the Alzheimer's Disease Detection Empowered with Transfer Learning (ADDTLA) model as a modified version of AlexNet. In the study, four different AD stages were used: moderate dementia (MoD), mild dementia (MD), very mild dementia (VMD), and non-dementia (ND). The model achieved an accuracy of 91.70 % for early disease detection. However, the paper lacked a comprehensive analysis of model accuracy and didn't fine-tune all convolutional layers, potentially limiting performance. Furthermore, the authors didn't explore alternative pre-trained networks or CNN architectures that could potentially yield better results. Consequently, the study relied on a specific dataset, which may have restricted the generalizability of its findings.

The study [160] evaluated the CascadeNet and ResNet algorithms and demonstrated their classification of AD stages in MRI. Utilizing Kaggle datasets, the proposed architecture achieved 99.02 % accuracy in quaternary classification, distinguishing between various AD subtypes. However, the proposed model operated without the need for feature extraction. Furthermore, the limitations included the absence of skull alignment and subtraction in MRI preprocessing, which may have impacted the results. Moreover, the study did not incorporate patient history information, limiting the context and connections that could have been established with the patient's past medical records.

The authors in Ref. [161] focused on two-class AD using a patch-based feature extraction approach and a 3-Dimensional ConvNet model. The MRI images were used from MIRIAD, and the 3D-CNN model

achieved 99.79 % accuracy for AD detection. However, the study lacks several limitations: 1) focus on a two-class classification task, 2) a reliance on patch size for improved accuracy, and 3) a lack of exploration of alternative neuroanatomical computation extraction methods, which may have restricted the generalizability of the findings in real-world clinical scenarios. Additionally, the authors did not provide an empirical assessment of the impact of visualizing problematic areas in scans on the effectiveness of the diagnostic method.

A modified InceptionV3 (AlzheimerNet) was proposed in Ref. [162] and implemented to classify all five AD stages and SMC using ADNI MRI data. Data preprocessing involved Contrast Limited Adaptive Histogram Equalization (CLAHE) enhancement and data augmentation to address dataset imbalance, resulting in 60,000 images across six classes. Initial training and testing of five existing models yield test accuracies ranging from 78.84 % to 96.31 %. Additionally, AlzheimerNet achieved the highest accuracy at 98.67 %. However, the study did not explore the potential of a hybrid model incorporating fMRI and PET datasets for diagnosing AD stages, which could have offered a more comprehensive understanding of the disease.

The authors in Ref. [163] developed a CNN architecture and utilized 2D MRI data from the OASIS. The reported results demonstrated the performance of the model with an overall testing accuracy of 99.68 %. However, the study did not address potential variations in AD prediction accuracy across different disease stages or age groups. Furthermore, advanced data mining algorithms and diverse datasets were not thoroughly explored. In the research [164], dynamic neuro-functional deep ensemble networks were introduced to predict various stages of AD. The authors utilized different frequency signals derived from rs-fMRI data. Among the three analyzed frequency features (slow4, slow5, and full-band), slow5 exhibited the highest performance when using the (Alexnet + Inception) model (named D2). The model achieved an accuracy of 96.61 % in distinguishing AD from normal (NL) subjects. The D2 model achieved 81.87 % and 82.67 % in tests (MCI vs. NL) and (AD vs. MCI), respectively. However, despite advancements in medical image classification, issues such as non-universality, sensitivity to lighting conditions, and reduced accuracy due to data quality could still affect the reliability of the models.

The authors in Ref. [165] utilized CNN on MRI image datasets for multi-classifying AD subjects into CN, MCI, EMCI, and LMCI categories. Transfer learning was employed using pre-trained health data classification models like MobileNet. The test results indicated that fine-tuning the MobileNet pre-trained model showed a higher accuracy of 96.6 % for multi-class AD stage classifications. Using the same ADNI dataset, the authors evaluated several models, including VGG16 and ResNet50. However, augmentation approaches, like downsampling, were not implemented, which potentially may impact model robustness. Additionally, the paper did not address multi-modal fusion techniques, which could have limited comprehensiveness.

The Federated Learning (FL) model was introduced in Ref. [166] involving two local clients, each using the MobileNet model. To train the MobileNet model efficiently, the authors employed a parallel approach where multiple devices trained the model and only shared model weights, not the raw data, with a central server. The proposed model achieved high accuracy: 95.24 % (OASIS), 81.94 % (ADNI), and 83.97 % (ADNI + OASIS). However, the model is limited classification to AD-positive and AD-negative without distinguishing AD sub-stages were not discussed. In Ref. [167], an early feature fusion approach was introduced. Using a 3-in-channel approach, the approach combined PET and MRI images from ADNI and trained a modified Resnet18 architecture on both datasets. The model achieved a testing accuracy of 96.73 % and a validation accuracy of 73.91 % on the ADNI database. Additionally, the authors developed an Explainable Artificial Intelligence (XAI) model to explain the results. However, the diagnostic accuracy of the document is not excellent, where relatively low classification accuracy was achieved.

A local cloud-based solution was developed in Ref. [168] for

classifying AD using MRI scans. The local server offered Software as a Service (SaaS), and the front end was designed for user-friendly access to the system. The multi-classification approach categorized AD into four stages. Utilizing transfer learning with the pre-trained GoogLeNet model, the accuracy achieved was 98 %, resulting in the development of a local cloud web application for Alzheimer's prediction. Furthermore, the authors developed an application that enables doctors to assess the presence of AD in patients remotely. Nevertheless, the paper acknowledged potential difficulties that doctors might have faced when remotely accessing the developed interface, leaving uncertainties about the usability and effectiveness of the solution.

A CNN model was utilized in Ref. [169] to classify the brain MRI dataset ADNI into four groups: MCI, AD, CN, and Healthy Person. The proposal classification accuracy rate reached an accuracy of 95.82 %. However, the study's limitations encompassed resource constraints and operational challenges associated with Google Colab. Additionally, the research was limited in scope as it primarily focused on Alzheimer's disease data, potentially hindering its application to broader medical classification issues and generalization across diverse datasets from various domains. In the article [170], various DL algorithms (all 26 pre-trained Keras DL models) were employed for the multi-class classification of AD. The authors utilized ADNI and identified ResNet-50v2 as the best performance, achieving 91.84 % accuracy and an f1-score of 0.97 for the AD class. Nevertheless, the primary limitation of the study was its focus on AD diagnosis without testing other types of dementia with overlapping symptoms. Also, the inclusion of demographic features could have enhanced prediction reliability. Another drawback was the lack of exploration of the potential for improved model performance by utilizing multimodal or ensemble models (which may reduce noise and capture both linear and nonlinear relationships in the data).

The paper [171] presented the Inception V3 model for identifying and classifying different AD types using brain MRI from OASIS. Although the model achieved higher accuracy (89 %), the model suffers from several limitations. The paper's limitations included not thoroughly exploring the impact of different transfer learning models on AD detection and having a narrow focus on a specific type of medical imaging data, potentially limiting the development of a more comprehensive diagnosis approach. In Ref. [172], the authors utilized three CNN models (VGG16 and Xception models and a fully customized CNN model) to detect and classify four stages of AD from 2D MRI on Kaggle. The custom model achieved the best performance with an accuracy of 94.77 % and an F1-score of 0.9481. However, the model needs more real-world images in training for reliable performance.

The authors in Ref. [173] proposed an Image-Fused multi-modality CNN architecture named "Image-Fused" for early diagnosis of AD. The study investigated the impact of hyper-parameters, including the training-testing split and learning rate, on performance. The model achieved a testing accuracy of 98.9 % with a learning rate of 0.0001 and an 80:20 % training-testing ratio. However, the limitations of the work included the omission of an exploration of fMRI data integration with the proposed model. Furthermore, there was a lack of integration of clinical and genetic data and a focus on simulation rather than real-world application. Moreover, there was a need for improved noise removal techniques and feature extraction methods to enhance the system's performance.

Building upon the above analysis, we observed that the ADNI dataset was the most prevalent choice among the reviewed papers. The OASIS dataset followed in popularity. To provide a clearer overview, we summarized the findings as follows.

1. Table 4 provides a critical summary of the works that utilized the ADNI dataset.
2. Table 5 summarizes the studies that utilized the OASIS dataset.
3. Table 6 lists the contributions of the papers that the studies that utilized datasets other than ADNI or OASIS.

Table 4

A critical summary of the works that utilized the ADNI dataset.

Ref	Year	Dataset	Classes	Algorithms	Results				Limitation	
					Acc (%)		Other			
[131]	2017	ADNI	NC, MCI, LMCI, and AD	GoogLeNet,	98.88	For MCI: Specificity = 99.9 % Sensitivity = 99.9 % PPV = 99.8 % For LMCI: Specificity = 99.9 % Sensitivity = 99.9 % PPV = 99.7 % For AD: Specificity = 99.9 % Sensitivity = 99.9 % PPV = 99.8 %			The absence of incorporating clinical data and multimodal data handling limited the system's robustness.	
				ResNet-18	98.01					
				ResNet-152	98.14					
[132]	2018	ADNI	NC, pMCI, sMCI, and AD	3D CNN	Multi-modality 95.68		Sensitivity = 81.08 % Specificity = 84.31 % AUC = 88.43 %		The method needed a large training dataset, which was challenging for multi-modality brain images. 1. Using MRI only, genomics and biomarkers might enhance the results. 2. The hierarchical feature extraction method still needs to be improved. 3. The complementarity between multimodal data was not taken into consideration.	
[135]	2019	ADNI-1	CN, EMCI, LMCI, and AD	CNN	AD vs. MCI 97.2 AD vs. NC 99.4 NC vs. MCI 98.9		The NC vs. MCT group has the highest results: Precision = 99.4 % Recall = 98.5 %.			
		ADNI-2			AD vs NC 98.6 AD vs. LMCI 99.7 AD vs. EMCI 100 NC vs. EMCI 99.7 NC vs. LMCI 98.5 EMCI vs. LMCI 98.0		The AD vs. EMCI group has a perfect score of 100 % for Precision, Recall and f1-score metrics.			
[136]	2019	ADNI (600 individuals)	CN, EMCI, and LMCI	CNN	CN vs. LMCI 94.54		The CN vs. LMCI group in the sagittal view achieved the highest results: Sensitivity = 91.70 % Specificity = 98.19 % F-score = 94.84 % AUC = 99.40 %		A small sample size (600 individuals, 200 per group) impacts the resulting robustness. 1. Overfitting in DL methods due to high parameter count. 2. Challenges with data scarcity in medical databases for network architecture construction.	
[137]	2019	ADNI	HC, pMCI, sMCI, and AD	CNN	Overall accuracy = 86		–			
[138]	2019	ADNI	NC, MCI, and AD	Modified VGG	Overall accuracy = 99.36		For CN vs. AD: Specificity = 100 % For MCI vs. NC: Sensitivity = 99.5 %.			
[140]	2020	ADNI	CN, EMCI, LMCI and AD	DCNN and CNN		DCNN	VCNN	1. The DCNN's lowest training loss is 2.02 % for the classification task of AD vs LMCI. 2. The lowest testing loss is 3.17 % for the AD vs CN classification task.	1. The VCNN lowest training loss is 1.74 % for the classification task of AD vs LMCI. 2. The lowest testing loss is 4.82 % for	1. Focuses on the sagittal view of T1-weighted MRI images, limiting its scope. 2. Model improvement requires optimizing parameters like layer count, filter number, size, pooling,
					CN vs EMCI	93.44	96.19			
					AD vs CN	91.40	93.78			
					CN vs LMCI	93.76	94.87			
(continued on next page)										

(continued on next page)

Table 4 (continued)

Ref	Year	Dataset	Classes	Algorithms	Results				Limitation		
					Acc (%)		Other				
					AD vs EMCI	91.56	92.80			the classification task of AD vs. CN	and dropout while keeping model size minimal
					AD vs LMCI	93.28	94.79				
					LMCI vs EMC	89.68	92.91				
[141]	2020	ADNI	CN, sMCI, pMCI, and AD	ResNet with different depths (D = 1, 2, 3, and 4)	Overall accuracy = 83.01		Sensitivity = 76 % Specificity = 87 % AUC = 0.88		1. Limited training data size. 2. Computationally expensive nature of training deep CNNs.		
[142]	2020	ADNI	NC, EMCI, LMCI, and AD	Autoencoders	Overall accuracy = 94.6		Sensitivity = 94.6 % Specificity = 96.7 %		1. IDLA relies on background information for predictions. 2. Limited exploration of brain region connections using R-fMRI data.		
[143]	2020	ADNI	NC, MCI and AD	Multi-task deep CNN +3D DenseNet	For AD vs. NC = 88.9 For MCI vs. NC = 76.2		A dice similarity coefficient = 87.0 % PPV = 84.6 % Sensitivity = 89.7 %		Lacks a detailed medical interpretation of learned features for practical clinical use in understanding brain abnormalities.		
[144]	2021	ADNI	NC, MCI, and AD	DCNN	Overall accuracy = 98.57		Validation accuracy = 87.72 %		1. Limited evaluation of performance metrics (specificity, sensitivity, recall, F1-score) for the DCNN model. 2. The absence of hyperparameter tuning may impact predictive accuracy.		
[145]	2021	ADNI	CN, MCI, and AD	DNN	Overall accuracy = 96.43		Sensitivity = 96.43 % Specificity = 96.43 % AUC = 0.964		Segmentation and classification challenges in CAD-based AD diagnosis require the development of more accurate approaches for improved diagnosis.		
[146]	2021	ADNI	CN, MCI, and AD	Modified LeNet	Overall accuracy = 96.64		Best performance, CN vs MCI in the 60–69 age group: F1 score = 0.97, Precision = 96 %, Recall = 98 %. Best performance, MCI vs AD in the 70–79 age group: F1 score = 0.98, Precision = 96 %, Recall = 97 %. Best performance, CN vs AD in the 70–79 age group: F1 score = 0.97, Precision = 0.96 %, Recall = 95 %.		1. There were increased memory requirements due to the modified LeNet model. 2. There was a limited exploration of various stages of AD. 3. Dependence on a single dataset source (ADNI) may have affected generalizability.		
[147]	2021	ADNI	NC, EMCI, LMCI and AD	Modified ResNet18	CN vs. EMCI: 98.74 CN vs. LMCI: 92.23 EMCI vs. LMCI: 99.76	97.24 90.23 99.56	Sensitivity (%) 94.21 99.97	Specificity (%)	This study did not investigate the model's performance on a multiclass classification case.		
[148]	2021	ADNI	NC, EMCI, LMCI and AD	VGG-19	AD vs. CN: 98.73 EMCI vs. LMCI: 83.72	98.19 82.09	Sensitivity (%) 99.09 85.13	Specificity (%)	The study used a small sample: 300 patients from ADNI. It had high computational complexity.		
[151]	2021	ADNI	HC, sMCI , pMCI and AD	DBN	HC vs. AD = 98.62 HC vs. pMCI = 96.67 HC vs. sMCI = 92.31 pMCI vs. AD = 91.89 sMCI vs. AD = 99.62 sMCI vs. pMC = 87.78		–		The paper did not utilize additional neuroimaging modalities like PET alongside MRI, limiting its understanding of synergies between modalities in the biomedical field.		

(continued on next page)

Table 4 (continued)

Ref	Year	Dataset	Classes	Algorithms	Results					Limitation	
					Acc (%)		Other				
[106]	2021	ADNI	CN, MCI, EMCI and AD	3D DenseNet-121	Overall accuracy = 86		Specificity = 86 % Precision = 86 % Recall = 86 % AUC = 0.91			1. Constraints from Google Colab limited data augmentation and epochs. 2. The dataset had class imbalance, particularly low LMCI instances. Imbalance and instance count impacted prediction accuracy. 3. There is low specificity due to multiple instances being misclassified as AD.	
[152]	2021	ADNI	NC, MCI and AD	SVM AlexNet VGG16 VGG19 GoogLeNet	74.25 85.14 88.92 90.02 87.29			–			There were very limited clinical trials of computational techniques.
[153]	2021	ADNI	AD, EMCI, LMCI and NC.	2D CNN 3D CNN VGG19	93.61 95.17 97			1. For AD 3D CNN and MCI and NC in 2D CNN: The highest Precision = 0.98 %. 2. For NC in 3D CNN: The highest Recall = 98 % F1 Score = 0.98			Additional pre-trained models and data augmentation techniques were not evaluated, leaving their effectiveness and limitations unexplored.
[155]	2022	ADNI	NC, MCI, EMCI, LMCI and AD	ResNet18+ DenseNet201	Multiclass classification: Overall accuracy: 98.8		Multiclass classification: Precision = 98.94 % Recall = 98.89 %			1. It should have considered multimodal input for better EMCI-LMCI differences. 2. The base domain model from brain images suited the target domain. 3. A large dataset was needed to reduce overfitting.	
[157]	2022	ADNI	CN, MCI, and AD	AlexNet EfficientNetB0 EfficientNetB1 EfficientNetB2 EfficientNetB3 EfficientNetB4 EfficientNetB5 EfficientNetB6 EfficientNetB7 LeNet MobileNet MobileNetV2 ResNet50 ResNet101 ResNet152 VGG16 VGG19 Xception	89.95 92.98 91.91 90.88 90.93 88.33 89.91 89.49 87.12 87.72 83.40 80.19 88.56 86.98 87.77 89.72 89.77 77.40			1. The highest rates in the models: EfficientNetB3: Precision = 89.78 % EfficientNetB2: Sensitivity = 94.42 % EfficientNetB3: Specificity = 97.28 % 2. The highest sensitivity rate for the NC class: VGG19 model: 93.80 %, 3. The highest sensitivity rate for the MCI class: ResNet152 model: 89.00 %. 4. The highest precision rate for the NC: VGG19 model: 90.98 %. 5. The highest precision rate for the MCI class: Xception model: 88.54 %. 6. The highest specificity rate for the NC: VGG19 model: 94.92 %. 7. The highest specificity rate for the MCI class: ResNet101 model: 95.22 %.			1. Global databases were used due to limitations, often being inadequate. 2. Hardware requirements posed challenges in DL studies.
[62]	2022	ADNI	CN and AD	3D ResNet-20	AD vs. CN Overall accuracy: 93.7		–			1. It is limited to tissue-labeled images and may not apply to other segmentation methods. 2. There’s no evaluation of overfitting or network stability.	
[162]	2023	ADNI	AD, EMCI, LMCI, MCI, SMC, and NC.	Modified InceptionV3 named AlzheimerNet	Overall accuracy: 98.6		Precision = 98.68 % Recall = 98.68 % Specificity = 99.74 % F1-score = 98.68 %			It did not explore a hybrid model using fMRI and PET data for AD stage diagnosis, limiting its scope.	
[164]	2023	ADNI	AD, MCI, and NL	Alexnet + Inception named (D2)	(Slow5) AD vs. NL: 96.61 AD vs. MCI: 82.67 NL vs. MCI: 81.87	Sensitivity 94.3 81.1 79.8	Specificity (%) 80.1 75.47	0.966 0.844 0.822	AUC	1. Medical image classification is vital and has challenges. 2. Existing methods lack universality and have issues. 3. Standard classification might not suit the real world. 4. The study used a small ADNI dataset.	

(continued on next page)

Table 4 (continued)

Ref	Year	Dataset	Classes	Algorithms	Results		Limitation
					Acc (%)	Other	
[165]	2023	ADNI	AD, EMCI, LMCI, and NC.	MobileNet VGG16 DenseNet121	96.22 91.66 92.44	1. for Final CN JPEG: The lowest F1-score = 0.93. 2. The macro average and weighted average of precision and recall are 96 %.	1. A small dataset may have affected accuracy. 2. A lack of discussion on augmentation. 3. Multi-modal fusion was not explored, limiting comprehensiveness.
[167]	2023	ADNI	NC, EMCI, LMCI and AD	ResNet50 Resnet18	66.22 Testing accuracy: 96.7 Validation accuracy: 73.9	3. The highest F1-score: 0.96 Specificity: 94.32 % Sensitivity: 66.74 %	The paper lacked optimal diagnostic accuracy, as indicated by the classification accuracy achieved by 73.90 % on the ADNI database.
[168]	2023	ADNI	LMCI, AD, NC, and EMCI.	Fine-tuned GoogLeNet	Overall accuracy: 98	–	Uncertainties existed regarding the usability and effectiveness of the solution due to potential difficulties faced by doctors when remotely accessing the interface.
[169]	2023	ADNI	Class 0: MCI, Class 1: AD, Class 2: CN, Class 3: Healthy Person	CNN	Overall accuracy: 95.82	1. Best in classifying the CN group, with a Precision of 93.7 %, Recall of 94.7 %, and F-score of 96.2 %. 2. Worst in classifying the MCI group, with a Precision of 90.7 %, Recall of 92.3 %, and F-score of 94.5 %.	1. Using Google Colab imposed limitations, including resource constraints, temporary storage, idle timeouts, and shared resources. 2. The study focused on AD data, limiting its applicability to other medical classification issues. 3. Limited exploration of diverse datasets from various domains affected the model's generalizability.
[170]	2023	ADNI	LMCI, AD, NC, MCI, and EMCI.	ResNet-50v2	Overall accuracy 91.84	1. The highest rates for AD class are: Precision = 95 %. Recall = 99 % F1-score = 0.97. 2. The lowest rates for NC class are: Precision = 75 %. Recall = 76 % F1-score = 0.76.	1. The study focused solely on AD diagnosis, neglecting other dementia types with overlapping symptoms. 2. The absence of demographic features limited prediction reliability. 3. Multimodal or ensemble models weren't explored, missing the potential for improved model performance and noise reduction.
[173]	2023	ADNI	NC, and AD	Multi-modality CNN (named Image Fused)	Overall accuracy: 98.9	Precision = 100 % Recall = 99 % F1-score = 0.99	1. Limited exploration of fMRI data integration. 2. Lack of clinical and genetic data integration. 3. Emphasis on simulation over real-world application. Improvement is needed in noise removal and feature extraction methods.

Table 5

A critical summary of the works that utilized the OASIS dataset.

Ref	Year	Dataset	Classes	Algorithms	Results		Limitation
					Acc (%)	Other	
[133]	2018	OASIS	Non-demented, Very mild dementia, Mild dementia and Moderate dementia	CNN (M1+M2+M3)	Overall accuracy: 93.18	Precision: 94 % Recall: 93 % F1-score: 0.92	The dataset had 416 sMRI data points, but a larger dataset is essential for effective DNN training.
[139]	2020	OASIS	CDR-0 (NoDementia), CDR-0.5 (Very Mild Dementia), CDR-1 (Mild-Dementia), and CDR-2 (Moderate AD)	SCNN inspired by VGG16	Overall accuracy: 99.05	–	1. There was a smaller number of annotated data. 2. Preprocessing steps such as skull stripping, segmentation, and normalization had convoluted parameters, which were the big problems in dealing with many parameters correctly.
[156]	2022	OASIS	No Dementia, Very Mild Dementia, Mild Dementia, and Moderate AD	Deep triplet network inspired by VGG16	Overall accuracy: 99.41	–	The lack of image samples in the datasets caused over-fitting and deteriorated the performance of DL models.
[163]	2023	OASIS	ND, MoD, MD, and VMD.	CNN	Testing: 99.68 Training: 100 Validation: 99.27	Sensitivity: 100 % Specificity: 100 % AUC: 100 %	1. The study did not explore AD prediction variations across stages or ages. 2. It lacked investigation of advanced data mining algorithms and diverse datasets.
[171]	2023	OASIS	Mild Demented, Moderate Demented, Non-Demented, and Very Mild Demented	Inception V3	Overall accuracy: 89	1. The highest results for "very mild demented" are specificity, sensitivity, and recall, which are all 100 %. 2. The lowest precision for "Moderate Demented": 78 % 3. The lowest recall and F1-score for "Mild Demented" are 0.79 and 0.82, respectively.	1. limited exploration of the impact of various transfer learning models on AD detection accuracy. 2. The study's narrow focus on specific imaging data (MRI) may have hindered a more comprehensive approach by integrating other data types like PET scans.

8. Challenges and open research directions

Balancing model complexity and clinical relevance is essential to obtain an optimal AD diagnosis using DL, which offers promise in AD diagnosis. This section addresses RQ6 and explores the main remaining challenges that hinder the optimal utilization of deep learning for AD diagnosis. Fig. 27 presents a breakdown into five key challenges as follows.

8.1. availability and generalizability of large datasets

The effectiveness of both ML and DL algorithms in disease diagnosis heavily depends on the availability of high-quality training models. Additionally, the scarcity of annotated data presents a significant challenge in AI-based medical diagnosis [174]. Despite the presence of on-line datasets like ADNI and OASIS, the limited availability of medical data hinders progress [175]. Achieving high generalizability and ensuring patterns extend to broader contexts remain difficult due to data privacy constraints, impacting the reliability of model predictions for new, unseen data. While data enhancement and transfer learning techniques mitigate over-fitting, the shortage of sufficient data samples may still lead to generalization issues [176]. Table 7 provides an overview of the most frequently used private and public AD datasets.

8.2. overfitting problem

Overfitting occurs when the model is trained on data that contains noise and excessive, irrelevant details. Additionally, using an insufficient amount of training data can lead to overfitting. Increasing the volume of available data is a recommended solution to address this issue [177]. To alleviate this issue, the dropout can drop the units (i.e., hidden and visible) in a neural network [178].

8.3. black box challenges

While ML and DL methods hold promise for brain disease diagnosis and predictions, they often lack transparency, making it challenging to understand their internal workings clearly [179]. One of the trending issues is the "black box" problem. Neural networks often likened to black boxes that transform input into output, are frequently employed in machine learning techniques [180]. While the mathematics behind constructing a neural network is straightforward, comprehending the intricate process by which it produces outputs can be challenging. ML algorithms collect vast amounts of data, identify patterns, and create predictive models. However, gaining a clear understanding of the inner workings of these models remains a challenge [181]. It cannot provide specific reasons or explanations for selecting particular features during the training process or for the variations in performance—whether high or low. Furthermore, it doesn't elucidate how the relationships within the training data are reflected in the feature selection process [182].

Recently, XAI has emerged as a crucial tool for enhancing the transparency of AI-based systems. The primary objective of XAI is to provide methods that produce more interpretable models while maintaining high performance levels. In the context of brain disease diagnosis, identifying suitable XAI approaches [183–185] can contribute to verified predictions, improved models, and new insights. Ultimately leading to more reliable and trustworthy brain disease detection systems.

8.4. interoperability and collaboration

In brain disease detection, numerous approaches exist for AI-based hardware and software solutions. However, manufacturers' differing rules, regulations, and interfaces can lead to interoperability issues [186]. To mitigate this issue, collaborative efforts involving healthcare providers, manufacturers, and AI scientists are essential for improving brain disease treatment quality and mitigating data scarcity. Leading global health organizations like the World Brain Alliance, World Health

Table 6

A critical summary of the works that did not utilize ADNI or OASIS datasets.

Ref	Year	Dataset	Classes	Algorithms	Results		Limitation
					Acc (%)	Other	
[134]	2019	ADNI + Milan	HC, s-MCI, c-MCI and AD	CNN	HC vs. AD: 98.2 c-MCI vs. s-MCI: 74.9 c-MCI vs. HC: 87.7 s-MCI vs. HC: 76.4 AD vs. c-MCI: 75.8 AD vs. s-MCI: 86.3	1. For AD vs. HC classification, the highest rates are: Sensitivity = 98.1 % Specificity = 98.3 % 2. For the AD vs. c-MCI classification: The lowest sensitivity = 74.8 %. 3. For the c-MCI vs. s-MCI classification: The lowest specificity = 74.9 %.	1.Future c-MCI presence was not excluded among s-MCI. 2. Longer follow-ups could have enhanced the diagnosis and the algorithm. 3. The model needed testing with various biomarkers for dementia prediction. 4. Typical AD presentations need effective diagnostic tools. 5. Preclinical phase accuracy testing and general population screening were needed for risk identification.
[149]	2021	Kaggle	MD, MOD, ND, and VMD.	CNN (named DEMNET)	Overall accuracy: 95.23	Precision = 96 % Recall = 95 % F1-score = 95.27 % AUC = 97 %	1. The omission of pre-processing steps such as skull stripping and intensity normalization without thorough may have impacted the model's effectiveness. 2. The study didn't explore the potential benefits of fine-tuning pre-trained convolutional layers, which could have enhanced the base model's performance.
		ADNI	NC, EMCI, LMCI and AD		Overall accuracy: 84.83	AUC = 95.62 %	
[150]	2021	GARD	NC and AD	CNN + DNN	LH-DNN: 94.82	Precision = 97 % Recall = 86 % F1-score = 91 %	1. It depended on the accuracy of previous localization and volume estimation methods. 2. There was validation on a small dataset.
					RH-DNN 94.02	Precision = 100 % Recall = 81 % f1-score = 90 %	
[154]	2022	Kaggle AD dataset1	MildDemented, ModerateDemented, NonDemented, and VeryMildDemented	DCNN and VGG16 (as feature extractor)	Overall accuracy 90.4	Precision = 90.5 % Recall = 90.4 % AUC = 96.9 % F1-score = 0.904	The dataset was limited in size and diversity, potentially impacting the model's ability to generalize its findings to broader populations or data distributions.
		Kaggle AD dataset2	non-demented, mild demented, and very mild demented		Overall accuracy 71.1	Precision = 71 % Recall = 71.1 % AUC = 85 % F1-score: 0.71	
[159]	2022	Kaggle	MD, MoD, ND, and VMD.	Modified AlexNet named ADDTLA	Overall accuracy 91.70	Sensitivity = 93.7 % Specificity = 96 % Precision = 91.5 % F1 score = 0.925	1. Limited model accuracy analysis and fine-tuning. 2. It ignored alternative pre-trained networks and architectures. 3. Findings were possibly limited due to dataset choice.
[160]	2022	Kaggle	dementia, very mild dementia, mild dementia, and moderate dementia	Cascade-ResNet-18	The accuracy of the quaternary classification: 99.02	Precision = 99.02 %	1. Lack of skull alignment and subtraction in MRI preprocessing may impact results. 2. The absence of patient history information limits contextual connections with medical records.
[161]	2022	MIRIAD	NC and AD	3D-CNN	Test accuracy: 99.79	Precision = 100 % AUC = 100 %	1. Limited to a two-class classification task. 2. It relied on patch size for improved accuracy. 3. There was no exploration of alternative neuroanatomical computation extraction methods, potentially limiting generalizability.
[166]	2023	ADNI+ OASIS	AD-positive and AD-negative	Modified MobileNet (Federated approach)	83.97	Precision = 74.29 % Sensitivity = 61.90 % Specificity = 92.11 %	Computational cost and limited classification of AD-positive and AD-negative without distinguishing AD sub-stages were not discussed.
[172]	2023	Kaggle	Non-demented, very mild demented, mild demented, and moderate demented	Fully customized CNN	94.77	1. The customized model achieved the highest rates in: Precision = 92.75 % Recall = 95 % F1-score of 0.948 2. VGG16 achieved the highest AUC = 99.3 %. 3. The Xception model achieved the lowest rates in: Precision = 91.51 % Recall = 91.41 %	1. A small number of images compared to traditional DL applications. 2. Class imbalance resulted from the unequal number of images in each class.

(continued on next page)

Table 6 (continued)

Ref	Year	Dataset	Classes	Algorithms	Results		Limitation
					Acc (%)	Other	
						AUC = 97.54 % F1- score = 0.915.	

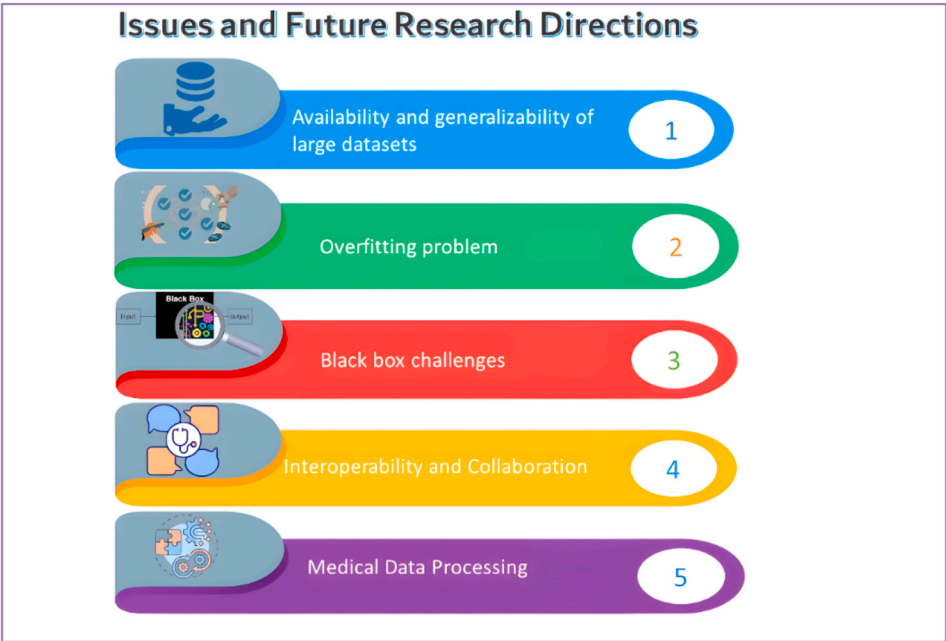


Fig. 27. Challenges and future directions in AD literature.

Table 7
AD Dataset state.

Datasets	State
ADNI	Public
OASIS	Public
MILAN	Private
NACC	Public
GARD	Private
MIRIAD	Public

Professions Alliance, and the World Health Organization could collaborate with international AI groups to tackle these challenges [187].

8.5. medical data processing

As mentioned earlier, AI applications often face hardware limitations, especially when dealing with medical data, due to the need for lossless data preservation. This becomes more critical as increased processing power demands more memory and computational resources [188]. Image pre-processing is a major concern, as its both time-consuming and space-intensive. Interestingly, object detection techniques can achieve high-accuracy AD prediction without pre-processing. Future efforts could focus on these methods to reduce overhead and costs. With the typically large volume of data used for brain disease detection, diverse data sources, and real-time sensor data, data processing platforms encounter significant challenges in effectively managing and maintaining generated data [189].

9. Main research key findings

9.1. Frequently use of CNN pre-trained AD diagnosis models

To highlight the prevalence of some models in academic literature, shedding light on the popular choices in this field. It is important to note that these models may often be modified, fine-tuned, customized, or integrated with other models to suit the specific requirements of AD diagnosis and detection studies. Therefore, we extensively investigated the most utilized models by supreme papers published in highly ranked publishers (such as, Science Direct, IEEE, and Springer). This review leverages the latest research by focusing on articles published within the past ten years (2013–2023). As shown in Fig. 28, the number of articles published in Science Direct outperformed other publishers in the context of AD detection. While VGG is the most utilized model, mobileNet is employed the least among the studied architectures. The best explanation of this utilization is the best-achieved performance via the VGG model.

9.2. ADNI vs OASIS datasets: a decade-long review (2013–2023)

The most commonly used publicly available online data sources are ADNI and OASIS. Therefore, from 2013 to 2023, we searched for original papers that utilized these two datasets from three highly reputable scientific databases: Science Direct, IEEE, and Springer. We carried out two search campaigns using defined keywords. First: ("Alzheimer's disease" OR "AD") AND ("Alzheimer's Disease Neuroimaging Initiative" OR "ADNI"). Second: ("Alzheimer's disease" OR "AD") AND ("Open Access Series of Imaging Studies" OR "OASIS"). As depicted in Fig. 29, ADNI is consistently used more frequently than OASIS across these publishers, suggesting its wider usage and relevance in academic research during this decade. Additionally, the analysis of the studies demonstrated that

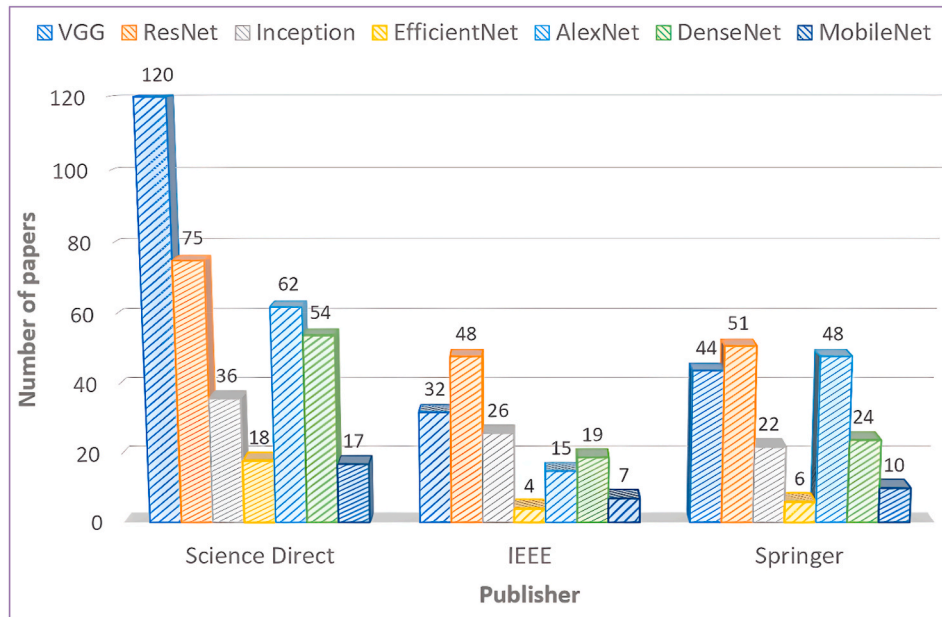


Fig. 28. Frequently used CNN pre-trained AD diagnosis models in Science Direct, IEEE, and Springer from 2013 to 2023).

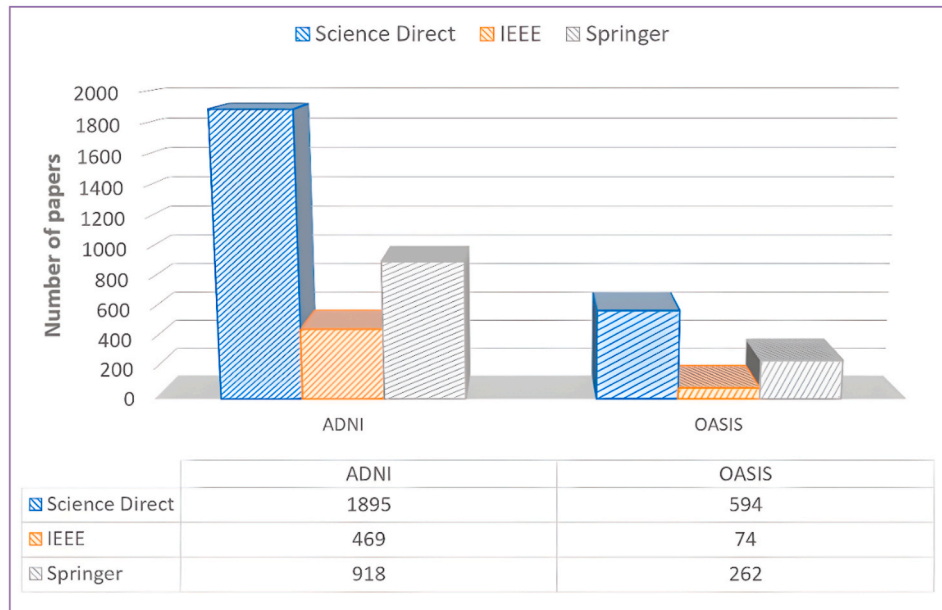


Fig. 29. Number of papers in the ranked scientific databases according to AD datasets from 2013 to 2023.

papers published in Science Direct journals are more than other publishers. Therefore, By analyzing these papers, researchers can explore and identify promising techniques for early AD detection in Science Direct.

9.3. MRI vs. PET

As illustrated in Fig. 30, MRI is the dominant imaging technique employed in AD detection studies. This dominance is because of the cost associated with radiopharmaceuticals required for PET scans [187].

9.4. Deep learning models

Fig. 31 Our analysis of 45 papers (Fig. 31) identified that the ResNet model and CNN models are the most prevalent DL architectures employed in AD diagnosis. VGG and other architectures were also utilized, although to a lesser extent. These models were employed in various versions and dimensions to achieve accurate classification, often customized, fine-tuned, or integrated with other models.

10. Conclusion

In this paper, we provided an extensive survey of the recent attempts to diagnose AD using DL. We particularly focus on the role of imaging

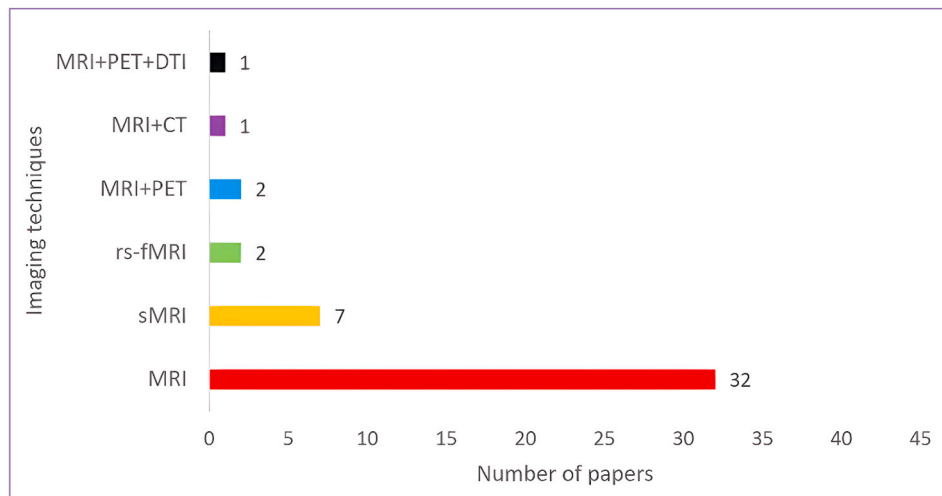


Fig. 30. Image modalities used in different articles.

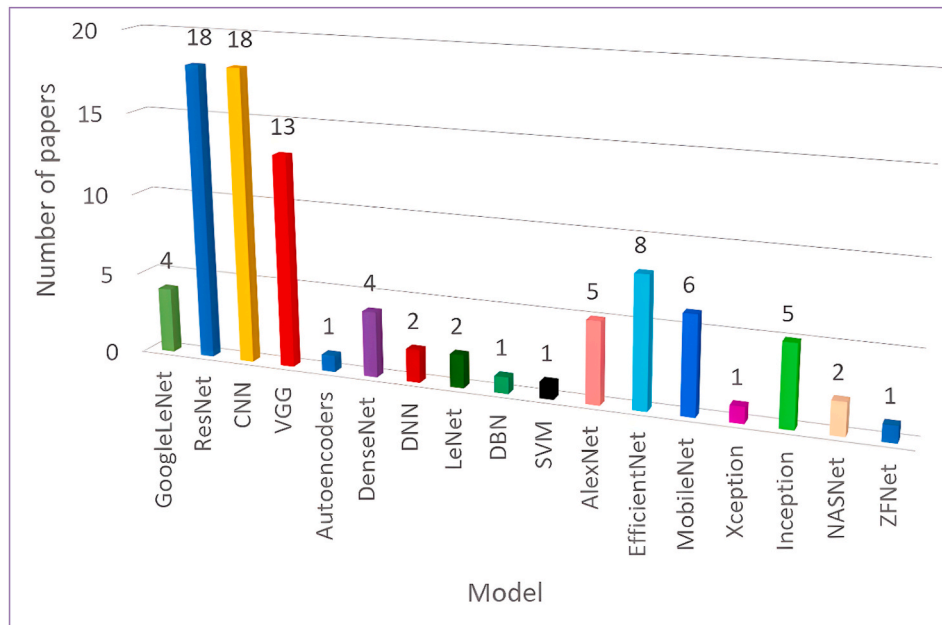


Fig. 31. The models of the selected articles.

modalities and preprocessing techniques in enhancing image quality for accurate AD diagnosis. Furthermore, we conducted a comparative analysis of recent research to clarify the latest DL developments. Therefore, one of the main findings from this survey, compared to other publishers, is that the papers published in Science Direct journals are the most utilized CNN pre-trained models and datasets in the last decade. Furthermore, the study showed that MRI imaging techniques have been utilized more than other techniques. Moreover, the survey presented practical implications and open research directions for diagnosing AD.

Nevertheless, the absence of comprehensive medical records and datasets in Arabic countries poses a significant challenge for AD diagnosis in Arab countries. Addressing this gap is crucial and will contribute to raising awareness and fostering local expertise in Alzheimer's research within the Arab healthcare landscape. Therefore, one of our primary focuses is creating an Iraqi AD dataset for patients from different cities. This comprehensive dataset will facilitate the development and validation of AI models specifically tailored for the diagnosis

of AD in the Iraqi population. Finally, we believe that this survey helps as a valuable resource for researchers in the AD context and acknowledges the existing challenges.

10.1. List of acronyms

Table 8 presents the list of acronyms used in this article.

Ethical statement

This paper is a systematic review (not an original work) and does not need any Ethical approval.

CRediT authorship contribution statement

Eqtidar M. Mohammed: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology,

Table 8
Nomenclature.

Acronym	Full Form
AD	Alzheimer's Disease
AAL	Anatomical Automatic Labeling
ADDTLA	Alzheimer's Disease Detection Empowered with Transfer Learning
ADNI	Alzheimer's Disease Neuroimaging Initiative
ADNI-GO	ADNI-Grand Opportunities
AE	Autoencoder
AI	Artificial Intelligent
AUC	Area Under the ROC Curve
aMCI	amnesic MCI
BOLD	Blood Oxygenation Level Dependent
CLAHE	Contrast Limited Adaptive Histogram Equalization
CAD	computer-aided diagnosis
CNN, ConvNet	Convolutional Neural Network
CN	Cognitively Normal
c-MCI	MCI converted to AD
CT	Computed Tomography
DBN	Deep Belief Network
DCNN	Deep CNN
DenseNet	Densely Connected Convolutional Network
DTI	Diffusion Tensor Imaging
DT	Decision Tree
DNN	Deep neural network
DL	Deep Learning
EMCI	Early MCI
ESSA	Enhanced Squirrel Search Algorithm
FL	Federated Learning
fMRI	functional Magnetic Resonance Imaging
GARD	The Gwangju Alzheimer's and Related Dementias
GRU	Gated Recurrent Units
HC	Healthy Control
IDLA	Improved Deep Learning Algorithm
KNN	K-Nearest Neighbor
LMCI	Late MCI
LSTM	Long short-term memory
LR	Logistic Regression
MAP	Rush Memory and Aging Project
MCC	Matthews's Correlation Coefficient
MCI	Mild Cognitive Impairment
MD	mild dementia
MoD	moderate dementia
MRI	Magnetic Resonance Imaging
NACC	National Alzheimer's Coordinating Center
NASNet	Neural Architecture Search
NB	Naïve Bayes
NC	Normal Cognition, Normal Control
ND	non-dementia
OASIS	Online Aerospace Supplier Information System
PET	Positron Emission Tomography
pMCI, MCIC	Progressive MCI (MCI patients who converted to AD)
PolyNet	Polymerization Network
ResNet	Residual Network
ReLU	Rectified Linear Unit
RBM	Restricted Boltzmann Machines
ResNeXt	Residual Networks with Lots of NeXt
RF	Random Forest
R-fMRI	Resting-state fMRI
ROI	Regions Of Interest
ROC	Receiver Operating Curve
RNN	Recurrent Neural Network
RVR	Relevance Vector Regression
SaaS	Software as a Service
SCNN	Siamese CNN
SMC	Subjective Memory Complaints
s-MCI, MCInc, sMCI	stable MCI (MCI patients who have not converted to AD)
SMOTE	Synthetic Minority Oversampling Technique
SVM	Support Vector Machine
sMRI	structural Magnetic Resonance Imaging
Tanh	Hyperbolic Tangent
VGG	Visual Geometry Group
VHFE	Voxel-based Hierarchical Function Extraction
VMD	very mild dementia
XAI	Explainable Artificial Intelligence
Xception	Extreme Inception
ZFNet	Zeiler and Fergus Network

Investigation, Formal analysis, Conceptualization. **Ahmed M. Fakh-rudeen:** Writing – review & editing, Visualization, Validation, Supervision, Software, Investigation, Formal analysis, Conceptualization. **Omar Younis Alani:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors would like to thank University of Salford _ UK and University of Kirkuk -Iraq for supporting and funding this research.

References

- [1] Saraiva C, Praça C, Ferreira R, Santos T, Ferreira L, Bernardino L. Nanoparticle-mediated brain drug delivery: Overcoming blood-brain barrier to treat neurodegenerative diseases. *J Contr Release* 2016;235:34–47. <https://doi.org/10.1016/J.JCONREL.2016.05.044>.
- [2] Chorawala MR, Shah AC, Pandya AJ, Kothari NR, Prajapati BG. Symptoms and conventional treatments of Alzheimer's disease. *Alzheimer's Disease and Advanced Drug Delivery Strategies* 2024;213–34. <https://doi.org/10.1016/B978-0-443-13205-6.00009-1>.
- [3] Dubois B, Picard G, Sarazin M. Early detection of Alzheimer's disease: new diagnostic criteria. *Dialogues Clin Neurosci* 2009;11(2):135–9. <https://doi.org/10.31887/DCNS.2009.11.2/BDUBOIS>.
- [4] Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. *Arch Neurol* 1999;56(3):303–8. <https://doi.org/10.1001/ARCHNEUR.56.3.303>.
- [5] Braak H, Braak E. Staging of alzheimer's disease-related neurofibrillary changes. *Neurobiol Aging* 1995;16(3):271–8. [https://doi.org/10.1016/0197-4580\(95\)00021-6](https://doi.org/10.1016/0197-4580(95)00021-6).
- [6] Mitchell AJ, Shiri-Feshki M. Temporal trends in the long term risk of progression of mild cognitive impairment: a pooled analysis. *J Neurol Neurosurg Psychiatry* 2008;79(12):1386–91. <https://doi.org/10.1136/JNNP.2007.142679>.
- [7] Shukla GP, Kumar S, Pandey SK, Agarwal R, Varshney N, Kumar A. Diagnosis and detection of alzheimer's disease using learning algorithm. *Big Data Mining and Analytics* 2023;6(4):504–12. <https://doi.org/10.26599/BDMA.2022.9020049>.
- [8] Hazarika RA, Maji AK, Sur SN, Paul BS, Kandar D. A survey on classification algorithms of brain images in alzheimer's disease based on feature extraction techniques. *IEEE Access* 2021;9:58503–36. <https://doi.org/10.1109/ACCESS.2021.3072559>.
- [9] ALZHEIMER'S ASSOCIATION REPORT. 2023 Alzheimer's disease facts and figures. *Alzheimer's Dementia* 2023;19:1598–695. <https://doi.org/10.1002/ALZ.13016>.
- [10] Jin D, Xu J, Zhao K, Hu F, Yang Z, Liu B, et al. Attention-based 3d convolutional network for Alzheimer's disease diagnosis and biomarkers exploration. In: *Proceedings - international Symposium on biomedical imaging*; 2019. p. 1047–51. <https://doi.org/10.1109/ISBI.2019.8759455>.
- [11] Han JY, Besser LM, Xiong C, Kukull WA, Morris JC. Cholinesterase inhibitors may not benefit mild cognitive impairment and mild Alzheimer disease dementia. *Alzheimer Dis Assoc Disord* 2019;33(2):87–94. <https://doi.org/10.1097/WAD.0000000000000291>.
- [12] Sundberg RJ. *The Chemical Century: Molecular Manipulation and its impact on the 20th Century*. Taylor & Francis; 2017. <https://doi.org/10.1080/00026980.2018.1469314>.
- [13] Keuck L. History as a biomedical matter: recent reassessments of the first cases of Alzheimer's disease. *Hist Philos Life Sci* 2018;40(1):1–26. <https://doi.org/10.1007/S40656-017-0177-7>.
- [14] Koh HK, Parekh AK. Toward a United States of health: implications of understanding the US Burden of disease. *JAMA* 2018;319(14):1438–40. <https://doi.org/10.1001/JAMA.2018.0157>.
- [15] Khan P, Kader MF, Islam SMR, Rahman AB, Kamal MS, Toha MU, et al. Machine learning and deep learning approaches for brain disease diagnosis: Principles and recent advances. *IEEE Access* 2021;9:37622–55. <https://doi.org/10.1109/ACCESS.2021.3062484>.
- [16] Pellegrini E, Ballerini L, Hernandez M, Chappell FM, González-Castro V, Anblagan D, et al. Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: a systematic review. *Alzheimer's Dementia: Diagnosis, Assessment & Disease Monitoring* 2018;10:519–35. <https://doi.org/10.1016/J.DADM.2018.07.004>.
- [17] Zhang L, Wang M, Liu M, Zhang D. A survey on deep learning for neuroimaging-based brain disorder analysis. *Front Neurosci* 2020;14:560709. <https://doi.org/10.3389/FNINS.2020.00779>.

- [18] Burgos N, Colliot O. Machine learning for classification and prediction of brain diseases: recent advances and upcoming challenges. *Curr Opin Neurol* 2020;33(4):439–50. <https://doi.org/10.1097/WCO.0000000000000838>.
- [19] Al-Shoukry S, Rassem TH, Makbol NM. Alzheimer's diseases detection by using deep learning algorithms: a mini-review. *IEEE Access* 2020;8:77131–41. <https://doi.org/10.1109/ACCESS.2020.2989396>.
- [20] Balne S, Elumalai A. Machine learning and deep learning algorithms used to diagnosis of Alzheimer's: review. *Mater Today Proc* 2021;47(15):5151–6. <https://doi.org/10.1016/J.MATPR.2021.05.499>.
- [21] Shetty A, Mehta D, Rane P, Dodani SN. Detection and Prediction of Alzheimer's disease using Deep learning: a review. In: 2021 international conference on Nascent Technologies in Engineering, ICNET 2021 - Proceedings. Institute of Electrical and Electronics Engineers Inc.; 2021. p. 1–7. <https://doi.org/10.1109/ICNET51185.2021.9487587>.
- [22] Gao S, Lima D. A review of the application of deep learning in the detection of Alzheimer's disease. *International Journal of Cognitive Computing in Engineering* 2022;3:1–8. <https://doi.org/10.1016/J.IJCC.2021.12.002>.
- [23] Patil V, Madgi M, Kiran A. Early prediction of Alzheimer's disease using convolutional neural network: a review. *Egyptian Journal of Neurology, Psychiatry and Neurosurgery* 2022;58(1):1–10. <https://doi.org/10.1186/S41983-022-00571-W>.
- [24] Arafat DA, Moustafa HED, Ali-Eldin AMT, Ali HA. Early detection of Alzheimer's disease based on the state-of-the-art deep learning approach: a comprehensive survey. *Multimed Tool Appl* 2022;81(17):23735–76. <https://doi.org/10.1007/s11042-022-11925-0>.
- [25] Sisodia PS, Ameta GK, Kumar Y, Chaplot N. A review of deep transfer learning approaches for class-wise prediction of alzheimer's disease using MRI images. *Arch Comput Methods Eng* 2023;30(4):2409–29. <https://doi.org/10.1007/S11831-022-09870-0>.
- [26] Deep Arya A, Singh Verma S, Chakrabarti P, Chakrabarti T, Elngar AA, Nami M, et al. A systematic review on machine learning and deep learning techniques in the effective diagnosis of alzheimer's disease. *Brain Inform* 2023;10(17). <https://doi.org/10.21203/RS.3.RS-2028945/V1>.
- [27] Muhammed Niyas KP, Thiagarajan P. A systematic review on early prediction of Mild cognitive impairment to Alzheimer's using machine learning algorithms. *International Journal of Intelligent Networks* 2023;4:74–88. <https://doi.org/10.1016/J.IJIN.2023.03.004>.
- [28] Shen J, Zhang CJP, Jiang B, Chen J, Song J, Liu Z, et al. Artificial intelligence versus Clinicians in disease diagnosis: systematic review. *JMIR Med Inform* 2019;7(3):e10010. <https://doi.org/10.2196/10010>.
- [29] Kumar DM, Satyanarayana D, Prasad MNG. MRI brain tumor detection using optimal possibilistic fuzzy C-means clustering algorithm and adaptive k-nearest neighbor classifier. *J Ambient Intell Hum Comput* 2021;12(2):2867–80. <https://doi.org/10.1007/S12652-020-02444-7>.
- [30] Noreen N, Palaniappan S, Qayyum A, Ahmad I, Imran M, Shoaib M. A deep learning model based on concatenation approach for the diagnosis of brain tumor. *IEEE Access* 2020;8:55135–44. <https://doi.org/10.1109/ACCESS.2020.2978629>.
- [31] Geetha M, Srinadh V, Janet J, Sumathi S. Hybrid Archimedes Sine Cosine optimization enabled Deep Learning for multilevel brain tumor classification using MRI images. *Biomed Signal Process Control* 2024;PartB(87):105419. <https://doi.org/10.1016/J.BSPC.2023.105419>.
- [32] Ramirez VM, Kmetzsch V, Forbes F, Dojat M. Deep learning models to study the early stages of Parkinson's disease. In: Proceedings - international Symposium on biomedical imaging. IEEE Computer Society; 2020. p. 1534–7. <https://doi.org/10.1109/ISBI45749.2020.9098529>.
- [33] Tian Z yao, Qian L, Fang L, Peng X hua, Zhu X hu, Wu M, et al. Frequency-specific changes of resting brain activity in Parkinson's disease: a machine learning approach. *Neuroscience* 2020;436:170–83. <https://doi.org/10.1016/J.NEUROSCIENCE.2020.01.049>.
- [34] Amprimo G, Masi G, Olmo G, Ferraris C. Deep Learning for hand tracking in Parkinson's Disease video-based assessment: current and future perspectives. *Artif Intell Med* 2024;154:102914. <https://doi.org/10.1016/J.ARTMED.2024.102914>.
- [35] Ahmed MS, Fakhrudeen AM. Deep learning-based COVID-19 detection: state-of-the-art in research. *Int J Nonlinear Anal Appl* 2023;14(1):1939–62. <https://doi.org/10.22075/IJNAA.2022.7119>.
- [36] Ahmed MS, Fakhrudeen AM. COVID-19IraqKirkukDataset: development and evaluation of an Iraqi dataset for COVID-19 classification based on deep learning. *Int J Nonlinear Anal Appl* 2023;14:2507–18. <https://doi.org/10.22075/IJNAA.2023.7317>.
- [37] Nawaz M, Saleem S, Masood M, Rashid J, Nazir T. COVID-ECG-RSNet: COVID-19 classification from ECG images using swish-based improved ResNet model. *Biomed Signal Process Control* 2024;89:105801. <https://doi.org/10.1016/J.BSPC.2023.105801>.
- [38] Beheshti I, Ganaie MA, Paliwal V, Rastogi A, Razzak I, Tanveer M. Predicting brain age using machine learning algorithms: a comprehensive evaluation. *IEEE J Biomed Health Inform* 2022;26(4):1432–40. <https://doi.org/10.1109/JBHI.2021.3083187>.
- [39] Barragán-Montero A, Javaid U, Valdés G, Nguyen D, Desbordes P, Macq B, et al. Artificial intelligence and machine learning for medical imaging: a technology review. *Phys Med* 2021;83:242–56. <https://doi.org/10.1016/J.EJMP.2021.04.016>.
- [40] Malik MM. A Hierarchy of Limitations in Machine Learning. *ArXiv Preprint ArXiv: 200205193* 2020. <https://doi.org/10.48550/arXiv.2002.05193>.
- [41] Ahsan MM, Luna SA, Siddique Z. Machine-learning-based disease diagnosis: a comprehensive review. *Healthcare* 2022;10(3):541. <https://doi.org/10.3390/HEALTHCARE10030541>.
- [42] Ian Goodfellow, Yoshua Bengio, Aaron Courville. Deep learning. MIT Press; 2016. <https://www.deeplearningbook.org/>. [Accessed 21 June 2024].
- [43] Sarvamangala DR, Kulkarni RV. Convolutional neural networks in medical image understanding: a survey. *Evol Intell* 2022;15(1). <https://doi.org/10.1007/s12065-020-00540-3>.
- [44] Ebrahimighahnavieh MA, Luo S, Chiong R. Deep learning to detect Alzheimer's disease from neuroimaging: a systematic literature review. *Comput Methods Progr Biomed* 2020;187:105242. <https://doi.org/10.1016/J.CMPB.2019.105242>.
- [45] Sharma S, Sharma S, Athaiya A. Activation functions in neural networks. *International Journal of Engineering Applied Sciences and Technology* 2020;4(12):310–6. <https://doi.org/10.33564/IJEAST.2020.v04i12.054>.
- [46] Ebrahim D, Ali-Eldin AMT, Moustafa HE, Arafat H. Alzheimer disease early detection using convolutional neural networks. *Proceedings of ICCES 2020 - 2020 15th International Conference on computer Engineering and systems. Institute of Electrical and Electronics Engineers Inc.*; 2020. p. 1–6. <https://doi.org/10.1109/ICCES51560.2020.9334594>.
- [47] Albawi S, Mohammed TA, Al-Zawi S. Understanding of a convolutional neural network. *Proceedings of 2017 international Conference on Engineering and technology, ICET 2017. Institute of Electrical and Electronics Engineers Inc.*; 2018. p. 1–6. <https://doi.org/10.1109/ICENGTECHNOL.2017.8308186>.
- [48] Han D, Liu Q, Fan W. A new image classification method using CNN transfer learning and web data augmentation. *Expert Syst Appl* 2018;95:43–56. <https://doi.org/10.1016/J.ESWA.2017.11.028>.
- [49] Weiss K, Khoshgoftaar TM, Wang DD. A survey of transfer learning. *J Big Data* 2016;3(1):1–40. <https://doi.org/10.1186/S40537-016-0043-6>.
- [50] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *ArXiv Preprint ArXiv:14091556*, <https://doi.org/10.48550/arXiv.1409.1556>; 2014.
- [51] LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE* 1998;86:2278–323. <https://doi.org/10.1109/5.726791>.
- [52] Sarraf S, Tofghi G, Org S. Classification of alzheimer's disease structural MRI data by deep learning convolutional neural networks. *ArXiv Preprint ArXiv: 160706583* 2016. <https://doi.org/10.48550/arXiv.1607.06583>.
- [53] Szegegy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. *Proceedings of the IEEE computer Society Conference on computer vision and pattern recognition. IEEE Computer Society*; 2014. p. 1–9. <https://doi.org/10.1109/CVPR.2015.7298594>.
- [54] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE computer Society conference on computer vision and pattern recognition. IEEE Computer Society*; 2015. p. 770–8. <https://doi.org/10.1109/CVPR.2016.90>.
- [55] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM* 2017;60(6):84–90. <https://doi.org/10.1145/3065386>.
- [56] Singh SP, Wang L, Gupta S, Goli H, Padmanabhan P, Gulyás B. 3D deep learning on medical images: a review. *Sensors* 2020;20(18):5097. <https://doi.org/10.3390/S20185097>.
- [57] Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. *Insights Imaging* 2018;9(4):611–29. <https://doi.org/10.1007/S13244-018-0639-9>.
- [58] Khan A, Sohail A, Zahoor U, Qureshi AS. A survey of the recent architectures of deep convolutional neural networks. *Artif Intell Rev* 2020;53(8):5455–516. <https://doi.org/10.1007/S10462-020-09825-6>.
- [59] Tan M, Le QV. EfficientNet: Rethinking model scaling for convolutional neural networks. In: *Proceedings of the 36th international conference on machine learning. PMLR*; 2019. p. 6105–14. <https://doi.org/10.48550/arXiv.1905.11946>.
- [60] Srivastava S, Divekar AV, Anilkumar C, Naik I, Kulkarni V, Pattabiraman V. Comparative analysis of deep learning image detection algorithms. *J Big Data* 2021;8(1):1–27. <https://doi.org/10.1186/S40537-021-00434-W>.
- [61] Zoph B, Vasudevan V, Shlens J, Le QV. Learning transferable architectures for scalable image recognition. In: *Proceedings of the IEEE computer Society conference on computer vision and pattern recognition. IEEE Computer Society*; 2018. p. 8697–710. <https://doi.org/10.1109/CVPR.2018.00907>.
- [62] Aaraji ZSh, Abbas HH. Automatic classification of alzheimer's disease using brain MRI data and deep convolutional neural networks. *ArXiv Preprint ArXiv: 220400068* 2022. <https://doi.org/10.48550/arXiv.2204.00068>.
- [63] Hoang VT, Jo KH. Practical analysis on architecture of EfficientNet. *International Conference on human system Interaction, HSI. IEEE Computer Society*; 2021. p. 1–4. <https://doi.org/10.1109/HSI52170.2021.9538782>.
- [64] Prakash Shukla S, Kumar Sharma L, Kiran Gulhare K, Shukla SP, Sharma LK. Deep neural network classification method to alzheimer's disease detection. *Int J Adv Res Comput Sci Software Eng* 2017;7(6):1–4. <https://doi.org/10.13140/RG.2.2.31207.75687>.
- [65] JayaLakshmi ANM, Krishna Kishore KV. Performance evaluation of DNN with other machine learning techniques in a cluster using Apache Spark and MLlib. *Journal of King Saud University - Computer and Information Sciences* 2022;34(1):1311–9. <https://doi.org/10.1016/J.JKSUCI.2018.09.022>.
- [66] Noor MBT, Zenia NZ, Kaiser MS, Mamun S AL, Mahmud M. Application of deep learning in detecting neurological disorders from magnetic resonance images: a survey on the detection of Alzheimer's disease, Parkinson's disease and schizophrenia. *Brain Inform* 2020;7(1):1–21. <https://doi.org/10.1186/S40708-020-00112-2>.

- [67] Ming Y, Cao S, Zhang R, Li Z, Chen Y, Song Y, et al. Understanding hidden Memories of recurrent neural networks. In: 2017 IEEE conference on visual Analytics science and technology, VAST 2017 - Proceedings. Institute of Electrical and Electronics Engineers Inc.; 2018. p. 13–24. <https://doi.org/10.1109/VAST.2017.8585721>.
- [68] Peng J, Guan J, Shang X. Predicting Parkinson's disease genes based on node2vec and autoencoder. *Front Genet* 2019;10:441295. <https://doi.org/10.3389/FGENE.2019.00226>.
- [69] Jo T, Nho K, Saykin AJ. Deep learning in alzheimer's disease: diagnostic classification and prognostic prediction using neuroimaging data. *Front Aging Neurosci* 2019;11:457612. <https://doi.org/10.3389/FNAGI.2019.00220>.
- [70] Pereira L, Pinto MP, Shah K, Khan S. Show and tell: a neural visual story-teller. In: Proceedings of the 3rd international conference on advances in science & technology (ICAST). Elsevier BV; 2020. <https://doi.org/10.2139/SSRN.3565282>.
- [71] Jain A, Zamir AR, Savarese S, Saxena A. Structural-RNN: deep learning on spatio-temporal graphs. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 5308–17. <https://doi.org/10.1109/CVPR.2016.573>.
- [72] Schmidt RM. Recurrent neural networks (RNNs): a gentle introduction and overview. *ArXiv Preprint ArXiv:191205911* 2019. <https://doi.org/10.48550/arXiv.1912.05911>.
- [73] Liu S, Liu S, Cai W, Pujol S, Kikinis R, Feng D. Early diagnosis of Alzheimer's disease with deep learning. In: 2014 IEEE 11th international symposium on biomedical imaging, ISBI 2014. Institute of Electrical and Electronics Engineers Inc.; 2014. p. 1015–8. <https://doi.org/10.1109/ISBI.2014.6868045>.
- [74] DiPietro R, Hager GD. Deep learning: RNNs and LSTM. Academic Press; 2020. <https://doi.org/10.1016/B978-0-12-816176-0.00026-0>.
- [75] Zhao H, Sun S, Jin B. Sequential Fault diagnosis based on LSTM neural network. *IEEE Access* 2018;6:12929–39. <https://doi.org/10.1109/ACCESS.2018.2794765>.
- [76] Cheng H, Yuan S, Li W, Yu X, Liu F, Liu X, et al. De-accumulated error collaborative learning framework for predicting Alzheimer's disease progression. *Biomed Signal Process Control* 2024;89:105767. <https://doi.org/10.1016/J.BSPC.2023.105767>.
- [77] Pathirage CSN, Li J, Li L, Hao H, Liu W, Ni P. Structural damage identification based on autoencoder neural networks and deep learning. *Eng Struct* 2018;172:13–28. <https://doi.org/10.1016/J.ENGSTRUCT.2018.05.109>.
- [78] Aamir M, Nawi NM, Mahdin H Bin, Naseem R, Zulqarnain M. Auto-encoder variants for solving handwritten digits classification problem. *International Journal of Fuzzy Logic and Intelligent Systems* 2020;20(1):8–16. <https://doi.org/10.5391/IJFIS.2020.20.1.8>.
- [79] Sun W, Shao S, Zhao R, Yan R, Zhang X, Chen X. A sparse auto-encoder-based deep neural network approach for induction motor faults classification. *Measurement* 2016;89:171–8. <https://doi.org/10.1016/J.MEASUREMENT.2016.04.007>.
- [80] Chaitanya CRA, Kaplanyan AS, Schied C, Salvi M, Lefohn A, Nowrouzezahrai D, et al. Interactive reconstruction of Monte Carlo image sequences using a recurrent denoising autoencoder. *ACM Trans Graph* 2017;36(4):1–12. <https://doi.org/10.1145/3072959.3073601>.
- [81] Chen Z, Yang D, Cui T, Li D, Liu H, Yang Y, et al. A novel imbalanced dataset mitigation method and ECG classification model based on combined 1D_CBAM-autoencoder and lightweight CNN model. *Biomed Signal Process Control* 2024;87 (PartB):105437. <https://doi.org/10.1016/J.BSPC.2023.105437>.
- [82] Morabito FC, Campolo M, Ieracitano C, Mammone N. Deep learning approaches to electrophysiological multivariate time-series analysis. Academic Press; 2019. <https://doi.org/10.1016/B978-0-12-815480-9.00011-6>.
- [83] An N, Jin L, Ding H, Yang J, Yuan J. A deep belief network-based method to identify proteomic risk markers for Alzheimer disease. *ArXiv Preprint ArXiv: 200305776*. <https://doi.org/10.48550/arXiv.2003.05776>; 2020.
- [84] Hill NL, Mogle J. Alzheimer's disease risk factors as mediators of subjective memory impairment and objective memory decline: protocol for a construct-level replication analysis 17 *Psychology and Cognitive Sciences* 1701 *Psychology. BMC Geriatr* 2018;18(1):1–8. <https://doi.org/10.1186/S12877-018-0954-5>.
- [85] Turek JJ, Lafayette W, Jeong K. Method and apparatus for motility contrast imaging. US Patent 2014. 8,886,295, <https://patents.google.com/patent/US8886295B2/en>. [Accessed 12 September 2023].
- [86] Makarets SJ, Quimby M, Collins J, Makris N, McGinnis S, Schultz A, et al. Flortaucipir tau PET imaging in semantic variant primary progressive aphasia. *J Neurol Neurosurg Psychiatry* 2018;89(10):1024–31. <https://doi.org/10.1136/JNPP.2017.316409>.
- [87] Waser M, Benke T, Dal-Bianco P, Garn H, Mosbacher JA, Ransmayr G, et al. Neuroimaging markers of global cognition in early Alzheimer's disease: a magnetic resonance imaging-electroencephalography study. *Brain Behav* 2019;9 (1):e01197. <https://doi.org/10.1002/BRB3.1197>.
- [88] Trojchanec K, Kitanovski I, Dimitrovski I, Loshkovska S. Longitudinal brain MRI retrieval for alzheimer's disease using different temporal information. *IEEE Access* 2017;6:9703–12. <https://doi.org/10.1109/ACCESS.2017.2773359>.
- [89] jith Tho D. Gaillard F. 1.5 T vs 3.0 T. *RadiopaediaOrg*. <https://doi.org/10.5334/7/RID-801>; 2008.
- [90] Sormaala MJ, Ruohola JP, Mattila VM, Koskinen SK, Pihlajamäki HK. Comparison of 1.5T and 3T MRI scanners in evaluation of acute bone stress in the foot. *BMC Musculoskel Disord* 2011;12(1):1–6. <https://doi.org/10.1186/1471-2474-12-128>.
- [91] Schick F, Pieper CC, Kupczyk P, Almansour H, Keller G, Springer F, et al. 1.5 vs 3 Tesla magnetic resonance imaging A review of favorite clinical applications for both field strengths—Part 1. *Invest Radiol* 2021;56(11):680–91. <https://doi.org/10.1097/RLI.0000000000000812>.
- [92] Chandra A, Dervenoulas G, Politis M. Magnetic resonance imaging in Alzheimer's disease and mild cognitive impairment. *J Neurol* 2019;266(6):1293–302. <https://doi.org/10.1007/S00415-018-9016-3>.
- [93] Frizzell TO, Glashutter M, Liu CC, Zeng A, Pan D, Hajra SG, et al. Artificial intelligence in brain MRI analysis of Alzheimer's disease over the past 12 years: a systematic review. *Ageing Res Rev* 2022;77:101614. <https://doi.org/10.1016/J.ARR.2022.101614>.
- [94] Varvatsoulas G. Voice-sensitive areas in the brain: a single participant study coupled with brief evolutionary psychological considerations. *Psychol Thought* 2014;7(1):66–79. <https://doi.org/10.23668/psycharchives.1948>.
- [95] Villemagne VL, Doré V, Burnham SC, Masters CL, Rowe CC. Imaging tau and amyloid- β proteinopathies in Alzheimer disease and other conditions. *Nat Rev Neurol* 2018;14(4):225–36. <https://doi.org/10.1038/nrneurol.2018.9>.
- [96] Drzezga A, Altomare D, Festari C, Arbizu J, Orini S, Herholz K, et al. Diagnostic utility of 18F-Fluorodeoxyglucose positron emission tomography (FDG-PET) in asymptomatic subjects at increased risk for Alzheimer's disease. *Eur J Nucl Med Mol Imag* 2018;45(9):1487–96. <https://doi.org/10.1007/S00259-018-4032-1>.
- [97] Salem Salamh AB, Salamah AA, Akyüz HI. A study of a new technique of the CT scan view and disease classification protocol based on level challenges in cases of coronavirus disease. *Radiol Res Pract* 2021;2021(1):1–9. <https://doi.org/10.1155/2021/5554408>.
- [98] Scheltens P, Kuiper M, Ch Wolters E, Barkhof F, Valk J, Weinstein HC, et al. Atrophy of medial temporal lobes on MRI in "probable" Alzheimer's disease and normal ageing: diagnostic value and neuropsychological correlates. *J Neurol Neurosurg Psychiatry* 1992;55(10):967–72. <https://doi.org/10.1136/JNPP.55.10.967>.
- [99] Johnson KA, Fox NC, Sperling RA, Klunk WE. Brain imaging in alzheimer disease. *Cold Spring Harb Perspect Med* 2012;2(4):a006213. <https://doi.org/10.1101/CSHPERSPECT.A006213>.
- [100] Zhang Y, Schuff N, Jahng GH, Bayne W, Mori S, Schad L, et al. Diffusion tensor imaging of cingulum fibers in mild cognitive impairment and Alzheimer disease. *Neurology* 2007;68(1):13–9. <https://doi.org/10.1212/01.WNL.0000250326.77323.01>.
- [101] Knight MJ, McCann B, Kauppinen RA, Coulthard EJ. Magnetic resonance imaging to detect early molecular and cellular changes in Alzheimer's disease. *Front Aging Neurosci* 2016;8:139. <https://doi.org/10.3389/FNAGI.2016.00139>.
- [102] Valotassiou V, Malamitsi J, Papatriantafyllou J, Dardiotis E, Tsougos I, Psimadas D, et al. SPECT and PET imaging in Alzheimer's disease. *Ann Nucl Med* 2018;32(9):583–93. <https://doi.org/10.1007/S12149-018-1292-6>.
- [103] Mosconi L, Walters M, Sterling J, Quinn C, McHugh P, Andrews RE, et al. Lifestyle and vascular risk effects on MRI-based biomarkers of Alzheimer's disease: a cross-sectional study of middle-aged adults from the broader New York City area. *BMJ Open* 2018;8(3):e019362. <https://doi.org/10.1136/BMJOPEN-2017-019362>.
- [104] Mussap M, Noto A, Cibecchini F, Fanos V. The importance of biomarkers in neonatology. *Semin Fetal Neonatal Med* 2013;18(1):56–64. <https://doi.org/10.1016/J.SINY.2012.10.006>.
- [105] Cedazo-Minguez A, Winblad B. Biomarkers for Alzheimer's disease and other forms of dementia: clinical needs, limitations and future aspects. *Exp Gerontol* 2010;45(1):5–14. <https://doi.org/10.1016/J.EXGER.2009.09.008>.
- [106] Solano-Rojas B, Villalón-Fonseca R. A low-cost three-dimensional DenseNet neural network for alzheimer's disease early discovery. *Sensors* 2021;21(4):1302. <https://doi.org/10.3390/S21041302>.
- [107] ADNI | About n.d. <https://adni.loni.usc.edu/about/> (accessed September 15, 2023).
- [108] ADNI | ACCESS DATA n.d. <https://adni.loni.usc.edu/data-samples/access-data/> (accessed September 25, 2023).
- [109] Marcus DS, Wang TH, Parker J, Csernansky JG, Morris JC, Buckner RL. Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *J Cognit Neurosci* 2007; 19(9):1498–507. <https://doi.org/10.1162/JOCN.2007.19.9.1498>.
- [110] OASIS Brains - Open Access Series of Imaging Studies n.d. <https://www.oasis-brains.org/#about> (accessed September 16, 2023).
- [111] Alabayati MHA. Early detection of Alzheimer's disease using machine learning algorithms. Altınbaş University Institutional Repository; 2022. <http://openaccess.altinbas.edu.tr/xmlui/handle/20.500.12939/3902>. [Accessed 16 September 2023].
- [112] McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's Dementia* 2011;7(3):263–9. <https://doi.org/10.1016/J.JALZ.2011.03.005>.
- [113] About NACC | National Alzheimer's Coordinating Center n.d. <https://naccdata.org/nacc-collaborations/about-nacc> (accessed June 22, 2024).
- [114] Katabathula S, Wang Q, Xu R. Predict Alzheimer's disease using hippocampus MRI data: a lightweight 3D deep convolutional network model with visual and global shape representations. *Alzheimer's Res Ther* 2021;13(1):1–9. <https://doi.org/10.1186/S13195-021-00837-0>.
- [115] AIBL Study ADNI Data n.d. <https://www.aibl.csiro.au/adni/index.html> (accessed September 16, 2023).
- [116] Fowler C, Rainey-Smith SR, Bird S, Bomke J, Bourgeat P, Brown BM, et al. Fifteen years of the Australian imaging, biomarkers and lifestyle (AIBL) study: progress and observations from 2,359 older adults spanning the spectrum from cognitive normality to alzheimer's disease. *J Alzheimers Dis Rep* 2021;5(1):443–68. <https://doi.org/10.3233/ADR-210005>.
- [117] Gupta Y, Lee KH, Choi KY, Lee JJ, Kim BC, Kwon GR. Early diagnosis of Alzheimer's disease using combined features from voxel-based morphometry and

- cortical, subcortical, and hippocampus regions of MRI T1 brain images. *PLoS One* 2019;14(10):e0222446. <https://doi.org/10.1371/JOURNAL.PONE.0222446>.
- [118] Kang S, Gim J, Lee J, Gunasekaran TI, Choi KY, Lee JJ, et al. Potential novel genes for late-onset alzheimer's disease in east-asian descent identified by APOE-stratified genome-wide association study. *J Alzheim Dis* 2021;82(4):1451–60. <https://doi.org/10.3233/JAD-210145>.
- [119] Park JY, Lee D, Lee JJ, Gim J, Gunasekaran TI, Choi KY, et al. A missense variant in SHARPIN mediates Alzheimer's disease-specific brain damages. *Transl Psychiatry* 2021;11(1):1–9. <https://doi.org/10.1038/s41398-021-01680-5>.
- [120] Malone IB, Cash D, Ridgway GR, MacManus DG, Ourselin S, Fox NC, et al. MIRIAD—public release of a multiple time point Alzheimer's MR imaging dataset. *Neuroimage* 2013;70:33–6. <https://doi.org/10.1016/J.NEUROIMAGE.2012.12.044>.
- [121] Chu X, Ilyas IF, Krishnan S, Wang J. Data cleaning: overview and emerging challenges. In: *Proceedings of the ACM SIGMOD international conference on management of data*, vol. 26. Association for Computing Machinery; 2016. p. 2201–6. <https://doi.org/10.1145/2882903.2912574>. June-2016.
- [122] Jo T, Nho K, Saykin AJ. Deep learning in alzheimer's disease: diagnostic classification and prognostic prediction using neuroimaging data. *Front Aging Neurosci* 2019;11:457612. <https://doi.org/10.3389/FNAGI.2019.00220>.
- [123] Challis E, Hurley P, Serra L, Bozzali M, Oliver S, Cercignani M. Gaussian process classification of Alzheimer's disease and mild cognitive impairment from resting-state fMRI. *Neuroimage* 2015;112:232–43. <https://doi.org/10.1016/J.NEUROIMAGE.2015.02.037>.
- [124] Lu D, Popuri K, Ding GW, Balachandrar R, Beg MF, Weiner M, et al. Multimodal and multiscale deep neural networks for the early diagnosis of alzheimer's disease using structural MR and FDG-PET images. *Sci Rep* 2018;8(1):1–13. <https://doi.org/10.1038/s41598-018-22871-z>. 2018;8(1).
- [125] Khademi A, Reiche B, DiGregorio J, Arezza G, Moody AR. Whole volume brain extraction for multi-centre, multi-disease FLAIR MRI datasets. *Magn Reson Imaging* 2020;66:116–30. <https://doi.org/10.1016/J.MRI.2019.08.022>.
- [126] Ortiz A, Munilla J, Górriz JM, Ramírez J. Ensembles of deep learning architectures for the early diagnosis of the alzheimer's disease. *Int J Neural Syst* 2016;26(7). <https://doi.org/10.1142/S0129065716500258>.
- [127] Subramoniam M, Aparna TR, Anurenjan PR, Sreeni KG. Deep learning-based prediction of alzheimer's disease from magnetic resonance images. *Intelligent Vision in Healthcare* 2022;145–51. https://doi.org/10.1007/978-981-16-7771-7_12.
- [128] Helaly HA, Badawy M, Haikal AY. Toward deep MRI segmentation for Alzheimer's disease detection. *Neural Comput Appl* 2022;34(2):1047–63. <https://doi.org/10.1007/S00521-021-06430-8>.
- [129] Cenek M, Hu M, York G, Dahl S. Survey of image processing techniques for brain pathology diagnosis: challenges and opportunities. *Frontiers Robotics AI* 2018;5:405977. <https://doi.org/10.3389/FROBT.2018.00120>.
- [130] Goshu YY, Kitaw D. Performance measurement and its recent challenge: a literature review. *Int J Bus Perform Manag* 2017;18(4):381–402. <https://doi.org/10.1504/IJBPM.2017.087103>.
- [131] Farooq A, Anwar S, Awais M, Rehman S. A deep CNN based multi-class classification of Alzheimer's disease using MRI. In: *Ist 2017 - IEEE international conference on imaging systems and techniques*, proceedings, vol. 2018- january. Institute of Electrical and Electronics Engineers Inc.; 2017. p. 1–6. <https://doi.org/10.1109/IST.2017.8261460>.
- [132] Liu M, Cheng D, Wang K, Wang Y. Multi-modality cascaded convolutional neural networks for alzheimer's disease diagnosis. *Neuroinformatics* 2018;16:295–308. <https://doi.org/10.1007/S12021-018-9370-4>.
- [133] Islam J, Zhang Y. Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks. *Brain Inform* 2018;5(2):1–14. <https://doi.org/10.1186/S40708-018-0080-3>.
- [134] Basaia S, Agosta F, Wagner L, Canu E, Magnani G, Santangelo R, et al. Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. *Neuroimage Clin* 2019;21:101645. <https://doi.org/10.1016/J.NICL.2018.101645>.
- [135] Yue L, Gong X, Li J, Ji H, Li M, Nandi AK. Hierarchical feature extraction for early Alzheimer's disease diagnosis. *IEEE Access* 2019;7:93752–60. <https://doi.org/10.1109/ACCESS.2019.2926288>.
- [136] Gorji HT, Kaabouch N. A deep learning approach for diagnosis of mild cognitive impairment based on MRI images. *Brain Sci* 2019;9(9):217. <https://doi.org/10.3390/BRANSI9090217>.
- [137] Spasov S, Passamonti L, Duggento A, Liò P, Toschi N. A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease. *Neuroimage* 2019;189:276–87. <https://doi.org/10.1016/J.NEUROIMAGE.2019.01.031>.
- [138] Khan NM, Abraham N, Hon M. Transfer learning with intelligent training data selection for prediction of alzheimer's disease. *IEEE Access* 2019;7:72726–35. <https://doi.org/10.1109/ACCESS.2019.2920448>.
- [139] Mehmood A, Maqsood M, Bashir M, Shuyuan Y. A deep siamese convolution neural network for multi-class classification of alzheimer disease. *Brain Sci* 2020;10(2):84. <https://doi.org/10.3390/BRANSI10020084>.
- [140] Suganth RC, Latha RS, Geetha M, Sreekanth GR. Diagnosis of alzheimer's disease from brain magnetic resonance imaging images using deep learning algorithms. *Adv Electr Comput Eng* 2020;20(3):57–64. <https://doi.org/10.4316/AECE.2020.03007>.
- [141] Abrol A, Bhattarai M, Fedorov A, Du Y, Pliis S, Calhoun V. Deep residual learning for neuroimaging: an application to predict progression to Alzheimer's disease. *J Neurosci Methods* 2020;339:108701. <https://doi.org/10.1016/J.JNEUMETH.2020.108701>.
- [142] Guo H, Zhang Y. Resting state fMRI and improved deep learning algorithm for earlier detection of alzheimer's disease. *IEEE Access* 2020;8:115383–92. <https://doi.org/10.1109/ACCESS.2020.3003424>.
- [143] Liu M, Li F, Yan H, Wang K, Ma Y, Shen L, et al. A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease. *Neuroimage* 2020;208:116459. <https://doi.org/10.1016/J.NEUROIMAGE.2019.116459>.
- [144] Kundaram SS, Pathak KC. Deep learning-based alzheimer disease detection. *Lecture Notes in Electrical Engineering* 2021;673:587–97. https://doi.org/10.1007/978-981-15-5546-6_50.
- [145] Raghavaiah P, Varadarajan S. A CAD system design to diagnosize alzheimers disease from MRI brain images using optimal deep neural network. *Multimed Tool Appl* 2021;80(17):26411–28. <https://doi.org/10.1007/S11042-021-10928-7>.
- [146] Hazarika RA, Abraham A, Kandar D, Maji AK. An improved LeNet-deep neural network model for alzheimer's disease classification using brain magnetic resonance images. *IEEE Access* 2021;9:161194–207. <https://doi.org/10.1109/ACCESS.2021.3131741>.
- [147] Odusami M, Maskeliūnas R, Damaševičius R, Krilavičius T. 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* 2021;11(6):1071. <https://doi.org/10.3390/DIAGNOSTICS11061071>.
- [148] Mehmood A, yang S, feng Z, wang M, Ahmad AS, khan R, et al. A transfer learning approach for early diagnosis of alzheimer's disease on MRI images. *Neuroscience* 2021;460:43–52. <https://doi.org/10.1016/J.NEUROSCIENCE.2021.01.002>.
- [149] Murugan S, Venkatesan C, Sumithra MG, Gao XZ, Elakkiya B, Akila M, et al. DEMNET: a deep learning model for early diagnosis of alzheimer diseases and dementia from MR images. *IEEE Access* 2021;9:90319–29. <https://doi.org/10.1109/ACCESS.2021.3090474>.
- [150] Basher A, Kim BC, Lee KH, Jung HY. Volumetric feature-based alzheimer's disease diagnosis from sMRI data using a convolutional neural network and a deep neural network. *IEEE Access* 2021;9:29870–82. <https://doi.org/10.1109/ACCESS.2021.3059658>.
- [151] Zeng N, Li H, Peng Y. A new deep belief network-based multi-task learning for diagnosis of Alzheimer's disease. *Neural Comput Appl* 2023;35(16):11599–610. <https://doi.org/10.1007/S00521-021-06149-6>.
- [152] Muhammed Raees PC, Thomas V. Automated detection of alzheimer's disease using deep learning in MRI. *J Phys Conf Ser* 2021;1921(1). <https://doi.org/10.1088/1742-6596/1921/1/012024>.
- [153] Helaly HA, Badawy M, Haikal AY. Deep learning approach for early detection of alzheimer's disease. *Cognit Comput* 2022;14(5):1711–27. <https://doi.org/10.1007/S12559-021-09946-2>.
- [154] Sharma S, Guleria K, Tiwari S, Kumar S. A deep learning based convolutional neural network model with VGG16 feature extractor for the detection of Alzheimer Disease using MRI scans. *Measurement: Sensors* 2022;24:100506. <https://doi.org/10.1016/J.MEASEN.2022.100506>.
- [155] Odusami M, Maskeliūnas R, Damaševičius R. An intelligent system for early recognition of alzheimer's disease using neuroimaging. *Sensors* 2022;22(3):740. <https://doi.org/10.3390/S22030740>.
- [156] Orouskhani M, Zhu C, Rostamian S, Shomal Zadeh F, Shafiei M, Orouskhani Y. Alzheimer's disease detection from structural MRI using conditional deep triplet network. *Neuroscience Informatics* 2022;2(4):100066. <https://doi.org/10.1016/J.NEURI.2022.100066>.
- [157] Savaş S. Detecting the stages of alzheimer's disease with pre-trained deep learning architectures. *Arabian J Sci Eng* 2022;47(2):2201–18. <https://doi.org/10.1007/S13369-021-06131-3>.
- [158] Lu B, Li HX, Chang ZK, Li L, Chen NX, Zhu ZC, et al. A practical Alzheimer's disease classifier via brain imaging-based deep learning on 85,721 samples. *J Big Data* 2022;9(1):1–22. <https://doi.org/10.1186/S40537-022-00650-Y>.
- [159] Ghazal TM, Abbas S, Munir S, Khan MA, Ahmad M, Issa GF, et al. Alzheimer disease detection empowered with transfer learning. *Comput Mater Continua (CMC)* 2022;70(3):5005–19. <https://doi.org/10.32604/CMC.2022.020866>.
- [160] Rezaee K, Khosravi MR, Sabri M, Al-Qawasmki K. A hybrid deep cascade-ResNet model for detecting alzheimer's stages in MR images. *1st international engineering conference on electrical, energy, and artificial intelligence, EICEEAI 2022*. Institute of Electrical and Electronics Engineers Inc.; 2022. p. 1–6. <https://doi.org/10.1109/EICEEAI56378.2022.10050454>.
- [161] Goenka N, Goenka A, Tiwari S. Patch-based classification for alzheimer disease using sMRI. In: *International conference on emerging smart computing and informatics, ESCI 2022*. Institute of Electrical and Electronics Engineers Inc.; 2022. p. 1–5. <https://doi.org/10.1109/ESCI53509.2022.9758317>.
- [162] Shamrat FMJM, Akter S, Azam S, Karim A, Ghosh P, Tasnim Z, et al. AlzheimerNet: an effective deep learning based proposition for alzheimer's disease stages classification from functional brain changes in magnetic resonance images. *IEEE Access* 2023;11:16376–95. <https://doi.org/10.1109/ACCESS.2023.3244952>.
- [163] EL-Geneedy M, Moustafa HED, Khalifa F, Khater H, Abdelhalim E. An MRI-based deep learning approach for accurate detection of Alzheimer's disease. *Alex Eng J* 2023;63:211–21. <https://doi.org/10.1016/J.AEJ.2022.07.062>.
- [164] Sethuraman SK, Malaiyappan N, Ramalingam R, Basheer S, Rashid M, Ahmad N. Predicting alzheimer's disease using deep neuro-functional networks with resting-state fMRI. *Electronics* 2023;12(4):1031. <https://doi.org/10.3390/ELECTRONICS12041031>.
- [165] Mohi ud din dar G, Bhagat A, Ansarullah SI, Othman MT Ben, Hamid Y, Alkahtani HK, et al. A novel framework for classification of different alzheimer's

- disease stages using CNN model. *Electronics* 2023;12(2):469. <https://doi.org/10.3390/ELECTRONICS12020469>.
- [166] Ghosh T, Palash MIA, Yousuf MA, Hamid MdA, Monowar MM, Allassafi MO. A robust distributed deep learning approach to detect alzheimer's disease from MRI images. *Mathematics* 2023;11(12):2633. <https://doi.org/10.3390/MATH11122633>.
- [167] Odusami M, Maskeliūnas R, Damaševičius R, Misra S. Explainable deep-learning-based diagnosis of alzheimer's disease using multimodal input fusion of PET and MRI images. *J Med Biol Eng* 2023;43(3):291–302. <https://doi.org/10.1007/S40846-023-00801-3>.
- [168] Pruthviraja D, Nagaraju SC, Mudiligrayappa N, Raisinghani MS, Khan SB, Alkhaldi NA, et al. Detection of alzheimer's disease based on cloud-based deep learning paradigm. *Diagnostics* 2023;13(16):2687. <https://doi.org/10.3390/DIAGNOSTICS13162687>.
- [169] Archana B, Kalirajan K. Alzheimer's disease classification using convolutional neural networks. In: International conference on innovative data communication technologies and application, ICIDCA 2023 - proceedings. Institute of Electrical and Electronics Engineers Inc.; 2023. p. 1044–8. <https://doi.org/10.1109/ICIDCA56705.2023.10100046>.
- [170] Srividhya L, Sowmya V, Vinayakumar R, Gopalakrishnan EA, Soman KP. Deep learning-based approach for multi-stage diagnosis of Alzheimer's disease. *Multimed Tool Appl* 2024;83:16799–822. <https://doi.org/10.1007/S11042-023-16026-0>.
- [171] Singh R, Sharma N, Gupta R. Detection of alzheimer's risk level using inception V3 transfer learning model. In: 2nd IEEE international conference on distributed computing and electrical circuits and electronics, ICDCECE 2023. Institute of Electrical and Electronics Engineers Inc.; 2023. p. 1–6. <https://doi.org/10.1109/ICDCECE57866.2023.10151235>.
- [172] Hridhee RA, Bhowmik B, Hossain QD. Alzheimer's disease classification from 2D MRI brain scans using convolutional neural networks. In: 3rd international conference on electrical, computer and communication engineering, ECCE 2023. Institute of Electrical and Electronics Engineers Inc.; 2023. p. 1–6. <https://doi.org/10.1109/ECCE57851.2023.10101539>.
- [173] Mishra SD, Dutta M. Image-Fused: improved Alzheimer's disease diagnosis using efficient deep neural network. *Soft Comput* 2023;1–12. <https://doi.org/10.1007/S00500-023-08108-W>.
- [174] Holzinger A, Langs G, Denk H, Zatlouk K, Müller H. Causability and explainability of artificial intelligence in medicine. *Wiley Interdiscip Rev Data Min Knowl Discov* 2019;9(4):e1312. <https://doi.org/10.1002/WIDM.1312>.
- [175] Elazab N, Soliman H, El-Sappagh S, Riazul Islam SM, Elmoghy M. Objective diagnosis for histopathological images based on machine learning techniques: classical approaches and new trends. *Mathematics* 2020;8(11):1863. <https://doi.org/10.3390/MATH8111863>.
- [176] Arbabiyazd L, Shen K, Wang Z, Hofmann-Apitius M, Ritter P, McIntosh AR, et al. Virtual connectomic datasets in alzheimer's disease and aging using whole-brain network dynamics modelling. *ENeuro* 2021;8(4). <https://doi.org/10.1523/ENEURO.0475-20.2021>.
- [177] Decuyper M, Stockhoff M, Vandenberghe S, al Ying X. An overview of overfitting and its solutions. *J Phys Conf Ser* 2019;1168(2):022022. <https://doi.org/10.1088/1742-6596/1168/2/022022>.
- [178] Srivastava N, Hinton G, Krizhevsky A. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res* 2014;15(56):1929–58. <https://jmlr.org/papers/volume15/srivastava14a/srivastava14a.pdf>.
- [179] Adadi A, Berrada M. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access* 2018;6:52138–60. <https://doi.org/10.1109/ACCESS.2018.2870052>.
- [180] Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, Pedreschi D. A survey of methods for explaining black box models. *ACM Comput Surv* 2018;51(5):1–41. <https://doi.org/10.1145/3236009>.
- [181] Lipton Zachary C. The mythos of model interpretability: in machine learning, the concept of interpretability is both important and slippery. *Queue* 3(16);2018 : 31–57. <https://doi.org/10.1145/3236386.3241340>.
- [182] Chakraborty S, Tomsett R, Raghavendra R, Harborne D, Alzantot M, Cerutti F, et al. Interpretability of deep learning models: a survey of results. In: 2017 IEEE SmartWorld, ubiquitous intelligence & computing, advanced & trusted computed, scalable computing & communications, cloud & big data computing, internet of people and smart city innovation; 2018. p. 1–6. <https://doi.org/10.1109/UIC-ATC.2017.8397411>. SmartWorld/SCALCOM/UIC/ATC/CBDCOM/IOP/SCID.
- [183] Kaur S, Singla J, Nkenyereye L, Jha S, Prashar D, Joshi GP, et al. Medical diagnostic systems using artificial intelligence (AI) algorithms: principles and perspectives. *IEEE Access* 2020;8:228049–69. <https://doi.org/10.1109/ACCESS.2020.3042273>.
- [184] Fellous JM, Sapiro G, Rossi A, Mayberg H, Ferrante M. Explainable artificial intelligence for neuroscience: behavioral neurostimulation. *Front Neurosci* 2019; 13:1346. <https://doi.org/10.3389/FNINS.2019.01346>.
- [185] Tjoa E, Guan C. A survey on explainable artificial intelligence (XAI): toward medical XAI. *IEEE Transact Neural Networks Learn Syst* 2021;32(11):4793–813. <https://doi.org/10.1109/TNNLS.2020.3027314>.
- [186] Razzak MI, Naz S, Zaib A. Deep learning for medical image processing: overview, challenges and the future BT - classification in BioApps: automation of decision making. *Spring* 2018;26:323–50. https://doi.org/10.1007/978-3-319-65981-7_12.
- [187] Srivastava S, Pant M, Agarwal R. Role of AI techniques and deep learning in analyzing the critical health conditions. *International Journal of System Assurance Engineering and Management* 2020;11(2):350–65. <https://doi.org/10.1007/S13198-019-00863-0>.
- [188] Fong JX, Shapiai MI, Tiew YY, Batool U, Fauzi H. Bypassing MRI pre-processing in alzheimer's disease diagnosis using deep learning detection network. In: Proceedings - 2020 16th IEEE international colloquium on signal processing and its applications, CSPA 2020. Institute of Electrical and Electronics Engineers Inc.; 2020. p. 219–24. <https://doi.org/10.1109/CSPA48992.2020.9068680>.
- [189] Hossain M, Islam SMR, Ali F, Kwak KS, Hasan R. An Internet of Things-based health prescription assistant and its security system design. *Future Generat Comput Syst* 2018;82:422–39. <https://doi.org/10.1016/J.FUTURE.2017.11.020>.