

Review

Transfer Learning for Alzheimer's Disease through Neuroimaging Biomarkers: A Systematic Review

Deevyankar Agarwal ^{1,*}, Gonçalo Marques ^{1,2}, Isabel de la Torre-Díez ¹, Manuel A. Franco Martín ³, Begoña García Zapirain ⁴ and Francisco Martín Rodríguez ⁵

¹ Department of Signal Theory and Communications and Telematics Engineering, University of Valladolid, Paseo de Belén 15, 47011 Valladolid, Spain; gonalosantosmarques@gmail.com (G.M.); isator@tel.uva.es (I.d.l.T.-D.)

² Polytechnic of Coimbra, ESTGOH, Rua General Santos Costa, 3400-124 Oliveira do Hospital, Portugal

³ Psychiatric Department, University Rio Hortega Hospital–Valladolid, 47011 Valladolid, Spain; mfrancom@saludcastillayleon.es

⁴ eVIDA Laboratory, University of Deusto, Avenida de las Universidades 24, 48007 Bilbao, Spain; mbgarciazapi@deusto.es

⁵ Advanced Clinical Simulation Center, School of Medicine, University of Valladolid, 47011 Valladolid, Spain; fmartin@saludcastillayleon.es

* Correspondence: deevyankar@gmail.com

Abstract: Alzheimer's disease (AD) is a remarkable challenge for healthcare in the 21st century. Since 2017, deep learning models with transfer learning approaches have been gaining recognition in AD detection, and progression prediction by using neuroimaging biomarkers. This paper presents a systematic review of the current state of early AD detection by using deep learning models with transfer learning and neuroimaging biomarkers. Five databases were used and the results before screening report 215 studies published between 2010 and 2020. After screening, 13 studies met the inclusion criteria. We noted that the maximum accuracy achieved to date for AD classification is 98.20% by using the combination of 3D convolutional networks and local transfer learning, and that for the prognostic prediction of AD is 87.78% by using pre-trained 3D convolutional network-based architectures. The results show that transfer learning helps researchers in developing a more accurate system for the early diagnosis of AD. However, there is a need to consider some points in future research, such as improving the accuracy of the prognostic prediction of AD, exploring additional biomarkers such as tau-PET and amyloid-PET to understand highly discriminative feature representation to separate similar brain patterns, managing the size of the datasets due to the limited availability.

Keywords: Alzheimer's disease; neuroimaging biomarkers; magnetic resonance imaging; positron emission tomography; transfer learning



Citation: Agarwal, D.; Marques, G.; de la Torre-Díez, I.; Franco Martín, M.A.; García Zapirain, B.; Martín Rodríguez, F. Transfer Learning for Alzheimer's Disease through Neuroimaging Biomarkers: A Systematic Review. *Sensors* **2021**, *21*, 7259. <https://doi.org/10.3390/s21217259>

Academic Editor: Stefanos Kollias

Received: 4 October 2021

Accepted: 27 October 2021

Published: 31 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The commonest form of dementia is Alzheimer's disease (AD), with no authenticated disease-modifying treatment [1]. AD progresses gradually for several years before clinical symptoms [1]. It is estimated that approximately 5.8 million individuals in the USA were living with AD in 2020. This number is anticipated to be 14 million by 2050 [2]. It is critical to identify individuals at an initial stage in AD, even before meeting the clinical criteria. Mild cognitive impairment (MCI) characterizes an effort to distinguish patients at an early clinical phase, and has been considered as a target for clinical trials [3]. MCI is a stage between healthy and AD. A subject with MCI will have changes in their cognitive ability, but still will be able to perform their daily activities. Nearly 20% of individuals aged 65 or above have MCI. In total, 35% of them develop AD in 3 to 5 years [2]. MCI subjects will either be converted to AD or will remain stable. A confirmed diagnosis of Alzheimer's disease (AD) can only be identified through autopsy [4].

Clinical experts are using magnetic resonance imaging (MRI) and positron emission tomography (PET) neuroimaging, neuronal injury, amyloid and tau biomarkers, cerebrospinal fluid, and degeneration [5–7] for AD diagnosis. Mini-Mental State Examination [8], and Clinical Dementia Rating [9] are the two recommended tests for AD diagnosis. The estimated cost of managing AD in the USA, comprising the communal, medical, and income loss to the patient family members in 2020, was approximately USD 305 billion. This cost is expected to rise as high as USD 1.1 trillion by 2050 [2]. The early detection of AD is critical before clinical manifestation for timely treatment. Therefore, a multi-class decision system to identify AD and its different stages from normal controls are required. Categorizing AD from MCI or normal patients is imperative, as AD is detectable without the aid of any experts or skills when it is too late for treatment. The main challenge is the progression prediction of MCI to AD and the distinction of MCI from normal controls in order to give timely treatment to patients. The progress made in neuroimaging techniques allowed physicians the opportunity to attempt to integrate highly dimensional multimodal neuroimaging data by using existing systems. Therefore, researchers recommended the use of computer-aided machine learning (ML) algorithms for AD diagnosis. ML pattern analysis methods that have performed well for AD detection include support vector machines, logistic regression, and support vector machine-recursive feature elimination [10]. Competitions such as Alzheimer’s disease Big Data Challenge [11] and the challenge of predicting MCI from MRI data [12] have revealed usefulness in the AD detection process by providing a common platform for researchers from all over the world. These competitions have produced several ML algorithms that have been implemented and assessed. To use these ML methods for classifications, the architectural design must be predefined. In general, four steps are required, namely the extraction of the features, the selection of the features, the reduction in the dimensions, and the implementation of a classification algorithm. Moreover, this process requires optimization and the involvement of experts in each stage [13]. The reproducibility of these implementations has been an issue [14].

The increasing capacity of GPUs enabled the spread of DL algorithms for image classification tasks. DL is the branch of ML that emulates the working of the human brain for identifying complex patterns. It uses raw neuroimaging data to learn features, hidden representation, and disease-related patterns, and explore associations in different parts of images through impulsive learning, which attracts researchers from the fields of highly dimensional medical image analysis, image segmentation, objects recognition, and disease detection [15]. On the one hand, DL models have been efficaciously applied to neuroimaging images such as MRI, functional MRI, PET, and diffusion-tensor imaging. On the other hand, DL methods, specifically convolutional neural networks (CNNs) have been shown to have better performance in comparison to the existing ML methods [13]. Recent studies suggest that transfer learning (TL) has turned out to be very popular in the domain of DL, as it enables DL training to be effective in the case of insufficient data [16,17]. TL is recommended by comparison with human behavior, since we can use knowledge learned in the previous tasks for solving new complex problems. Classifying AD and normal controls is simple when compared to classifying progressive MCI and stable MCI. Additionally, the amount of neuroimaging data for the classification of progressive MCI vs. stable MCI (pMCI vs. sMCI) is comparatively less as compared to the first task. However, both types of tasks shared common biomarkers [18]. Based on this perception, local TL has been implemented in many studies. The concept of local TL is to use the finalized weights of the AD vs. normal control classifier as the initial weights for the classification of sMCI vs. pMCI. Moreover, some included studies also used pre-trained 3D CNN-based architectures such as VGG16, ResNet for initializing the weights for the classification of AD vs. normal control, and local TL for initializing the weights for the classification of stable MCI vs. progressive MCI, and obtained better results. TL helps to speed up the training procedure and improve the performance of DL architectures [19].

The main objective of this study is to explore the neuroimaging biomarkers and their features, available databases, and pre-processing and input management techniques,

the DL and TL methods that are adept for apprehending the disease-related patterns and perform efficiently even in the case of insufficient neuroimaging data, the accuracy achieved to date, the limitations, and gaps. Consequently, we systematically reviewed the papers of the last 10 years. These papers were reviewed, evaluated, and analyzed based on the DL and TL algorithms and neuroimaging biomarkers. Furthermore, we discussed the outcomes and gaps for DL applications in the AD research domain. This study aims to support future researchers that want to work in this field.

The rest of this article is structured as follows: Section 2 provides the research methodology, including search strategies, research questions, and inclusion and exclusion criteria. Section 3 provides the answers to research questions, along with results and a discussion, followed by the conclusion section.

2. Materials and Methods

A systematic literature review was carried out by the authors, considering studies published over the last ten years. The included studies focus on DL architectures, TL, neuroimaging data to classify AD, and predicting the transformation from MCI to AD. This study was conducted in different steps according to PRISMA methodology [20]. The main objective was to identify the gaps in the state of the art. Firstly, the research questions (RQs) were identified. Secondly, the search strategy was implemented, and the inclusion and exclusion criteria were identified for selecting the relevant articles. Finally, data extraction was carried out to answer the RQs. Furthermore, answers to these RQs were given while highlighting the challenges, limitations, and future scope in the field. These steps were explained in the following sub-sections, respectively.

2.1. Research Questions

To support future research activities, it is mandatory to examine the critical aspects of the current literature. The objective of this is to help develop an improved decision support system for the early diagnosis of AD. The following RQs were identified to represent critical aspects needed.

RQ 1: Which DL models and TL approaches have been used for capturing disease-related patterns?

RQ 2: Which are the neuroimaging biomarkers and the parameters used?

RQ 3: Which are the pre-processing techniques used to handle neuroimaging biomarkers?

RQ 4: What is the current state of accuracy and other performance measures for the diagnostic classification and prognostic prediction of AD?

RQ 5: How many and what structures use the opensource, available datasets?

RQ 6: Which software platforms are needed for pre-processing neuroimaging data and implementing DL algorithms through a transfer learning approach?

RQ 7: What are the techniques that have been used to reduce overfitting?

RQ 8: What are the current research gaps, challenges, and opportunities?

2.2. Search Strategy

Science Direct, IEEE Xplore, Web of Science, ACM Digital Library, and PubMed were chosen to identify contributions in DL applications for the early diagnosis of AD by using TL in neuroimaging data. These databases were selected since they include the most relevant peer-reviewed studies on DL and are the more reputable sources. The search process was carried out by using the following query:

“Deep Learning” AND “Alzheimer” AND “neuroimaging” AND “Transfer Learning”

The first author performed the search by using Sysrev [21]. This is a platform for the collaborative extraction of data from documents available online that provides support by building article filters through ML. In Figure 1, the relevance of important terms identified by Sysrev for our search is given. Maximum relevance is given to early and late MCI, which is fortunately also of utmost concern in this work. A total of 215 articles were acquired, including 161 articles from Science Direct, 15 articles from the Web of Science,

6 articles from PubMed, 25 articles from ACM, and 8 articles from IEEE Xplore. The documents include all types of documents, such as research papers published in peer reviewed journals or conference proceedings, book chapters, survey papers, theses, patents, handbooks, or encyclopedias. The articles were further studied and analyzed according to the inclusion and exclusion criteria. A study was included upon the common agreement of all authors. Authors took extensive care to avoid redundancy during the filtering process. The screening process was supported by Sysrev [22].

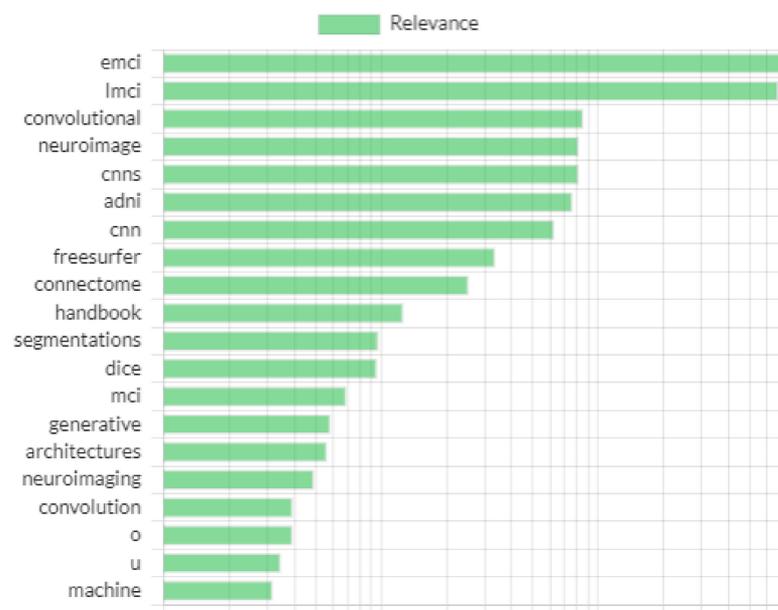


Figure 1. Relevance of important terms identified by Sysrev [22].

2.3. Inclusion and Exclusion Criteria

The inclusion and exclusion criteria are critical to selecting the papers that are significant and aligned with the study objectives. We included articles that implemented DL architecture with TL and neuroimaging biomarkers for the diagnosis and early detection of AD. The issues in the conventional ML, DL and hybrid methods for the early detection of AD by using neuroimaging data, which enforced this inclusion criterion, are given below:

1. The accuracy achieved by ML, DL and hybrid methods is summarized in Table 1. While traditional ML methods such as SVM have shown good results, DL methods such as CNN and sparse autoencoders have surpassed ML methods. Hybrid methods use deep learning for feature extraction and traditional ML methods such as SVM for classification. The combination of SVM for classification and stacked autoencoder for feature selection produced accuracies of up to 83.3% for early detection of AD. The maximum accuracy achieved by DL method for early detection of AD is 82.9. This has been further improved by using TL, as is evident in Section 3.
2. One issue with traditional ML and DL models is that they utilize samples only from a single domain, which reduces their accuracy, when the numbers of samples is lower. TL utilizes samples from the target domain as well as from various supplementary domains.
3. In traditional ML, results are influenced by well-defined features; selecting the optimal features from the highly dimensional neuroimaging data required the involvement of experts and optimization in multiple stages, which may be time consuming. DL identifies important features from neuroimaging data automatically, but the problem in training a deep neural network model from scratch is that it can take too long to converge and needs a dataset of sufficient size. Image classification datasets used for object classification have millions of images, but neuroimaging datasets are limited to hundreds. Specifically, the size of datasets for sMCI and pMCI subjects is very

- limited, and physician's interpreted data can be expensive or inaccessible due to ethical reasons. A limited number of training data can lead to overfitting.
4. The hybrid methods yield good performance with a limited amount of data, but they do not take complete advantage of DL. Here, the idea of TL by using pre-trained CNN networks came into the picture. Pre-trained CNN networks are trained on datasets of millions of images, on multiple servers for weeks, and facilitated by professional experts for public use. These networks can be used with small datasets by performing the fine tuning of the final layers of the CNN. Even TL has been implemented for cross-domain applications in the included studies.

Table 1. Comparison of research for the early detection of AD by using ML, DL and hybrid methods.

Approach	Method	Accuracy of AD vs. NC	Accuracy of sMCI vs. pMCI	Reference
ML Only	SVM (RBF)	81%	70%	[23]
	SVM(LINEAR)	88.9%	70.7%	[24]
	TS-SVM	-	82.5%	[25]
	SVM(LINEAR)	93.1%	75%	[26]
	(GROUP LASSO) SVM	95.1%	65.4%	[27]
	SVM (RBF)	82.5%	69.23%	[28]
DL Only	3D CNN	92.87%	76.21%	[29]
	DBN	90%	78%	[30]
	DNN	84.6%	82.4%	[31]
	3DCNN	91.09%	76.9%	[32]
Hybrid Methods	SAE + SVM	98.8%	83.3%	[33]
	DBM + SVM	95.35%	75.92%	[34]
	RBM+SVM	91.4%	57.4%	[35]

Abbreviations: SVM—State Vector Machine; RBF—Radial Basis Function; TS-SVM—Temporally Structured SVM; DBM—Deep Boltzmann Machine; DNN—Deep Neural Network.

The authors believe that the clinically acceptable system for the early detection of AD will make exhaustive use of the TL concept. It can be seen in the further discussion that the recent publications with good accuracy for the early detection of AD used TL.

Cross validation and at least two prediction matrices were also considered. The clear explanation of data sources and pre-processing methods applied for biomarkers were also considered as one of the inclusion criteria. Articles such as survey papers, thesis, patents, handbooks, encyclopedias were excluded. We did not find any book chapters or conference proceedings that satisfied the inclusion criteria.

2.4. Study Selection

The search strategy and study selection were delineated in detail using the PRISMA flow diagram [20] (Figure 2). After the identification of the articles through the search strategy, 215 articles were obtained. In total, 15 duplicates were removed. The remaining 200 items were screened for a second time. The studies' importance was identified through the title, abstract, and 165 articles were excluded as they met at least one of the exclusion criteria. Several papers were excluded as they classified other diseases, such as autism and breast cancer, and did not focus on Alzheimer's disease. Out of the 35 articles retrieved, in total, 12 articles [36–47] were excluded as they did not use TL approaches or did not perform the prognostic prediction of AD. Five articles [48–52] were excluded as they did not provide at least two prediction matrices or did not perform cross-validation of the data. Two studies [53,54] were excluded as they did not provide the insights for input

management. The other two papers [55,56] were excluded as they did not use neuroimaging data. Finally, the study proposed by [57] was excluded as they tried to distinguish AD from Parkinson's. The remaining 13 papers [58–70] were included in this systematic review process. These studies are relevant to answering the research questions identified at the start of this section. In this research area, the leading journals are *Journal of Neuroscience Methods*, *NeuroImage: Clinical and Behavioural Brain Research*, and *Medical Image Analysis*. According to Google Scholar, references [58,65,66,68] are the most cited publications.

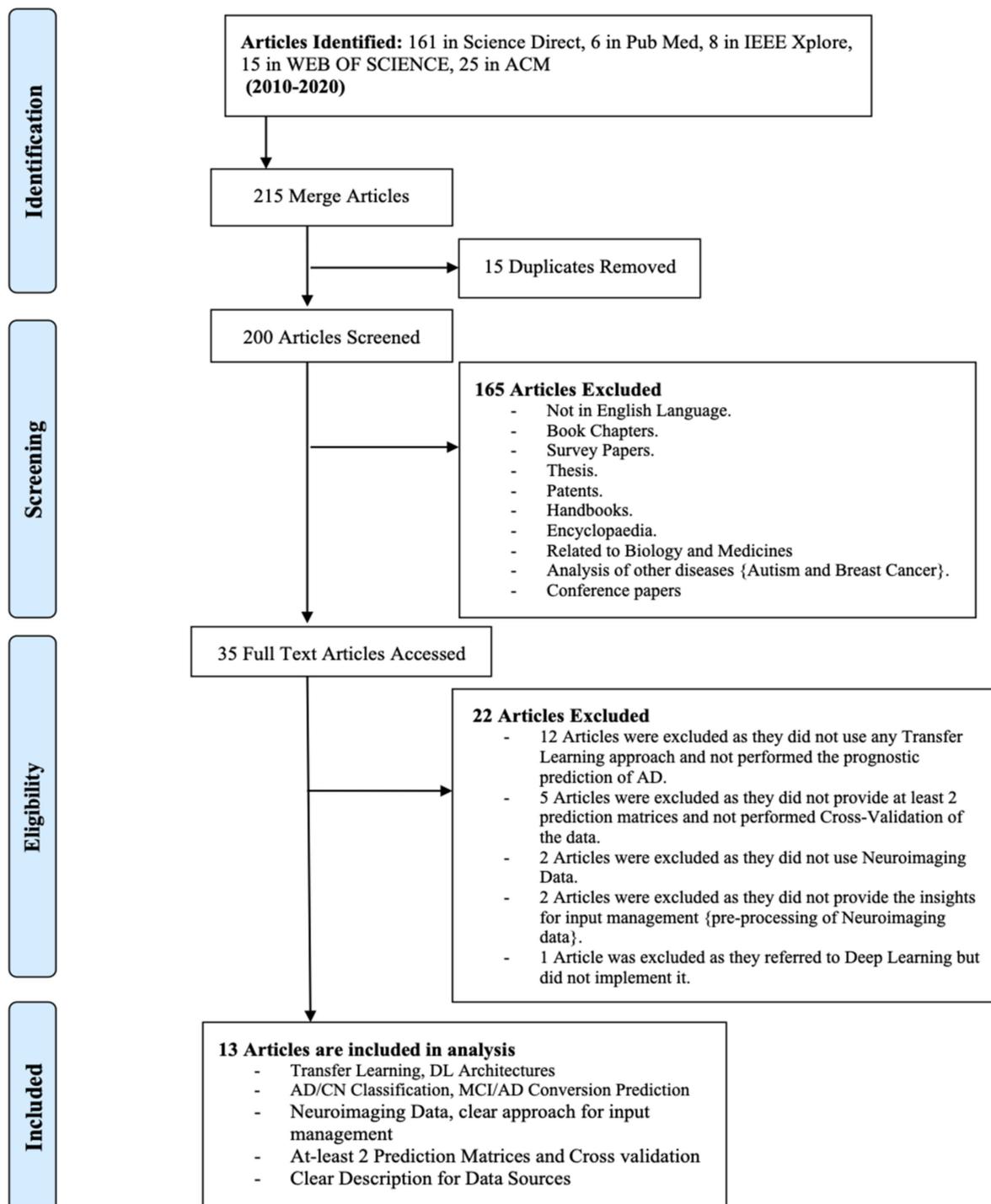


Figure 2. PRISMA diagram of the proposed systematic review.

2.5. Data Extraction Parameters

The data collected were extracted based on the following parameters to conduct the systematic review.

1. Neuroimaging modality.
2. DL algorithm and TL approach.
3. Input management techniques.
4. Pre-processing techniques for neuroimaging biomarkers.
5. Dataset sources and number of neuroimaging biomarkers.
6. Data validation and augmentation techniques.
7. Software used for the pre-processing of neuroimaging biomarkers and implementing algorithms
8. Classifier type.
9. Performance metrics of the proposed system.

2.6. Risk of Bias

The main limitation of conducting a systematic literature review is that it might be biased. To avoid this scenario to the maximum possible extent, the authors designed the search key using the keywords and logic operators. Based on the PRISMA guidelines, the authors tried to follow the best possible criteria and processes for completing the review. Only articles that satisfied all the inclusion criteria were included in the final evaluation. Although early researchers conducted several reviews [71,72] of ML and DL methods for AD classification, they did not consider all these RQs, specifically RQ1 and RQ4. In [71], the authors conducted an in-depth analysis of studies published between 2013 and 2018, and they performed a comparison of the diagnostic classification accuracy of pure DL and a combination of DL and ML methods. They also compared the change in accuracy based on the type of biomarkers. In [72], the authors presented the review of studies published between 2005 and 2019 and conducted an in-depth analysis according to the three categories, namely state vector machine, artificial neural network, and deep learning models. The main differences in the existing studies and this work are that in the previous studies, the authors did not perform the comparative analysis of the change in accuracy according to the DL and TL combinations, and according to the type of biomarker and input management technique. Therefore, this study can help future researchers in identifying the most effective combination of the DL model with TL and in choosing the most suitable preprocessing method. Additionally, in the previous studies, no discussion was evident regarding the emerging biomarkers such as amyloid-PET and tau-PET. In the current study, the authors explored the importance of these biomarkers in future research for the early diagnosis of AD.

In Section 3, answers to all the RQs are discussed in detail. We hope that this work will provide useful information for the scientific community in understanding the research gaps, opportunities, and challenges in this field, which will help researchers in implementing a clinically acceptable solution for the early diagnostic of AD. Several relevant acronyms are provided in the Supplementary Materials (Table S1).

3. Results and Discussion

This systematic review includes 13 studies from four different databases. Among these, nine studies were included from Science Direct, two from IEEE Xplore, one from PubMed, and one from Web of Science. Based on the keywords, we found acceptable studies to have been published only from 2017 to 2020—one study in 2017, five studies in 2018, three studies in 2019, and four studies in 2020. The details regarding the included work, year, journal name, database, country, and citations are provided in the Supplementary Materials (Table S2). The review process, which focused on extracting information from the included studies, is presented in the Supplementary Materials. Table S3 presents the technical insights and different datasets of neuroimaging biomarkers of the included

studies. Table S4 provides a category-wise analysis of different pre-processing techniques for neuroimaging biomarkers with the required software.

3.1. Answer to RQ1

The first research question concerned the types of DL architectures and TL methods used to capture the disease-related patterns from neuroimaging modalities to classify AD and predict the conversion of MCI to AD. DL architectures can be divided into two categories, either supervised learning or unsupervised learning. The unsupervised methods can be divided into convolutional autoencoders (CAEs), stacked autoencoders (SAEs), and inception convolutional autoencoder (ICAEs). The supervised methods can be divided into DenseNets, deep neural networks (DNNs), recurrent neural networks (RNNs), Google Net, CaffeNet, AlexNet, and 2D/3D convolutional neural networks. In Figure 3, the dissemination of different DL architectures and their combinations used in the analyzed studies is given.

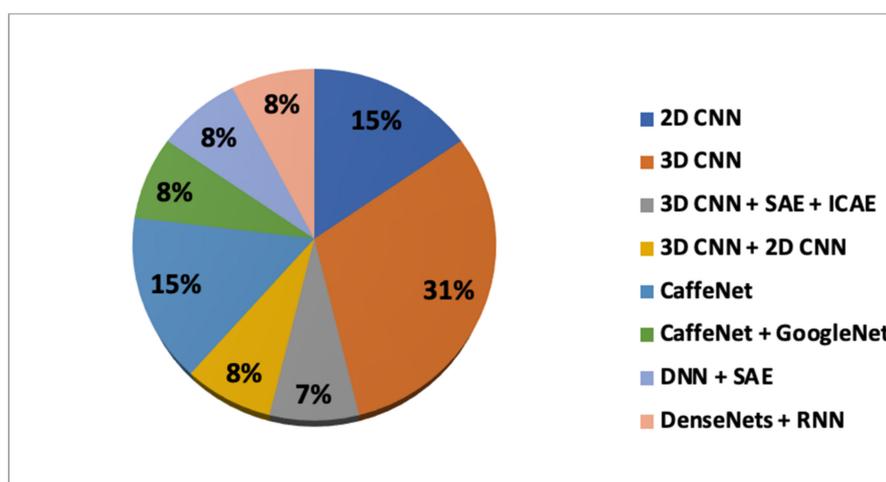


Figure 3. Dissemination of DL architectures in the included studies.

3.1.1. Unsupervised Deep Learning

Unsupervised learning is an indispensable component when confronted with data shortages and high dimensionality. It is used to attain a task-specific depiction from neuroimaging data and for transfer learning, particularly beneficial when one has restricted labeled data but a bigger set of unlabeled data. Autoencoders (AEs) are the most common approach to be used in such cases. In general, numerous AEs are piled on top of each other, and the result is called a stacked AE. Researchers have used different variations of AE, such as CAE, ICAE, and SAE, for feature extraction and dimension reduction, and transfer learning in an unsupervised way. KanghanOh et al. [60] proposed in their work CAE- and ICAE-centered unsupervised learning to extract sparse representations for AD and NC. Initially, MRI scans of patients with AD and NCs were pre-trained based on CAE-based unsupervised learning, followed by supervised fine-tuning to build the classifier to differentiate AD vs. NC. After that, they transfer learned knowledge for classifying the pMCI vs. sMCI.

In [60], the authors applied an inception module to the convolutional autoencoder (ICAЕ) to discover highly advanced representations and lessen the dependency on the fully connected layer. The outcomes of the CAE and ICAE-based models exhibited improved performances, as compared to the baseline models [73–75]. Donghuan Lu et al. [65] used SAE for the pre-training of each DNN of the proposed multiscale DNN. They used greedy layer-wise training [76], training a single hidden layer at a time. This was followed by supervised fine-tuning that was steered to build the classifier to differentiate pMCI vs. sMCI, which showed better performance than the published literature.

3.1.2. Supervised Deep Learning

Researchers have used the combination of unsupervised learning for the future extraction and initialization of weight and biases of the target networks, with supervised fine-tuning as explained below.

Deep Neural Networks: DNNs are the most extensively used in diverse research areas to learn formerly unknown abstract patterns. DNN provide some of the top performances amongst the machine learning methods recommended for recognition tasks. Deep learning networks have also been applied to distinguish the AD-related progression patterns. Liu et al. [77] trained a stacked autoencoder to study the concealed representation trailed by a softmax output layer for classification. Donghuan Lu et al. [65] suggested a novel framework comprising of multiscale DNN to study the patterns of metabolism alterations due to AD pathology as discriminative from the patterns of metabolism in normal controls. For a single DNN, they implemented training in two steps—firstly, unsupervised pre-training by using SAE, and secondly, supervised fine-tuning by keeping the three encoder layers and adding a softmax output layer for each DNN. For improving the accuracy of the progression prediction of AD, the authors used the weights of the DNN, used to classify AD vs. NC.

Convolutional Neural Networks: CNNs are stimulated by the visual cortex of the brain. CNNs are the most effective model for image analysis. They are used to better exploit spatial information by taking 2D/3D images as the input and extracting features by stacking convolutional layers. The main advantage of a CNN is to amalgamate feature extraction and classification. In comparison to DNNs, the number of parameters is drastically reduced because of shared weights and pooling layers in CNNs. The necessity for a large dataset can be contemplated as a flaw of these models. The CNN's architecture received increased attention after a deep CNN achieved excellent results in the ILSVRC competition. Although many studies preferred to construct their own CNN structures, it is common in the literature to use well-established pre-trained CNN structures such as ResNet, deep-ResNet [78], CaffeNet [79], AlexNet [80], DenseNet [81], VGG16 [82], GoogleNet [83], and inception-v4 [84] for transfer learning. CNNs are designed to identify patterns in 2D images. Several researchers have used 2D CNNs for 3D neuroimaging data. Farheen Ramzan et al. [59] use ResNet-18 for transfer learning followed by a 2D CNN for the classification of AD vs. NC classes. They executed transfer learning in two approaches. Firstly, TL is used by substituting the last dense layer of the original network with the new dense layer to counterpart the number of classes. Secondly, all the layers, excluding the last layer (classifier) of the network, are used for feature extraction, and the weights of the last layer are reformed to the new task. Moreover, fine-tuning, in which more than one layer of the network is re-trained from the samples of the new task. Lan Lin et al. [64] extracted intermediate representations of structural MRI (sMRI) by means of a pre-trained 2D CNN model (CaffeNet). Moreover, they combined principal component analysis (PCA) and sequential feature selection (SFS). Finally, they adopted a support vector machine (SVM) for the progression prediction of MCI to AD. As neuroimaging data are 3D, and there is a spatial relationship amongst the images, 3D CNNs were used in the most recent research. For AD detection, CNNs essentially take the whole image or region of interest as the input. This may necessitate training an enormous number of parameters on a small dataset, which may lead to overfitting. To avoid overfitting, researchers used augmentation and cross-validation techniques [58–61,63,69]. Silvia Basaia et al. [59] used a 3D CNN without any previous feature engineering and irrespective of the erraticism of imaging protocols and scanners. To improve the performance of the prognostic progression of AD, they transferred the weights of the CNN used to classify AD vs. NC to the other CNNs to classify sMCI vs. pMCI. Their network architecture contains 12 recurrent blocks of convolutional layers (2 blocks with 50 kernels of size $5 \times 5 \times 5$ with alternating strides 1 and 2 and 10 blocks with 100 to 1600 kernels of size $3 \times 3 \times 3$ with alternating strides 1 and 2), a rectified linear unit (activation layer), a fully connected layer and one output (logistic regression) layer. Anees Abrol et al. [62] proposed an adapted form of ResNet for

studying neuroimaging data to predict MCI to AD progression. The proposed method starts by training the deep models using MCI individuals only, followed by a TL version that is further trained on AD and controls. This framework has 3D convolutional units, 3D batch-normalization units, and non-linear activation units, and a fully connected (FC) layer featuring 512 output nodes. Marcia Hon et al. [68] used two different pre-trained 3D CNN-based architectures VGG16 and inception V4. Only the last fully connected layer is re-trained with the training data. For the transfer learning models, pre-trained network weights from ImageNet were obtained. FeiGao et al. [70] proposed an age-adjusted 3D CNN-based pre-training model. The proposed method aims to extract and transfer features. Moreover, the authors used a pre-trained network for feature extraction and age prediction and a fine-tuned network to transfer both features and knowledge from age prediction for MCI converter prediction. The pre-trained network takes 3D MRI images from NC subjects as inputs and predicts age and extracts related features. In the fine-tuning model, the age-related information from the pre-training is transferred. Together with the features from the pre-training model, the risk of the MCI subject converting to AD is predicted. Junhao Wen et al. [61] implemented four different CNN architectures—3D subject-level CNN, 3D ROI-based CNN, 3D patch-level CNN, and 2D slice-level CNN. Moreover, two different approaches were used for transfer learning: (i) AE pre-training for 3D CNNs, and (ii) ResNet pre-trained on ImageNet for 2D CNNs. Three approaches (3D subject-level, 3D ROI-based, 3D patch-level) provided approximately the same level of performance. On the other hand, the 2D-slice approach was less efficient.

Zhiguang Yang et al. [67] used the AlexNet pre-trained CNN architecture of the Caffe framework [85]. Caffe was implemented in C++ language and had Python and Matlab interfaces, and it could speedily realize the CNN on GPU and expediently convert between GPU and CPU. The three basic results of the Caffe framework are blobs, layers, and nets. The network structure of AlexNet included eight layers of neural networks, five convolution layers, and three fully connected layers. Congling Wu et al. [63] explored and evaluated two CNN architectures, GoogleNet and CaffeNet, in multiple classifications and estimations of conversion risk through TL from pre-trained ImageNet (via fine-tuning). CaffeNet is a revised version of the classic CNN architecture of AlexNet [79]. GoogleNet model is the current state-of-the-art CNN framework proposed for the ILSVRC challenge. The architecture of GoogleNet [83] is considerably composite and deep, involving a 22-layer architecture. CaffeNet attained promising accuracies in the NC, sMCI, and pMCI classifications in comparison to GoogleNet.

Recurrent Neural Networks: RNNs [86] can efficiently model the temporal behavior for a time sequence, with the output being reliant on the preceding computations. Fan Li et al. [50] proposed a hybrid method including convolutional and recurrent neural networks for hippocampus analysis. They proposed 3D DenseNets based on the decomposed patches of the 3D image of the hippocampus to assess the intensity and shaping of the features and applied the bidirectional GRU (BGRU), which consists of a forward GRU and a backward GRU, to capture correlation features between the left and right hippocampus from two directions. The BGRUs are cascaded on the DenseNet and the inputs of each BGRU are the features of the left and right hippocampus learned by 3D DenseNets from both the interior and exterior hippocampus.

3.1.3. Transfer Learning Methods

Training a DL model from the initial stage was performed in some of the studies. However, the training process consumes considerable time, and the size of neuroimaging datasets is restricted, which may lead to overfitting. As we have seen in the previous discussion of DL architectures, TL is less time-consuming and shows better performance in comparison to training from scratch. Recent studies [58,62–64,67–69] used initialization or feature extraction by CNN-based TL architectures, such as ResNet, deep ResNet, CaffeNet, AlexNet, DenseNets, VGG16, and inception with or without fine-tuning, followed by an extended network architecture. On the other hand, other approaches [59,65,66,70] used

local TL. The weights of the CNN used to classify AD vs. NC were transferred to the other CNNs and used as initial weights to predict the progression of AD. Furthermore, Refs [60,61] used unsupervised DL architectures such as SAE and ICAE for the pre-training of 3D CNNs. Figure 4 shows the distribution of each TL approach and their combinations used in the analyzed studies.

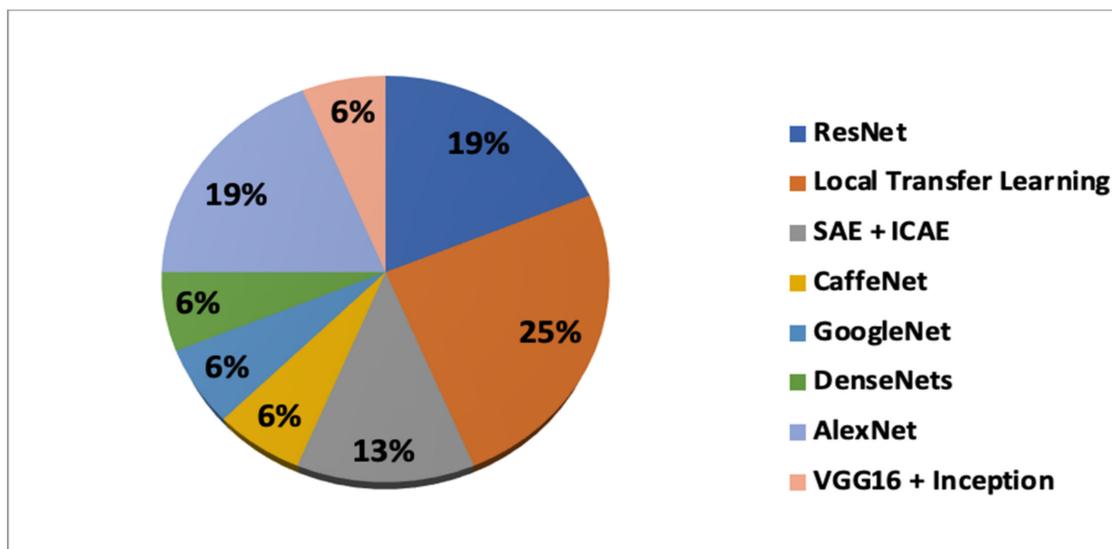


Figure 4. Distribution of TL methods in the included studies.

3.1.4. Comparative Analysis of Supervised/Unsupervised DL Networks with TL

According to our review, the unsupervised DL architectures SAE, CAE, and ICAE are best for extracting features and for classification. They can characterize highly non-linear and complex patterns and can be used for finding initialization parameters for CNNs. AEs have limitations, as they learn to capture as much information as possible and do not only capture relevant information.

Supervised methods were mostly used in the analyzed studies, as they enabled feature selection and categorization being combined into a single algorithm. TL methods have been used in the most suitable initialization and most informative feature extraction methods, followed by extended supervised DL architectures. 3D CNNs can obtain 3D neuroimaging information and has exhibited better performance in local feature extraction when used with ResNet for TL. However, the intricacy of training is a limitation, which can be solved by using ROI or patch-based pre-processing methods for neuroimaging modalities. On the contrary, 2D CNN is easier to train, but is not competent enough in encoding the spatial information of 3D images across the third dimension. Therefore, several studies use all three views of neuroimaging images or use RNNs to capture 3D information. RNNs show reliable performance for sequential 2D image slices. However, they have problems associated with the training process owing to the fading gradients. DNNs show a decent performance for vector-based problems, as they detect complex non-linear relationships, but have a sluggish training process, generalization issues, and are not ideal for images. The performance of the different DL architectures for the included studies concerning the accuracy, sensitivity, and specificity, and technical insights related to DL architectures and TL methods are reported in the Table S3. The accuracy achieved to date by using different combinations of DL and TL approaches is discussed in Section 3.4.

3.2. Answer to RQ2

The second research question concerned the neuroimaging biomarkers and parameters used for the classification and prediction of the progression of AD. For AD detection, non-invasive biomarkers, such as MRI and PET, have been used commonly. During the

feature selection process, the most commonly chosen features from these biomarkers are the mean sub cortical volumes, gray matter densities, cortical thickness, brain glucose metabolism, and cerebral amyloid- β accumulation in regions of interest. The neuroimaging data acquired from these technologies have been used for providing a computer-aided system to help physicians to improve healthcare systems. MRI is well-known and uses biomarkers and has demonstrated excellent performance in the implementations of AD vs. NC classifications. MRI provides the ability to study pathological brain alterations accompanying AD. It uses a magnetic field and radiofrequency pulses to generate a 3D representation of soft tissues, bones. Different types of MRI images have been used in the included studies. The sMRI was used in [62–64,68]. sMRI facilitates the tracing of brain atrophy and aids in finding out the possible cause of AD; it has additional advantages related to its high spatial resolution and its availability. Functional MRI (fMRI) also reflects the changes associated with blood flow. Resting-state fMRI (rs-fMRI) captures the fluctuations in blood oxygenation levels of subjects in the rest state. As a result, the brain regions affected by neuron-degeneration demonstrate diverse patterns of blood oxygenation levels. The fMRI has been used in one included study [58]. Several of the authors have used the T1-weighted sequence (T1w) of sMRI scans. T1w sequences are part of the MRI protocols and are thought of as the most anatomical of images. These sequences approximate the closest appearance of tissues such as white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) [87]. These measurements are one of the most significant biomarkers to detect AD progression [68]. T1w images were used in [59–61,69,70]. The T1w sequences used are of different types, for example, 3T or 1.5 T baseline. Here, 3T/1.5T implies a 3Tesla/1.5 Tesla MRI, which is produced by a magnetic field of 3 Tesla or 1.5 Tesla and results in a cleaner and more complete image [88]. The authors of [89] used T1w MP-RAGE sequences of MRI and obtained excellent results. Magnetization-prepared rapid gradient echo (MP-RAGE) is a sequence of sMRI which captures high tissue contrast and offers high spatial resolution with whole-brain coverage in a short time period. PET is nuclear medicine functional imaging technology used to observe the metabolic process for AD detection. It has been used for many years for research and clinical purposes. Meanwhile, 18F-FDG-PET is the most widely used type of PET image. It provides a measure of cellular glucose metabolism. It can be used to assist the neurocognitive disorder due to AD [90]. It is mainly beneficial for the early detection of AD, as it can demonstrate the distinctive patterns of AD earlier than MRI for MCI subjects [91]. Another type of PET is amyloid-PET, which is used to assess brain amyloid deposition, one of the main neuropathological milestones of AD, with an elevated sensitivity and specificity in patients with established Alzheimer's disease who underwent an autopsy within one year of PET imaging [91]. Several studies used different combinations of biomarkers to improve the accuracy of AD detection [92,93]. PET images were used in [65–67]. In [65], the authors used 18F-FDG-PET with sMRI, and in [66], the combination of 18F-FDG-PET with AV-45 PET obtained comparatively good accuracy for detecting the progression of AD. Another type of future PET is tau-PET imaging [94]. Abnormal deposition of tau is a key contributing factor to neurodegenerative diseases such as AD. The advances in neuroimaging technologies have steered the advancement of tau-specific tracers for PET such as THK5351 and THK5357 [95]. These images are specifically helpful in monitoring the progression of the disease. We did not find any article where the authors used tau-PET for predicting the AD progression by using DL and TL methods. Additionally, a few additional factors that are pertinent to AD detection are gender, age, the pattern of speech, EEG, tau protein, A β protein, retinal abnormalities, postural kinematics analysis, MMSE and CDR score, logical memory test, and some genes believed to be responsible for AD [23]. A detailed explanation of the discussed neuroimaging modalities for AD detection was given in [91]. Figure 5 presents healthy and AD images. Their clinical acceptability decreases from left to right. In the sMRI, grey matter volume is displayed in blue, in the 18F-FDG PET images, reduced metabolism is displayed in green, in the amyloid-PET images, the small quantity of deposited amyloid is displayed in green or blue, and in the

tau-PET images of AD patients, low tau-tracer retention is displayed in green or blue. Two other types of future PET images are TSPO-PET [95] and SV2A-PET [96], which can be integrated into future research with DL techniques for the detection of AD progression.

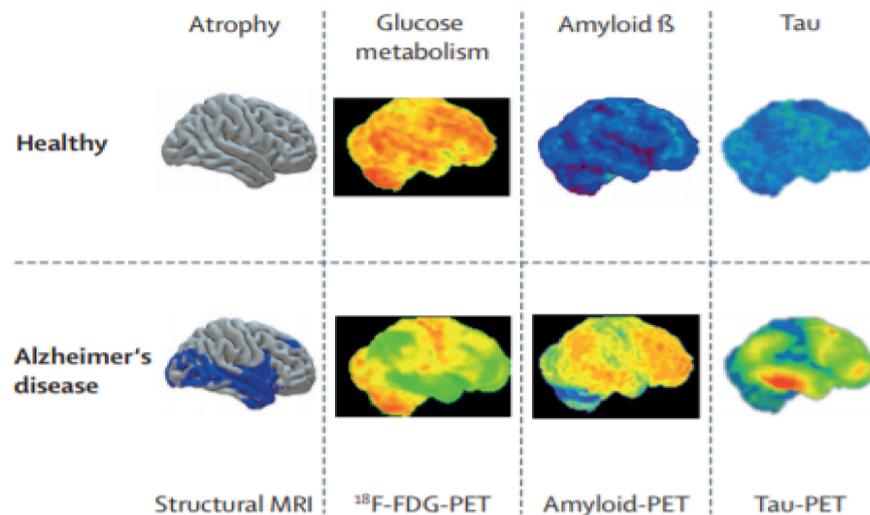


Figure 5. MRI and PET using different radiotracers. Reprinted with permission from ref [91] Copyright 2021 Elsevier.

The distribution of neuroimaging modalities used in the included studies is given in Table 2.

Table 2. Distribution of neuroimaging modalities.

Modalities	Percentage	References	N
rs-fMRI	7%	[97–99]	1
sMRI	31%	[62–64,68]	4
3T T1wMRI	23%	[59,69,70]	3
1.5T T1wMRI	15%	[60,61]	2
18-F FDG PET	8%	[67]	1
18-F FDG PET + sMRI	8%	[65]	1
18-F FDG PET + AV-45 PET	8%	[66]	1

In Table 3, the accuracies achieved by the type of neuroimaging biomarkers in the included studies are given.

Table 3. Accuracy achieved by the type of neuroimaging biomarkers.

Type of Biomarker	AD vs. NC	pMCI vs. sMCI	References
rs-fMRI	97.92%	-	[58]
sMRI	92.78%	77.85%	[62–64,68]
3T T1wMRI	94.70%	75.16%	[59,69,70]
1.5T T1wMRI	87.30%	75.97%	[60,61]
18-F FDG PET	-	72.19%	[67]
18-F FDG PET + sMRI	93.58%	82.51%	[65]
18-F FDG PET + AV-45 PET	96.10%	84.20%	[66]

3.3. Answer to RQ3

The third research question focuses on the used pre-processing techniques and the input management of neuroimaging modalities for the classification and prediction of the progression of AD. Pre-processing is one of the main contributing factors to determine the success of any AI-based system. It removes artifacts and noise from the data to enhance the quality of the image and improves feature extraction. Pre-processing steps are not needed in some DL architectures [98,99]. Nevertheless, most of the included studies used some of the following pre-processing methods on the raw neuroimaging data.

1. **Intensity Normalization:** The use of different scanners and parameters for examining subjects at different times may cause large variations in the intensities. It can cause performance degradation in subsequent processing, such as registration, segmentation, and tissue volume measurement [99]. Normalization relates to plotting the intensities of all the voxels against a reference scale so that alike structures have identical intensities. The most common method is to use the non-parametric, non-uniform intensity, normalization (N3) algorithm. This is used for correcting the intensity's non-uniformity, by sharpening histogram peaks [100]. It has been applied in the included articles [58,63,69]. In [60], the authors normalized the intensities of the MRI scan to [0, 1]. In [66], the brainstem region was utilized for internal normalization, as it is expected not to be affected by AD, hence the mean intensity in the brainstem region is calculated and used to distribute the average intensities respectively for each subject.

2. **Spatial Smoothing:** An FWHM Gaussian filter between 5 and 8 mm reduces noise level and preserves the signals present in the image [101]. It was applied in three of the included studies [58,62,64].

3. **Registration (Spatial Normalization):** Registration is the process of mapping a subject's neuroimaging image to a reference brain space. It allows comparison among subjects of varying anatomy [102]. It standardizes neuroimaging images regarding a common template such as MNI [103]. The authors registered the neuroimaging images into the MNI template by using different methods. In [58], linear transformation with 12 DOF (such as translation, scaling, shear, and rotation) was used. The Diffeomorphic Anatomical Registration Exponentiated Lie Algebra (DARTEL) registration [104] was used in several studies [59,60,64]. In [62], linear (affine) registration was performed using the SyN algorithm [105]. In [61], the authors used non-linear registration by using the unified segmentation approach [106]. In [62], the images were spatially normalized to the 152 average T1 MNI template. In [65], the authors performed image registration via a highly dimensional, highly accurate non-rigid registration method (LDDMM) [107], widely thought of as being the prominent registration tool. Another registration process found in the literature is co-registering multiple modalities. Anterior commissure (AC) and posterior commissure (PC) are two main landmarks in the brain. AC-PC line [108] has been embraced as a standard by the neuroimaging community, and in many examples, the reference plane for axial imaging to achieve the comparison between different subjects. In [66], the authors registered PET images such that the forward-facing and the rearmost axis of the subjects were parallel to the AC-PC line. In [69], the authors implemented the affine registration method [109] to linearly align the MR images to a template. In [70], the authors conducted rigid registration to the MNI152 space to ensure consistency of orientation and position. The authors of [64,67,70] applied interpolation methods [110] to transfer data into the same voxel sizes and dimensions.

4. **Gradwrap:** Correct geometry distortion due to gradient non-linearity. It was used in one included study [63].

5. **Tissue segmentation:** Tissue segmentation aims at partitioning a neuroimaging modality into segments corresponding to different tissues (grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF)), probability maps and metabolic intensities of the brain region in the case of PET images. Probability maps are used as the input of the classification model. Probability maps contribute the quantitative measurement of the spatial distribution of these tissues in the brain. Neurodegeneration causes changes in the

volume of GM at its initial stages, specifically in the temporal lobe area [111]. This process was conducted in [59,62,64,65,67] using different software tools.

6. Bias-field correction: This is a low-frequency, smooth signal that degrades neuroimaging images, particularly those formed by old MRI machines [112]. A pre-processing step is needed for correcting bias-field inhomogeneities. This process was applied in three of the included studies [60,61,63] by using the N41TK [113] algorithm.

7. BrainExtraction (Skull Stripping): For eliminating non-brain tissues, such as neck and skull tissues. It was used in three of the included studies [58,61,69]. In [61], the authors applied a mask-based approach for skull-stripping.

8. Motion Correction: Used to eliminate and make the impact of head motions more precise during the data acquisition process. It was used in one included study [58].

9. High-pass filtering: Used to eliminate the low-frequency noise signal, which is generated because of psychological artifacts. It was used in one of the included studies [58].

In [68], the authors used the entropy-based sorting mechanism to choose the most informative images from the axial plane of each 3D MRI scan, and they did not use any other pre-processing methods.

In Table 4, the dissemination of each pre-processing method from the included studies is given; as we can see, almost 50% of studies performed intensity normalization and registration.

Table 4. Distribution of pre-processing methods.

Pre-Processing Method	Used %	References	N
Intensity Normalization	54%	[43,45,46,48,50,54,55]	7
Spatial Smoothing	23%	[58,62,64]	3
Registration	46%	[58–61,65,67]	6
Gradwrap	8%	[63]	1
Brain Extraction	23%	[58,61,69]	3
Tissue Segmentation	38%	[59,62,64,65,67]	5
High Pass Filtering	8%	[58]	1
Motion Correction	8%	[58]	1
Bias-Field Correction	23%	[60,61,63]	3
Entropy-Based Sorting	8%	[66]	1
No Pre-processing	8%	[68]	1

Based on the extracted features from the neuroimaging modalities, the management of input data can be allocated to the following four categories.

1. Voxel-based input management: In this approach, voxel intensity values of the complete neuroimaging modalities or tissue components are used. The registration method is necessary for this approach to map all the images into a standard 3D space. In [59], the authors performed tissue segmentation and registration on T1-weighted MRI to generate the GM, WM, and CSF tissue probability maps in MNI space. Several research studies [60,61,70] performed a full brain image analysis of T1W sMRI. In [66], the authors performed the full brain analysis in multimodality mode (18 FDG-PET and AV-45 PET) and co-registration by using the AC–PC line. The advantage of full brain analysis is that the spatial information is fully integrated, and we can obtain 3D information from neuroimaging scans. The drawback of this approach is the high dimensionality and computational load. In [64], the authors segmented the sMRI images into GM, WM, and CSF by using the DARTEL method and used PCA and SFS for the precise selection of the best features.

2. Slice-based input management: In this approach, 2D slices are extracted from 3D brain scans either by using the researcher’s logic or standard projections such as the horizontal plane, frontal plane, and median plane. Full brain analysis is not possible,

as all the information of 3D brain scans cannot be converted into a 2D slice. For tissue segmentation, generally, slice-based methods take tissues from the central part of the brain. In [58], the authors performed many pre-processing steps, as given in Table S4, to convert 3D rs-fMRI scans to 2D images, along with the image height and time axis, and they obtained 6160 2D images for each fMRI scan and saved them in PNG format. In [63], the authors used Matlab and NifTi to convert 3DMRI scans of each subject into 160 2D image slices. In [67], the authors converted the NifTi images (created from the gray matter density maps of each subject) averaged from the Z-axis into 65 PNG images. In [68], the authors performed the entropy-based sorting mechanism to select the 32 most informative images from the axial plane view of each 3D scan. The advantage of this approach is to decrease the number of parameters to thousands from millions during the training and testing phase of the model. Because of using 2D representations of 3D scans, the adjacent 2D images lose spatial dependency.

3. ROI-based input management: Studies based on the region of interest focused on the areas that are identified to be informative. In this context, the informative parts that are affected in the early stage of AD have to be recognized. The hippocampus, amygdala, and entorhinal cortex, grey matter in the entire brain, and the temporal and parietal lobe must be treated with greater precedence for AD or MCI classification, whereas the amygdala and hippocampus can be used for predicting the conversion of MCI to AD [114]. Meanwhile, the occipital lobe, thalamus, globus pallidus, and putamen ought to be low priority choices for the early diagnosis of AD. This approach also requires the previous knowledge of the brain atlas, such as automated anatomical labeling (AAL) [115] or Kabbani reference work [116]. In [62], the authors identified the utmost discriminative brain regions for classifying the pMCI vs. sMCI by approximating occlusion sensitivity utilizing the network occlusion approach [117]. They occluded brain networks in reference to the AAL brain atlas one at a time, and the significance of each brain region was assessed. The most relevant weights were perceived in the “hippocampus, temporal superior, par hippocampal gyrus, middle and inferior gyrus, occipital superior, fusiform gyrus, middle and inferior gyrus including cuneus and calcarine, lingual gyrus, frontal middle, and inferior gyrus regions, precuneus, and cerebellum 6, crux 1 and 2 regions” [62]. They concluded that the hippocampus and amygdala subcortical regions in the medial temporal lobe are the most informative regions in the early AD. In [65], the authors segmented the sMRI into GM and WM and divided the GM into 85 cortical and subcortical anatomical ROIs. In [69], the authors performed the segmentation of the hippocampus from other regions and create a binary mask for the individual hippocampus. After that, the centroid of each hippocampus was calculated, and patches were extracted from the centroid for further implementation. The ROI-based approach has the advantage of low feature dimension and easy interpretation, but it ignores all the details of abnormalities. Still, research is ongoing to identify the most informative or affected part of the brain due to AD [33,118].

4. Patch-Based Input Management: A patch is a 3D cube; in this approach, disease-related patterns are identified by mining the features from patches. The core issue is to select the most discriminative patches for expressing both patch-level and image-level features [32]. In [65], the segmented ROI is additionally sectioned into smaller regions of variable dimensions, termed as patches, for implementing the multiscale method. This method states that the signal extracted at multiple scales from the fine-scale to the coarse-scale can be applied concurrently to detect deviations due to AD. The sizes of the patches were decided based on the consideration to retain sufficient comprehensive information as well as to avoid large feature dimensions. The subdivision of ROI into patches is performed by the k-means clustering algorithm [119], where ROI can be clustered by using spatial coordinates. In [69], a binary mask was created by means of the segmentation of each hippocampus, and 3D patches were created by cropping at the centroid of each mask. Three-dimensional patches decomposed further into two patches, for the internal and external hippocampus. The benefit of this approach is that it is sensitive to small changes, but selecting the most informative patch is still a challenge. The input management method-

wise analysis of different pre-processing techniques for neuroimaging biomarkers has been summarized in Table S4.

We conclude that input management is also a critical issue. A 2D-slice input has the advantage of fewer training parameters and a less complex network, but also has the drawback of losing spatial dependency between nearby slices. Voxel-based input management considers all the brain information but treats all regions equally, without any consideration of the anatomical structure, and has the drawback of high feature dimensionality and load on the processors. An ROI-based input is easily interpretable and has been used in clinical diagnosis, but in this method, the entire brain is represented by fewer features, and it provides knowledge about only a particular part of the brain, such as the hippocampus. Research is still ongoing to recognize the uses of other brain regions in detecting AD. Sometimes the ignorance of small abnormalities in the ROI may be the cause of damage to discriminative information and restrict the true command of extracted features. Nevertheless, patch-based input management can competently handle large feature dimensions and are responsive to small changes [120]. In Table 5, the distribution of input management approaches in the included studies is given.

Table 5. Dissemination of input management methods.

Input Management Approach	Percentage	References	N
Voxel-Based	38%	[59,60,64,66,70]	5
Slice-Based	31%	[58,63,67,68]	4
Patch-Based + ROI Based	15%	[65,69]	2
ROI Based	8%	[62]	1
Voxel-based + ROI-based + Patch-based + Slice-based	8%	[61]	1

3.4. Answer to RQ4

Out of 13, two studies performed multiclass classification and 11 studies performed binary classification. The accuracy performance metric was used consistently in all the included studies. In [58], the authors performed multiclass (AD, NC, sMCI, pMCI, SMC, MCI, respectively) classification by using the softmax classifier and rs-fMRI, and achieved an average accuracy of 97.92% and 97.98% for the off the shelf and fine-tuned models, respectively. In [63], the authors performed three-way discrimination among the NC, sMCI, and pMCI, respectively, by using the GoogleNet/softmax and CaffeNet/softmax, and obtained an overall accuracy score of 83.23% and 87.78%, respectively. In [59,60], the authors implemented all possible binary classifications for distinguishing NC, sMCI, and pMCI (AD vs. NC, NC vs. sMCI, pMCI vs. sMCI, NC vs. pMCI, and AD vs. sMCI, AD vs. pMCI) by using T1wMRI. In [61], the authors performed binary classification for AD vs. NC and sMCI vs. pMCI by using three input management approaches (ROI-based, patch-based, subject-level); for implementing all these methods, three different datasets and T1wMRI were used. Datasets are discussed in detail in Section 3.5. In [62], the authors performed binary classifications tasks (AD vs. NC, sMCI vs. pMCI, NC vs. pMCI, AD vs. sMCI) by using ResNet, SVM classifier and sMRI. In [64], the authors performed only one binary classification task (sMCI vs. pMCI) by using CaffeNet and sMRI biomarkers. They combined PCA and SFS for feature selection and finally adopted SVM for the prediction. In [65], the authors used multimodality (FDG-PET metabolism imaging and sMRI) and an ensemble of classifiers, multiple networks were trained, and then they chose the final classification result (PCA followed by SVM) to identify subjects classes (AD vs. NC, sMCI vs. pMCI). In [66], the authors performed the binary classification for AD vs. NC and sMCI vs. pMCI by using multimodality (FDG and AV-45 PET) and softmax classifier. In [67], the authors performed the binary classification sMCI vs. pMCI by using FDG-PET images, AlexNet, and SFS feature selection followed by an SVM classifier. In [68], the

authors used three different models (VGG16-from scratch, VGG16-TL, InceptionV4-TL), softmax classifier, and sMRI biomarkers for the binary classification of AD vs. NC. In [69], the authors conducted the multi-class classification (AD vs. NC, sMCI vs. pMCI, and sMCI vs. NC) by constructing DenseNets on the decomposed patches of the internal and external hippocampus of T1wMRI images, cascaded by RNN to learn the features from both hippocampi and the softmax classifier. The authors also compared the proposed method by implementing other commonly used hippocampus analysis methods such as disease detection by shape and volume analysis of the hippocampus. In [70], the authors performed a binary classification task (sMCI vs. pMCI) by extending feature-based transfer learning with knowledge transfer. They introduced a surrogate age marker that was captured through pre-training, and then they applied fine-tuning for the final prediction by combining the output of 3D CNN and surrogate age marker. They used T1wMRI and a fully connected layer obtained through flattening for predicting the value of surrogate age marker.

In Table 6, the accuracy achieved according to the DL and TL combinations is presented. It can be seen that the maximum accuracy achieved for classifying the sMCI vs.pMCI is 87.78%, obtained through the combination of CaffeNet (training) and ImageNet(transfer learning).

Table 6. The accuracy achieved according to the DL and TL combinations.

DL and TL Combination	AD vs. NC	pMCI vs. sMCI	Ref.
2D CNN + ResNet-18	97.92%	-	[58]
3D CNN + Local Transfer Learning	98.20%	74.90%	[59]
3D CNN + CAE + Local Transfer Learning	85.24% ± 3.97%	73.23% ± 4.21%	[60]
3D CNN + ICAE + Local Transfer Learning	86.60% ± 3.66%	73.95% ± 4.82%	[61]
3D CNN + Local Transfer Learning	88.00% ± 3.00%	78.00%±7.00%	[61]
2DCNN + Local Transfer Learning	79.00% ± 4.00%	-	[62]
3D CNN + Deep ResNet	91.30%	77.80%	[62]
CaffeNet + ImageNet	-	87.78%	[63]
GoogleNet + ImageNet	-	83.23%	[64]
2D CNN + CaffeNet	-	77.98%	[64]
DNN +SAE + Local Transfer Learning	93.58%	82.50%	[65]
Deep CNN + Local Transfer Learning	96.00%	84.20%	[66]
AlexNet + Local Transfer Learning	-	72.19%	[67]
VGG16(from scratch)	74.12%	-	[68]
VGG16 + AlexNet(ImageNet)	92.30%	-	[68]
Inception V4 + ALEXNet(ImageNet)	96.25%	-	[69]
DenseNets + RNN+ Local Transfer Learning	89.10%	72.50%	[69]
3D CNN + local Transfer Learning	-	76.00%	[70]

In Table 7, the accuracy achieved according to the input management and neuroimaging biomarkers is presented. It can be seen that the maximum accuracy achieved for classifying the sMCI vs. pMCI is 87.78%, obtained through the combination of slice-based and sMRI.

Table 7. The accuracy achieved according to the neuroimaging data and input management approach.

Input Management with Biomarker	AD vs. NC	pMCI vs. sMCI	References
Slice-based + rs-fMRI	97.92%	-	[58]
Voxel-based + 3T T1wMRI	98.20%	74.90%	[59]
Voxel-based + 1.5T T1w MP-RAGE MRI	86.60% ± 3.66%	73.95% ± 4.82%	[60]
Voxel-based + T1wMRI	85.00% ± 4.00%	73.00% ± 5.00%	[61]
ROI-based + T1wMRI	88.00% ± 3.00%	78.00% ± 7.00%	
Patch-based + T1wMRI	83.00% ± 2.00%	77.00% ± 4.00%	
Slice-based + T1wMRI	79.00% ± 4.00%	-	[62]
ROI-based + sMRI	91.30%	77.80%	
Slice-based + sMRI	-	87.78%	[63]
Voxel-based + sMRI	-	77.98%	[64]
ROI-based + Patch-based + sMRI + 18FDG-PET	93.58%	82.50%	[65]
Voxel-based + 18F FDG-PET + AV-45 PET	96.00%	84.20%	[66]
Slice-based + 18F FDG-PET	-	72.19%	[67]
Slice-based + sMRI	96.25%	-	[68]
ROI-based + Patch-based + 3T T1wMRI	89.10%	72.50%	[69]
Voxel-based + 3T T1wMRI		76.00%	[70]

3.5. Answer to RQ5

Numerous online datasets of neuroimaging biomarkers are available that have been developed with the motive of providing the free availability of neuroimaging data to the community of scientists to assist in forthcoming discoveries in the neurosciences. Datasets such as ADNI [121], OASIS [122], AIBL [123], IXI [124], and MIRIAD [125] are freely available online.

ADNI: ADNI was developed by longitudinal multicenter studies. It was released in 2004 by the National Institute of Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), and some other national institutes as a USD 60 million, 5-year partnership project, based in North America. The acquisition of all the data was achieved through the ADNI protocol. To date, they have four variants: ADNI-1 (400 MCI, 200 sMCI, and 200 NC contribute to the research and were followed for 2–3 years, with MRI and FDG-PET biomarkers only), ADNI-Grand Opportunities (the existing ADNI-1 cohort + 200 EMCI subjects), ADNI-2 (ADNI-1, ADNI-GO cohort + 150 NC, 100 EMCI, 150 LMCI, and 150 mild AD subjects, 107 SMC subjects and also amyloid PET images have been added), and ADNI-3 (which began in 2016 and will remain up to 2022, and tau-PET has already been added as a key indicator for AD, while further research is ongoing for the discovery, optimization, standardization, and validation of clinical trial measures for AD detection). In this work, this is the most used dataset, being used in 92% of studies by itself or in combination with other datasets.

OASIS: OASIS-3 is the latest release in the OASIS series. The formerly released OASIS-Cross-sectional and OASIS-Longitudinal datasets have been utilized for AD research, OASIS-3 is the longitudinal, cognitive biomarker dataset for normal aging and AD. It contains the neuroimaging data generated from 2168 MRI sessions and 1608 PET sessions by using 1098 subjects (609 NC and 489 people at different stages of cognitive debility) in the age range of 43-95 years. It was used in 16% of the studies in our review.

AIBL: This is known as Australian ADNI, and adds scientific value to the ADNI cohort. It contains 250 MR/flutemetamol, 200 MR/florbetapir, and 50 MR/PIB images. It was used in 8% of the studies in our review.

IXI: This contains 600 images from normal, healthy subjects. Researchers can download T1wMRI, T2wMRI, and DTI images directly in NIFTI format. It was utilized in 8% of the studies in our review.

MIRIAD: This is a database consisting of MRI scans of 46 AD and 23 NC subjects, and scans were collected at interims from 2 weeks to 2 years. This study has the objective to examine the viability of using MRI for clinical trials of AD.

Meanwhile, several studies preferred to use their own datasets. In [59], the authors designed a self-regulating dataset “MILAN” of 3D T1 weighted images, acquired from 229 subjects (124 AD, 50 MCI, 55 NC), recruited at the Neurology Department, Scientific Institute, and University Vita-Salute San Raffaele, Milan. The distribution of the discussed datasets in the included studies is given in Table 8.

Table 8. Distribution of datasets.

Datasets	Used Percentage	Reference	N
ADNI	69.23%	[43,45,47–52,54]	9
OASIS	7.69%	[68]	1
IXI + ADNI	7.69%	[70]	1
OASIS + ADNI + AIBL	7.69%	[60]	1
Local (MILAN) + ADNI	7.69%	[59]	1

Only some studies used the combination of different types of datasets. For developing a clinically acceptable system, researchers should use data of different homogeneities [126]. The size of the datasets must be balanced either by taking the same number of images for training and testing or by redefining the loss function. Other points such as patient overlapping, set sampling or ground truth should also be considered for the uniform distribution of data and avoiding data leakage during training and testing. Techniques for avoiding data leakage and overfitting are discussed in Section 3.7. Furthermore, all the other details, such as sample size and the combination of different datasets from all included articles, has been summarized in Table S3.

3.6. Answer to RQ 6

Different software packages are available to assist researchers in performing the pre-processing of neuroimaging data and the implementation of DL and TL methods. The following brain image analysis packages were used for the pre-processing of neuroimaging biomarkers in the included articles. The pre-processing pipeline implemented by these packages is given in Table S4.

FSL [127]: This is the all-inclusive library of tools for analyzing sMRI, fMRI, and DTI neuroimaging data. Most of the tools can be run either by command line or GUI interface. It was used in [58,59].

SPM12 [128]: This is intended for the analysis of neuroimaging data sequences; the sequences can be a succession of images from diverse allies. The current release, updated on 13 January 2020, is specifically intended for the analysis of fMRI and PET images. It was used in [59,60,62,64].

Nipype [129]: This is an open-source Python project developed by the Nipy community [130], provides an even interface to the currently available brain imaging software, and facilitates communication amongst these packages (for example, AFNI [131], ANTS [132], Camino [133], FreeSurfer [134], FSL, MNE [135], Slicer [136], SPM [137]) within a single flow. It eases the learning curve required to use diverse packages. One of the most important advantages is that it makes one's research reproducible and shares one's processing workflows with the research community. It was used in [61].

MATLAB [138]: The image processing toolbox permits one's to mechanize common image processing workflows. The researcher can section image data, perform batch processing for large datasets, perform the registration of images, create histograms, and manipulate ROIs. MATLAB was used in [63,68].

NifTi_2014 toolkit [139]: This is a tool for analyzing and processing neuroimaging images that can be loaded by MRIcro software [140] and MATLAB2015b [141]. It was used in [67].

FreeSurfer [134]: This is an open-source collection for analyzing and processing MRI images, developed by the Laboratory for Computational Neuroimaging (LCN), USA. It is a set of automated tools for the renewal of the brain cortical surface from sMRI and the overlap of fMRI onto the recreated surfaces. It was used in [65].

MRICron [140]: This allows users to view medical images in various formats, and creates format headers for exporting neuroimaging biomarkers to other platforms. It is a stand-alone program but includes tools to analyze MRI, fMRI, and PET images. It can be used for the efficient viewing and exporting of neuroimaging data and to identify ROIs. It was used in [70].

Moreover, in [66], the authors did not perform any pre-processing of FDG and AV-45 PET images. Pre-processed images were downloaded from the ADNI at the most progressive pre-processing stage, and were used for deep CNN training and testing. The dissemination of pre-processing software in the included articles is given in Table 9.

Table 9. Distribution of pre-processing software.

Pre-Processing Software	Percentage	References	N
FSL	17%	[58,69]	2
SPM12	34%	[59,60,62,64]	4
MATLAB	17%	[63,68]	2
FreeSurfer	8%	[65]	1
NifTi_2014 toolkit	8%	[67]	1
Nipype	8%	[61]	1
MRICron	8%	[70]	1

Software packages such as CAFFE [85], Keras [142], Tensorflow [143], Theano [144], Pytorch [145], MatConvNet [146], and the Deep Learning toolbox [147] have been used for implementing DL and TL algorithms. In Table 10, details of the packages used in the included articles are given. The distribution of software packages in the included articles is given in Table 11.

Table 10. Software packages used for implementing DL and TL algorithms.

Ref	Software Architectures Used for Implementing DL and TL Algorithms
[58]	Implemented on CAFFE Framework and trained on theFloydHub cloud service with GPU Tesla K80.
[59]	Python using Theano by using Dell Poweredge T630 Linux, including high-performance GPU NVIDIA Tesla K40, with 2880 CUDA cores and High-Frequency Intel Xeon E5 –2623 v3 with 78 GB memory overall.
[60]	Not mentioned
[61]	The DL models were built using the Pytorch library 5. TensorboardX 6 was embedded into the current framework to dynamically monitor the training process.
[62]	Training and testing routines were implemented on an NVIDIA CUDA parallel computing platform using GPU-accelerated CUDA toolkit/compilation and Pytorch python package tensor libraries
[63]	CAFFE Framework with NVIDIA DIGITS and NVIDIA Quadro M4000 (GPU).
[64]	Not mentioned
[65]	Open-source Deep Learning toolbox
[66]	MatConvNet
[67]	CAFFE Framework
[68]	Keras with TensorFlow backend
[69]	PC with Ubuntu14.04-x64/GPU NVIDIA GTX1080Ti of 11GB memory, python 2.7.9, Keras library with the Tensorflow backend.
[70]	Not mentioned

Table 11. Dissemination of software packages used for implementing DL models.

Software Packages	Percentage	References	N
CAFFE framework	30%	[58,63,67]	3
Tensorflow + Keras	20%	[68,69]	2
Pytorch	20%	[61,62]	2
Deep Learning Tool-Box	10%	[65]	1
Theano	10%	[59]	1
MatConvNet	10%	[66]	1

3.7. Answer to RQ7

Dataset size has a notable impact on the performance of classifiers on an unseen test dataset [44]. The available datasets of AD and MCI subjects are of a relatively small size, with only a few hundred samples. DL algorithms tend to be easily over-fitted when trained on a lower number of samples due to a large number of training parameters. In this section, we discuss the techniques used in the included articles for reducing overfitting.

Data Augmentation: This is a way of increasing the heterogeneity of training datasets without collecting new data. It generates new data samples from the existing data. It can be categorized into two methods. The first category is transformation methods, which cover a mixture of simple transformations such as random translation, rotation, reflection, distortion, blurring and flipping, cropping, noise injection, gamma correction, scaling, and intensity variations by arbitrarily adjusting brightness, contrast, saturation, and hue on the training data. Transformation methods were used in [59,60] to improve the classifier performance. The second type of method is neuroimaging data synthesis [148], intended to generate a new dataset that shares features with the source dataset [68]. It can be implemented by using autoencoders (AE) [149] and generative adversarial networks

(GAN) [150,151] for neuroimaging biomarkers. Nevertheless, this area needs to be explored and the effectiveness of the synthesis images for predicting AD still must be proved. In [58], the authors obtained 6120 2D images from each fMRI scan by using pre-processing techniques. In [63], the authors employed a novel strategy for data augmentation, based on the image integration method Shin [152], to create sample image patches from MRI. It uses the information from training data adequately to implement the constraint of the three-channel input of CNN. In [66], the authors performed the augmentation of PET images by flicking them in the left–right direction. In [69], the hippocampus centroid is shifted by ± 2 voxels in the x and z directions to extract more patches. Technical insights of the augmentation approaches from included articles are given in Table S3. Nevertheless, some studies did not perform augmentation at all [62,64,65,67,68,70], and relied only on TL methods to reduce overfitting.

Transfer Learning: This is the idea of utilizing a model trained on a certain chore (for example, the ImageNet classification task or unsupervised learning task), conducive to achieving the task of AD classification or predicting the progression of MCI to AD. TL methods may reduce the overfitting by providing a better weight initialization of DL models. It also reduces the time needed for training the DL models and provides better performance compared to training from scrap [153].

Regularization: This can be performed by using different approaches, such as dropout, weight decay, and L1/L2 regularization. Dropout relies on the idea of arbitrarily and independently dropping neurons, setting their output value equal to zero, which makes the network less intricate and less vulnerable to overfitting. It improves the generalization capability of the DL model. The number of nodes chosen varied from 25% to 50% from one study to another. Weight decay also increases the generalizability of the model by regularizing the updated weights, multiplying them by a factor slightly smaller than one; it also reduces the complexity of the model. Dropout and weight were employed in several of the included studies [61,62,65,69,70].

Batch Normalization: This is a technique for training DL models, which standardizes the input to a layer for each mini-batch. It speeds up the training process and increases the performance and generalizability of the models.

Early stopping: This involves stopping the training process at a former point. It helps in determining the number of iterations needed for training the model, before being critically overfitted. It was used in several of the included studies [62,65]. In [65], the validation set was only used to define the early stopping time-point.

Cross-Validation: In this technique, the database is divided into training and testing sets many times by keeping the same proportion but rotating the instances every time. It is a statistical method for testing the performance of classifiers. The most-used cross-validation method is k-fold. In our review, the value of k ranges between 5 and 10. It was used in 85% of included articles

Another major issue that was taken care of by the researchers is data leakage, which also leads to the overfitting and biased evaluation of classification algorithms. It occurs because of using test data in the training process [10]. The following four main causes of data leakage were noticed.

Wrong data split: This implies that images of similar subjects are used at several points in time (training, validation, and testing) [154]. For unbiased evaluation or to reduce overfitting, the researchers have to split the data at the subject level, not at the image level.

Late Split: Perform data augmentation after splitting the datasets for an unbiased evaluation. If augmentation is conducted before the split, the same images may be found at multiple points in time.

Prejudice Transfer Learning: When the source and destination targets overlap, when a model is trained during the task of AD vs. NC classification and used for initializing the weights of another task MCI vs. AD, in this case, AD subjects used in the training set of the source task can be in the testing or validation set of the target task. Researchers have to use different datasets for the source and destination, as the authors did in [59,60,65,66].

Absence of an independent validation set: An unrelated validation set (from the test set) should be used for hyper parameter optimization.

3.8. Answer to RQ8

Although DL methods with the TL approach have shown better results, the following points still need to be taken care of for developing a clinically acceptable system. The progression detection of MCI to AD is more significant when compared to others for subjects and their loved ones, for future planning, and also for doctors to target those patients that should undergo suitable treatments. As can be seen from Tables 7 and 8, the maximum accuracy achieved for pMCI vs. sMCI is 87.78% to date, which needs to be improved. Table 7 shows that multimodality performs better than a single modality. The new combination of existing biomarkers and unexplored biomarkers such as tau-PET can be used in the future for predicting the progression of AD.

Despite applying augmentation and TL to avoiding overfitting, the non-availability of large neuroimaging datasets causes generalizability issues. Neuroimaging data synthesis can be employed by researchers in the future to produce new images from prevailing images by using GAN and AE. Still, the viability of synthesis neuroimaging data has to be proven. Owing to the complexity of DL models and 3D neuroimaging biomarkers, it is problematic to display which precise features have been extracted and also to regulate how those features lead to the inference and comparative prominence of specific features [155]. This makes it complex to correct any prejudices that arise from the input datasets. Filter visualization and activation maps [156] were used in [65] to understand the features that contributed significantly to determining the output. It is a challenge to explore the importance of the anatomy of neuroimaging data in predicting the progression of AD for implementing a transparent DL model. The changeableness of configuration and the fortuity involved in training make it challenging to replicate the study and attain identical results. In the future, researchers must reproduce important findings from DL and TL approaches for entirely independent datasets.

4. Conclusions

Alzheimer's disease is one of the leading reasons for death, specifically in developed countries. The timely identification of AD is a demanding task. Therefore, computer-based systems are needed to support physicians in accurate disease classification. DL with TL approaches have been employed in recent years to solve this problem. We started this review with the concept of AD, MCI, tailed by the discussion of existing standards for diagnosis of AD by using MRI, fMRI and PET neuroimaging biomarkers. These biomarkers can be used with DL and TL methods and can be added as an additional tool with some other factors, such as genetic data, MMSE, and Mini-Cog Test results, for better classification. Minimum pre-processing (intensity normalization, registration) is suggested for neuroimaging modalities. Multi-modality can be used for better feature extraction. CNNs with different TL approaches were used in most of the studies. CNN-based DL architectures such as CaffeNet or GoogleNet can be combined with pre-trained models such as ImageNet, VGG16, or inception v4 for implementing more accurate systems. ROI- and patch-based methods are reliable due to their ability to collect relevant information from a brain scan. In addition, GAN may be used for producing novel images from prevailing images for data augmentation. Nevertheless, some challenges such as overfitting concerning the use of small datasets and reproducibility regarding the randomness involved in the training need to be resolved soon. Moreover, it is difficult to determine which features are more significant for an accurate decision support system. This research presents a comparative analysis of the state of the art on DL and TL methods and neuroimaging biomarkers for the early detection of AD. The authors have also analyzed the different input management and pre-processing techniques of neuroimaging biomarkers, different sources of neuroimaging datasets, and the limitations that have been discussed.

State-of-the-art techniques used by researchers for handling overfitting and reproducibility issues have been discussed, along with the software platforms needed to develop DL and TL methods. The pre-processing of biomarkers has also been analyzed. The results show that due to the large size of 3D data, pure DL models took a lot of time in training, as weight initialization was performed randomly. In studies that reported good accuracy, researchers applied only local TL or the combination of pre-trained 3D CNN networks with local TL. These results can be further improved by identifying the most effective combinations of DL and TL with the appropriate biomarkers by conducting continuous experimentation. We identified that the maximum accuracy achieved to date for the early detection of AD is 87.78%, which should be improved in future research. Around 31% of studies used sMRI, while 38% used T1wMRI and other biomarkers such as rs-fMRI, 18F-FDG-PET, and AV45-PET. We did not find any implementations of DL with TL that use tau-PET or amyloid-PET. In total, 77% of the studies used voxel and slice-based input management approaches for handling neuroimaging biomarkers, and only 23% of studies used ROI- and patch-based approaches. Around 50% of studies used intensity normalization and registration pre-processing methods, while only 8% of the studies used the biomarkers without any pre-processing methods. The CAFFE framework was used by 30% of the studies for implementing DL architectures, and SPM12 was used by 34% of the studies for the pre-processing of biomarkers, 68% studies used ADNI datasets for training as well as testing purposes, only 8% of the studies used their local datasets in combination with ADNI. Future researchers must use different datasets for training and testing to develop a reproducible system.

We hope that this paper will help researchers to obtain a reliable overview of the state of the art for developing a clinically acceptable system for the early detection of AD.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/s21217259/s1>.

Author Contributions: D.A. and G.M. participated in the conception and methodology of the study and manuscript writing. I.d.I.T.-D., M.A.F.M., B.G.Z. and F.M.R. participated in the review and manuscript writing. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been partially supported by European Commission and the Ministry of Industry, Energy and Tourism under the project AAL-20125036 named BWetake Care: ICTbased Solution for (Self-) Management of Daily Living.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hardy, J. Amyloid, the presenilins and Alzheimer's disease. *Trends Neurosci.* **1997**, *20*, 154–159. [[CrossRef](#)]
2. Facts and Figures. Alzheimer's Disease and Dementia. Available online: <https://www.alz.org/alzheimers-dementia/facts-figures> (accessed on 16 January 2021).
3. Jicha, G.A.; Parisi, J.E.; Dickson, D.W.; Johnson, K.; Cha, R.; Ivnik, R.J.; Tangalos, E.G.; Boeve, B.F.; Knopman, D.S.; Braak, H. Neuropathologic outcome of mild cognitive impairment following progression to clinical dementia. *Arch. Neurol.* **2006**, *63*, 674–681. [[CrossRef](#)]
4. Klöppel, S.; Stonnington, C.M.; Chu, C.; Draganski, B.; Scahill, R.I.; Rohrer, J.D.; Fox, N.C.; Jack, C.R., Jr.; Ashburner, J.; Frackowiak, R.S.J. Automatic classification of MR scans in Alzheimer's disease. *Brain* **2008**, *131*, 681–689. [[CrossRef](#)]
5. Petersen, R.C. Mild cognitive impairment as a diagnostic entity. *J. Intern. Med.* **2004**, *256*, 183–194. [[CrossRef](#)] [[PubMed](#)]
6. Dubois, B.; Feldman, H.H.; Jacova, C.; DeKosky, S.T.; Barberger-Gateau, P.; Cummings, J.; Delacourte, A.; Galasko, D.; Gauthier, S.; Jicha, G.; et al. Research criteria for the diagnosis of Alzheimer's disease: Revising the NINCDS-ADRDA criteria. *Lancet Neurol.* **2007**, *6*, 734–746. [[CrossRef](#)]
7. McKhann, G.; Drachman, D.; Folstein, M.; Katzman, R.; Price, D.; Stadlan, E.M. Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* **1984**, *34*, 939. [[CrossRef](#)] [[PubMed](#)]

8. Folstein, M.F.; Folstein, S.E.; McHugh, P.R. “Mini-mental state”: A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* **1975**, *12*, 189–198. [[CrossRef](#)]
9. Morris, J.C. The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology* **1993**, *43*, 2412–2414. [[CrossRef](#)] [[PubMed](#)]
10. Rathore, S.; Habes, M.; Iftikhar, M.A.; Shacklett, A.; Davatzikos, C. A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer’s disease and its prodromal stages. *NeuroImage* **2017**, *155*, 530–548. [[CrossRef](#)] [[PubMed](#)]
11. Allen, G.I.; Amorosso, N.; Anghel, C.; Balagurusamy, V.; Bare, C.J.; Beaton, D.; Bellotti, R.; Bennett, D.A.; Boehme, K.L.; Boutros, P.C.; et al. Crowdsourced estimation of cognitive decline and resilience in Alzheimer’s disease. *Alzheimer’s Dement.* **2016**, *12*, 645–653. [[CrossRef](#)]
12. Cerasa, A.; Sarica, A.; Calhoun, V.; Quattrone, A. A Machine Learning Neuroimaging Challenge for Automated Diagnosis of Mild Cognitive Impairment. Available online: <https://www.researchgate.net/project/A-Machine-learning-neuroimaging-challenge-for-automated-diagnosis-of-Mild-Cognitive-Impairment> (accessed on 18 January 2021).
13. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 7553. [[CrossRef](#)] [[PubMed](#)]
14. Samper-González, J.; Burgos, N.; Bottani, S.; Fontanella, S.; Lu, P.; Marcoux, A.; Routier, A.; Guillon, J.; Bacci, M.; Wen, J.; et al. Reproducible evaluation of classification methods in Alzheimer’s disease: Framework and application to MRI and PET data. *NeuroImage* **2018**, *183*, 504–521. [[CrossRef](#)] [[PubMed](#)]
15. Plis, S.M.; Hjelm, D.R.; Salakhutdinov, R.; Allen, E.A.; Bockholt, H.J.; Long, J.D.; Johnson, H.J.; Paulsen, J.S.; Turner, J.A.; Calhoun, V.D. Deep learning for neuroimaging: A validation study. *Front. Neurosci.* **2014**, *8*, 229. [[CrossRef](#)] [[PubMed](#)]
16. Dauphin, G.M.Y.; Glorot, X.; Rifai, S.; Bengio, Y.; Goodfellow, I.; Lavoie, E.; Muller, X.; Desjardins, G.; Warde-Farley, D.; Vincent, P.; et al. Unsupervised and transfer learning challenge: A deep learning approach. In *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*; JMLR Workshop and Conference Proceedings; MIT Press: Cambridge, MA, USA, 2012; pp. 97–110. Available online: <http://proceedings.mlr.press/v27/mesnil12a.html> (accessed on 19 January 2021).
17. Raina, R.; Ng, A.Y.; Koller, D. Constructing informative priors using transfer learning. In *Proceedings of the 23rd International Conference on Machine Learning*, New York, NY, USA, 25–29 June 2006; pp. 713–720. [[CrossRef](#)]
18. Zhou, L.; Wang, Y.; Li, Y.; Yap, P.-T.; Shen, D. Alzheimer’s Disease Neuroimaging Initiative (ADNI). Hierarchical Anatomical Brain Networks for MCI Prediction: Revisiting Volumetric Measures. *PLoS ONE* **2011**, *6*, e21935. [[CrossRef](#)]
19. Ker, J.; Wang, L.; Rao, J.; Lim, T. Deep Learning Applications in Medical Image Analysis. *IEEE Access* **2018**, *6*, 9375–9389. [[CrossRef](#)]
20. PRISMA. Available online: <http://prisma-statement.org/prismastatement/flowdiagram> (accessed on 20 January 2021).
21. Bozada, T.; Borden, J.; Workman, J.; Del Cid, M.; Malinowski, J.; Luechtefeld, T. Sysrev: A Fair Platform for Data Curation and Systematic Evidence Review. *Front. Artif. Intell.* **2021**, *4*, 105. [[CrossRef](#)] [[PubMed](#)]
22. Sysrev. Alzheimer’s Survey. Available online: <https://sysrev.com/u/1468/p/36874> (accessed on 22 October 2021).
23. Cuingnet, R.; Gerardin, E.; Tessieras, J.; Auzias, G.; Lehéricy, S.; Habert, M.-O.; Chupin, M.; Benali, H.; Colliot, O.; Alzheimer’s Disease Neuroimaging Initiative. Automatic classification of patients with Alzheimer’s disease from structural MRI: A comparison of ten methods using the ADNI database. *NeuroImage* **2011**, *56*, 766–781. [[CrossRef](#)]
24. Retico, A.; Bosco, P.; Cerello, P.; Fiorina, E.; Chincarini, A.; Fantacci, M.E. Predictive Models Based on Support Vector Machines: Whole-Brain versus Regional Analysis of Structural MRI in the Alzheimer’s Disease. *J. Neuroimaging* **2015**, *25*, 552–563. [[CrossRef](#)]
25. Zhu, Y.; Zhu, X.; Kim, M.; Shen, D.; Wu, G. Early diagnosis of Alzheimer’s disease by joint feature selection and classification on temporally structured support vector machine. In *Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2016—19th International Conference*, Athens, Greece, 17–21 October 2016; pp. 264–272. [[CrossRef](#)]
26. Beheshti, I.; Demirel, H.M. Alzheimer’s Disease Neuroimaging Initiative. Classification of Alzheimer’s disease and prediction of mild cognitive impairment-to-Alzheimer’s conversion from structural magnetic resource imaging using feature ranking and a genetic algorithm. *Comput. Biol. Med.* **2017**, *83*, 109–119. [[CrossRef](#)]
27. Sun, Z.; Qiao, Y.; Lelieveldt, B.P.F.; Staring, M. Integrating spatial-anatomical regularization and structure sparsity into SVM: Improving interpretation of Alzheimer’s disease classification. *NeuroImage* **2018**, *178*, 445–460. [[CrossRef](#)]
28. Zeng, N.; Qiu, H.; Wang, Z.; Liu, W.; Zhang, H.; Li, Y. A new switching-delayed-PSO-based optimized SVM algorithm for diagnosis of Alzheimer’s disease. *Neurocomputing* **2018**, *320*, 195–202. [[CrossRef](#)]
29. Li, R.; Zhang, W.; Suk, H.-I.; Wang, L.; Li, J.; Shen, D.; Ji, S. Deep Learning Based Imaging Data Completion for Improved Brain Disease Diagnosis. In *Medical Image Computing and Computer Assisted Intervention*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2014.
30. Ortiz, A.; Munilla, J.; Górriz, J.M.; Ramírez, J. Ensembles of Deep Learning Architectures for the Early Diagnosis of the Alzheimer’s Disease. *Int. J. Neural. Syst.* **2016**, *26*, 1650025. [[CrossRef](#)]
31. Lu, D.; Popuri, K.; Ding, G.W.; Balachandar, R.; Beg, M.F.; Alzheimer’s Disease Neuroimaging Initiative. 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*, 5697. [[CrossRef](#)] [[PubMed](#)]
32. Liu, M.; Zhang, J.; Adeli, E.; Shen, D. Landmark-Based Deep Multi-Instance Learning for Brain Disease Diagnosis. *Med Image Anal.* **2018**, *43*, 157–168. [[CrossRef](#)] [[PubMed](#)]
33. Suk, H.-I.; Lee, S.-W.; Shen, D.; The Alzheimer’s Disease Neuroimaging Initiative. Latent feature representation with stacked auto-encoder for AD/MCI diagnosis. *Brain. Struct. Funct.* **2015**, *220*, 841–859. [[CrossRef](#)] [[PubMed](#)]

34. Hierarchical Feature Representation and Multimodal Fusion with Deep Learning for AD/MCI Diagnosis—ScienceDirect. Available online: <https://www.sciencedirect.com/science/article/abs/pii/S1053811914005540?via%3Dihub> (accessed on 23 October 2021).
35. Li, F.; Tran, L.; Thung, K.-H.; Ji, S.; Shen, D.; Li, J. A Robust Deep Model for Improved Classification of AD/MCI Patients. *IEEE J. Biomed. Health Inform.* **2015**, *19*, 1610–1616. [[CrossRef](#)] [[PubMed](#)]
36. Shakarami, A.; Tarrach, H.; Mahdavi-Hormat, A. A CAD system for diagnosing Alzheimer’s disease using 2D slices and an improved AlexNet-SVM method. *Optik* **2020**, *212*, 164237. [[CrossRef](#)]
37. Ieracitano, C.; Mammone, N.; Bramanti, A.; Hussain, A.; Morabito, F.C. A Convolutional neural network approach for classification of dementia stages based on 2D-spectral representation of EEG recordings. *Neurocomputing* **2019**, *323*, 96–107. [[CrossRef](#)]
38. Basheera, S.; Ram, M.S.S. A novel CNN based Alzheimer’s disease classification using hybrid enhanced ICA segmented gray matter of MRI. *Comput. Med. Imaging Graph.* **2020**, *81*, 101713. [[CrossRef](#)] [[PubMed](#)]
39. Basheera, S.; Ram, M.S.S. Convolution neural network-based Alzheimer’s disease classification using hybrid enhanced independent component analysis based segmented gray matter of T2 weighted magnetic resonance imaging with clinical valuation. *Alzheimer’s Dement. Transl. Res. Clin. Interv.* **2019**, *5*, 974–986. [[CrossRef](#)]
40. Shahamat, H.; Abadeh, M.S. Brain MRI analysis using a deep learning based evolutionary approach. *Neural Netw.* **2020**, *126*, 218–234. [[CrossRef](#)] [[PubMed](#)]
41. Wee, C.-Y.; Liu, C.; Lee, A.; Poh, J.S.; Ji, H.; Qiu, A. Cortical graph neural network for AD and MCI diagnosis and transfer learning across populations. *NeuroImage Clin.* **2019**, *23*, 101929. [[CrossRef](#)] [[PubMed](#)]
42. Qiu, S.; Chang, G.H.; Panagia, M.; Gopal, D.M.; Au, R.; Kolachalama, V.B. Fusion of deep learning models of MRI scans, Mini-Mental State Examination, and logical memory test enhances diagnosis of mild cognitive impairment. *Alzheimer’s Dement.* **2020**, *10*, 737–749. [[CrossRef](#)]
43. Tufail, A.B.; Ma, Y.; Zhang, Q.-N. Multiclass classification of initial stages of Alzheimer’s disease through neuroimaging modalities and convolutional neural networks. In Proceedings of the IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC), Chongqing, China, 12–14 June 2020; pp. 51–56. [[CrossRef](#)]
44. Wang, H.; Shen, Y.; Wang, S.; Xiao, T.; Deng, L.; Wang, X.; Zhao, X. Ensemble of 3D densely connected convolutional network for diagnosis of mild cognitive impairment and Alzheimer’s disease. *Neurocomputing* **2019**, *333*, 145–156. [[CrossRef](#)]
45. Bi, X.; Li, S.; Xiao, B.; Li, Y.; Wang, G.; Ma, X. Computer aided Alzheimer’s disease diagnosis by an unsupervised deep learning technology. *Neurocomputing* **2020**, *392*, 296–304. [[CrossRef](#)]
46. Li, W.; Lin, X.; Chen, X. Detecting Alzheimer’s disease Based on 4D fMRI: An exploration under deep learning framework. *Neurocomputing* **2020**, *388*, 280–287. [[CrossRef](#)]
47. Zhang, F.; Li, Z.; Zhang, B.; Du, H.; Wang, B.; Zhang, X. Multi-modal deep learning model for auxiliary diagnosis of Alzheimer’s disease. *Neurocomputing* **2019**, *361*, 185–195. [[CrossRef](#)]
48. Zhang, B.; Wang, J.; Lin, L.; Wu, S. The Effect of Smoothing Filter on CNN based AD Classification. In Proceedings of the 2019 8th International Conference on Bioinformatics and Biomedical Science, New York, NY, USA, 29–31 October 2019; pp. 54–58. [[CrossRef](#)]
49. Khan, N.M.; Hon, M.; Abraham, N. Transfer Learning with intelligent training data selection for prediction of Alzheimer’s Disease. *arXiv* **2019**, arXiv:1906.01160. [[CrossRef](#)]
50. Rohini, M.; Surendran, D. Classification of Neurodegenerative Disease Stages using Ensemble Machine Learning Classifiers. *Procedia Comput. Sci.* **2019**, *165*, 66–73. [[CrossRef](#)]
51. Kruthika, K.R.; Maheshappa, H.D.; Alzheimer’s Disease Neuroimaging Initiative. CBIR system using Capsule Networks and 3D CNN for Alzheimer’s disease diagnosis. *Inform. Med. Unlocked* **2019**, *14*, 59–68. [[CrossRef](#)]
52. Jabason, E.; Ahmad, M.O.; Swamy, M.N.S. Hybrid Feature Fusion Using RNN and Pre-trained CNN for Classification of Alzheimer’s Disease (Poster). In Proceedings of the 2019 22th International Conference on Information Fusion (FUSION), Ottawa, ON, Canada, 2–5 July 2019; pp. 1–4.
53. Liu, M.; Li, F.; Yan, H.; Wang, K.; Ma, Y.; Alzheimer’s Disease Neuroimaging Initiative; Shen, L.; Xu, M. A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer’s disease. *Neuroimage* **2020**, *208*, 116459. [[CrossRef](#)] [[PubMed](#)]
54. Ji, H.; Liu, Z.; Yan, W.Q.; Klette, R. Early Diagnosis of Alzheimer’s Disease Using Deep Learning. In Proceedings of the 2nd International Conference on Control and Computer Vision, New York, NY, USA, 15–18 June 2019; pp. 87–91. [[CrossRef](#)]
55. Gao, F.; Yoon, H.; Wu, T.; Chu, X. A feature transfer enabled multi-task deep learning model on medical imaging. *Expert Syst. Appl.* **2020**, *143*, 112957. [[CrossRef](#)]
56. El-Sappagh, S.; Abuhmed, T.; Islam, S.M.R.; Kwak, K.S. Multimodal multitask deep learning model for Alzheimer’s disease progression detection based on time series data. *Neurocomputing* **2020**, *412*, 197–215. [[CrossRef](#)]
57. Yagis, E.; De Herrera, A.G.S.; Citi, L. Generalization Performance of Deep Learning Models in Neurodegenerative Disease Classification. In Proceedings of the 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), San Diego, CA, USA, 18–21 November 2019; pp. 1692–1698. [[CrossRef](#)]
58. Ramzan, F.; Khan, M.U.G.; Rehmat, A.; Iqbal, S.; Saba, T.; Rehman, A.; Mehmood, Z. A Deep Learning Approach for Automated Diagnosis and Multi-Class Classification of Alzheimer’s Disease Stages Using Resting-State fMRI and Residual Neural Networks. *J. Med. Syst.* **2019**, *44*, 37. [[CrossRef](#)]

59. Basaia, S.; Agosta, F.; Wagner, L.; Canu, E.; Magnani, G.; Santangelo, R.; Filippi, M.; Alzheimer's Disease Neuroimaging Initiative. Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. *Neuroimage Clin.* **2019**, *21*, 101645. [[CrossRef](#)] [[PubMed](#)]
60. Classification and Visualization of Alzheimer's Disease using Volumetric Convolutional Neural Network and Transfer Learning | Scientific Reports. Available online: <https://www.nature.com/articles/s41598-019-54548-6> (accessed on 13 September 2020).
61. Wen, J.; Thibeau-Sutre, E.; Diaz-Melo, M.; Samper-González, J.; Routier, A.; Bottani, S.; Dormont, D.; Durrleman, S.; Burgos, N.; Colliot, O. Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation. *Med. Image Anal.* **2020**, *63*, 101694. [[CrossRef](#)] [[PubMed](#)]
62. Abrol, A.; Bhattarai, M.; Fedorov, A.; Du, Y.; Plis, S.; Calhoun, V.; Alzheimer's Disease Neuroimaging Initiative. Deep Residual Learning for Neuroimaging: An application to Predict Progression to Alzheimer's Disease. *J. Neurosci. Methods* **2020**, *339*, 108701. [[CrossRef](#)] [[PubMed](#)]
63. Wu, C.; Guo, S.; Hong, Y.; Xiao, B.; Wu, Y.; Zhang, Q. Discrimination and conversion prediction of mild cognitive impairment using convolutional neural networks. *Quant. Imaging. Med. Surg.* **2018**, *8*, 992–1003. [[CrossRef](#)] [[PubMed](#)]
64. Lin, L.; Zhang, B. MCI Conversion Prediction Based on Transfer Learning. *DEStech Trans. Comput. Sci. Eng.* **2018**. [[CrossRef](#)]
65. Lu, D.; Popuri, K.; Ding, G.W.; Balachandar, R.; Beg, M.F.; Alzheimer's Disease Neuroimaging Initiative. Multiscale deep neural network based analysis of FDG-PET images for the early diagnosis of Alzheimer's disease. *Med. Image Anal.* **2018**, *46*, 26–34. [[CrossRef](#)] [[PubMed](#)]
66. Choi, H.; Jin, K.H. Predicting cognitive decline with deep learning of brain metabolism and amyloid imaging. *Behav. Brain Res.* **2018**, *344*, 103–109. [[CrossRef](#)] [[PubMed](#)]
67. Yang, Z.; Liu, Z. The risk prediction of Alzheimer's disease based on the deep learning model of brain 18F-FDG positron emission tomography. *Saudi. J. Biol. Sci.* **2020**, *27*, 659–665. [[CrossRef](#)] [[PubMed](#)]
68. Hon, M.; Khan, N.M. Towards Alzheimer's disease classification through transfer learning. In Proceedings of the 2017 IEEE International Conference On Bioinformatics And Biomedicine (Bibm), Kansas City, MO, USA, 13–16 November 2017; pp. 1166–1169. [[CrossRef](#)]
69. Li, F.; Liu, M. A hybrid Convolutional and Recurrent Neural Network for Hippocampus Analysis in Alzheimer's Disease. *J. Neurosci. Methods* **2019**, *323*, 108–118. [[CrossRef](#)]
70. Gao, F.; Yoon, H.; Xu, Y.; Goradia, D.; Luo, J.; Wu, T.; Su, Y. AD-NET: Age-adjust neural network for improved MCI to AD conversion prediction. *NeuroImage Clin.* **2020**, *27*, 102290. [[CrossRef](#)] [[PubMed](#)]
71. Frontiers. Deep Learning in Alzheimer's Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data | Frontiers in Aging Neuroscience'. Available online: <https://www.frontiersin.org/articles/10.3389/fnagi.2019.00220/full> (accessed on 11 September 2020).
72. Tanveer, M.; Richhariya, B.; Khan, R.U.; Rashid, A.H.; Khanna, P.; Prasad, M.; Lin, C.T. Machine Learning Techniques for the Diagnosis of Alzheimer's Disease: A Review. *ACM Trans. Multimed. Comput. Commun. Appl.* **2020**, *16*, 1–35. [[CrossRef](#)]
73. Hosseini-Asl, E.; Ghazal, M.; Mahmoud, A.; Aslantas, A.; Shalaby, A.M.; Casanova, M.F.; Barnes, G.N.; Gimel'farb, G.; Keynton, R.; El-Baz, A. Alzheimer's disease diagnostics by a 3D deeply supervised adaptable convolutional network. *Front. Biosci.* **2018**, *23*, 584–596. [[CrossRef](#)]
74. Korolev, S.; Safiullin, A.; Belyaev, M.; Dodonova, Y. Residual and plain convolutional neural networks for 3D brain MRI classification. In Proceedings of the 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), Melbourne, Australia, 18–21 April 2017; pp. 835–838. [[CrossRef](#)]
75. Rieke, J.; Eitel, F.; Weygandt, M.; Haynes, J.-D.; Ritter, K. Visualizing Convolutional Networks for MRI-based Diagnosis of Alzheimer's Disease. *arXiv* **2012**, arXiv:1808.02874.
76. Bengio, Y. Practical Recommendations for Gradient-Based Training of Deep Architectures. In *Neural Networks: Tricks of the Trade*, 2nd ed.; Montavon, G., Orr, G.B., Müller, K.-R., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 437–478. [[CrossRef](#)]
77. Liu, S.; Liu, S.; Cai, W.; Pujol, S.; Kikinis, R.; Feng, D. Early diagnosis of Alzheimer's disease with deep learning. In Proceedings of the 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI), Beijing, China, 29 April–2 May 2014; pp. 1015–1018. [[CrossRef](#)]
78. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *arXiv* **2015**, arXiv:1512.03385.
79. CaffeNet 1.2. Azure AI Gallery. Available online: <https://gallery.azure.ai/Model/CaffeNet-1-2-2> (accessed on 25 March 2021).
80. Wei, J. AlexNet: The Architecture that Challenged CNNs. Available online: <https://towardsdatascience.com/alexnet-the-architecture-that-challenged-cnns-e406d5297951> (accessed on 23 January 2021).
81. Huang, G.; Liu, Z.; Van der Maaten, L.; Weinberger, Q.K. Densely Connected Convolutional Networks. *arXiv* **2018**, arXiv:1608.06993.
82. VGG16—Convolutional Network for Classification and Detection. Available online: <https://neurohive.io/en/popular-networks/vgg16/> (accessed on 22 January 2021).
83. Alake, R. Deep Learning: GoogLeNet Explained. Available online: <https://towardsdatascience.com/deep-learning-googlenet-explained-de8861c82765> (accessed on 22 January 2021).
84. Tsang, S.-H. Review: Inception-v4—Evolved From GoogLeNet, Merged with ResNet Idea (Image Classification). Available online: <https://towardsdatascience.com/review-inception-v4-evolved-from-googlenet-merged-with-resnet-idea-image-classification-5e8c339d18bc> (accessed on 22 January 2021).

85. Caffe. Deep Learning Framework. Available online: <https://caffe.berkeleyvision.org/> (accessed on 22 January 2021).
86. Sherstinsky, A. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network. *Phys. D Nonlinear Phenom.* **2020**, *404*, 132306. [[CrossRef](#)]
87. Gaillard, F. MRI Sequences (Overview). Radiology Reference Article. Available online: <https://radiopaedia.org/articles/mri-sequences-overview> (accessed on 26 January 2021).
88. What Is 3T MRI. Medical Associates of Northwest Arkansas. Available online: <https://www.mana.md/what-is-3t-mri/> (accessed on 25 January 2021).
89. Wang, J.; He, L.; Zheng, H.; Lu, Z.-L. Optimizing the Magnetization-Prepared Rapid Gradient-Echo (MP-RAGE) Sequence. *PLoS ONE* **2014**, *9*, e96899. [[CrossRef](#)] [[PubMed](#)]
90. Slough, C.; Masters, S.C.; Hurley, R.A.; Taber, K.H. Clinical Positron Emission Tomography (PET) Neuroimaging: Advantages and Limitations as a Diagnostic Tool. *JNP* **2016**, *28*, 66–71. [[CrossRef](#)]
91. Chételat, G.; Arbizu, J.; Barthel, H.; Garibotto, V.; Law, I.; Morbelli, S.; Van de Giessen, E.; Agosta, F.; Barkhof, F.; Brooks, D.J.; et al. Amyloid-PET and 18F-FDG-PET in the diagnostic investigation of Alzheimer’s disease and other dementias. *Lancet Neurol.* **2020**, *19*, 951–962. [[CrossRef](#)]
92. Frisoni, G.B.; Boccardi, M.; Barkhof, F.; Blennow, K.; Cappa, S.; Chiotis, K.; Démonet, J.-F.; Garibotto, V.; Giannakopoulos, P.; Gietl, A.; et al. Strategic roadmap for an early diagnosis of Alzheimer’s disease based on biomarkers. *Lancet Neurol.* **2017**, *16*, 661–676. [[CrossRef](#)]
93. Laforce, R.; Soucy, J.-P.; Sellami, L.; Dallaire-Théroux, C.; Brunet, F.; Bergeron, D.; Miller, B.L.; Ossenkoppele, R. Molecular imaging in dementia: Past, present, and future. *Alzheimer’s Dement.* **2018**, *14*, 1522–1552. [[CrossRef](#)]
94. Saint-Aubert, L.; Lemoine, L.; Chiotis, K.; Leuzy, A.; Rodriguez-Vieitez, E.; Nordberg, A. Tau PET imaging: Present and future directions. *Mol. Neurodegener.* **2017**, *12*, 19. [[CrossRef](#)]
95. Werry, E.L.; Bright, F.M.; Pigué, O.; Ittner, L.M.; Halliday, G.M.; Hodges, J.R.; Kiernan, M.C.; Loy, C.T.; Kril, J.T.; Kassiou, M. Recent Developments in TSPO PET Imaging as A Biomarker of Neuroinflammation in Neurodegenerative Disorders. *Int. J. Mol. Sci.* **2019**, *20*, 13. [[CrossRef](#)] [[PubMed](#)]
96. Cai, Z.; Li, S.; Matuskey, D.; Nabulsi, N.; Huang, Y. PET imaging of synaptic density: A new tool for investigation of neuropsychiatric diseases. *Neurosci. Lett.* **2019**, *691*, 44–50. [[CrossRef](#)] [[PubMed](#)]
97. Cheng, D.; Liu, M. Classification of Alzheimer’s Disease by Cascaded Convolutional Neural Networks Using PET Images. In *Machine Learning in Medical Imaging*; Springer: Cham, Switzerland, 2017; pp. 106–113. [[CrossRef](#)]
98. Liu, M.; Cheng, D.; Wang, K.; Wang, Y.; The Alzheimer’s Disease Neuroimaging Initiative. Multi-Modality Cascaded Convolutional Neural Networks for Alzheimer’s Disease Diagnosis. *Neuroinform* **2018**, *16*, 295–308. [[CrossRef](#)] [[PubMed](#)]
99. Sun, X.; Shi, L.; Luo, Y.; Yang, W.; Li, H.; Liang, P.; Li, K.; Mok, V.C.T.; Chu, W.C.W.; Wang, D. Histogram-based normalization technique on human brain magnetic resonance images from different acquisitions. *BioMed. Eng. OnLine* **2015**, *14*, 73. [[CrossRef](#)] [[PubMed](#)]
100. Sled, J.G.; Zijdenbos, A.P.; Evans, A.C. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans. Med. Imaging* **1998**, *17*, 87–97. [[CrossRef](#)]
101. Smoothing—Introduction to MRI. Available online: http://jpeelle.net/mri/image_processing/smoothing.html (accessed on 4 February 2021).
102. Spatial Normalization—An Overview. ScienceDirect Topics. Available online: <https://www.sciencedirect.com/topics/medicine-and-dentistry/spatial-normalization> (accessed on 5 February 2021).
103. About the MNI Space(s)—Lead-DBS’. Available online: <https://www.lead-dbs.org/about-the-mni-spaces/> (accessed on 5 February 2021).
104. Ashburner, J. A fast diffeomorphic image registration algorithm. *Neuroimage* **2007**, *38*, 95–113. [[CrossRef](#)] [[PubMed](#)]
105. Avants, B.B.; Epstein, C.L.; Grossman, M.; Gee, J.C. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* **2008**, *12*, 26–41. [[CrossRef](#)] [[PubMed](#)]
106. Ashburner, J.; Friston, K.J. Unified segmentation. *NeuroImage* **2005**, *26*, 839–851. [[CrossRef](#)]
107. Beg, M.F.; Miller, M.I.; Trounev, A.; Younes, L. Computing Large Deformation Metric Mappings via Geodesic Flows of Diffeomorphisms. *Int. J. Comput. Vis.* **2005**, *61*, 139–157. [[CrossRef](#)]
108. Gaillard, F. Anterior Commissure—Posterior Commissure Line. Radiology Reference Article. Available online: <https://radiopaedia.org/articles/anterior-commissure-posterior-commissure-line-1> (accessed on 5 February 2021).
109. Köser, K. Affine Registration. In *Computer Vision: A Reference Guide*; Ikeuchi, K., Ed.; Springer: Boston, MA, USA, 2014; pp. 22–25. [[CrossRef](#)]
110. Interpolation Methods For Image Registration (Biomedical Image Analysis). Available online: <http://what-when-how.com/biomedical-image-analysis/interpolation-methods-for-image-registration-biomedical-image-analysis/> (accessed on 10 February 2021).
111. Jack, C.R.; Knopman, D.S.; Jagust, W.J.; Petersen, R.C.; Weiner, M.W.; Aisen, P.S.; Shaw, L.M.; Vemuri, P.; Wiste, H.J.; Weigand, S.D.; et al. Tracking pathophysiological processes in Alzheimer’s disease: An updated hypothetical model of dynamic biomarkers. *Lancet Neurol.* **2013**, *12*, 207–216. [[CrossRef](#)]
112. Juntu, J.; Sijbers, J.; Van Dyck, D.; Gielen, J. Bias Field Correction for MRI Images. In *Computer Recognition Systems*; Springer: Heidelberg/Berlin, Germany, 2005; pp. 543–551. [[CrossRef](#)]

113. Tustison, N.J.; Avants, B.B.; Cook, P.A.; Zheng, Y.; Egan, A.; Yushkevich, P.A.; Gee, J.C. N4ITK: Improved N3 Bias Correction. *IEEE Trans. Med. Imaging* **2010**, *29*, 1310–1320. [[CrossRef](#)] [[PubMed](#)]
114. Long, X.; Chen, L.; Jiang, C.; Zhang, L. Prediction and classification of Alzheimer disease based on quantification of MRI deformation. *PLoS ONE* **2017**, *12*, e0173372. [[CrossRef](#)] [[PubMed](#)]
115. Rolls, E.T.; Huang, C.-C.; Lin, C.-P.; Feng, J.; Joliot, M. Automated anatomical labelling atlas 3. *NeuroImage* **2020**, *206*, 116189. [[CrossRef](#)]
116. Kabani, N.J.; MacDonald, D.J.; Holmes, C.J.; Evans, A.C. 3D Anatomical Atlas of the Human Brain. *NeuroImage* **1998**, *7*, S717. [[CrossRef](#)]
117. Zeiler, M.D.; Fergus, R. Visualizing and Understanding Convolutional Networks. In *Computer Vision—ECCV 2014*; Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2014; pp. 818–833. [[CrossRef](#)]
118. Suk, H.-I.; Shen, D. Deep Learning-Based Feature Representation for AD/MCI Classification. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2013*; Springer: Heidelberg/Berlin, Germany, 2013; pp. 583–590. [[CrossRef](#)]
119. Raamana, P.R.; Weiner, M.W.; Wang, L.; Beg, M.F. Thickness network features for prognostic applications in dementia. *Neurobiol. Aging* **2015**, *36*, S91–S102. [[CrossRef](#)]
120. Li, F.; Liu, M. Alzheimer’s disease diagnosis based on multiple cluster dense convolutional networks. *Comput. Med Imaging Graph.* **2018**, *70*, 101–110. [[CrossRef](#)] [[PubMed](#)]
121. ADNI. Alzheimer’s Disease Neuroimaging Initiative. Available online: <http://adni.loni.usc.edu/> (accessed on 20 February 2021).
122. LaMontagne, P.J.; Keefe, S.; Lauren, W.; Xiong, C.; Grant, E.A.; Moulder, K.L.; Morris, J.C.; Benzinger, T.L.S.; Marcus, D.S. OASIS-3: Longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer’s disease. *Alzheimer’s Dement.* **2018**, *14*, 1097. [[CrossRef](#)]
123. AIBL Study ADNI Data. Available online: <https://aibl.csiro.au/adni/index.html> (accessed on 20 February 2021).
124. IXI Dataset—Brain Development’. Available online: <https://brain-development.org/ixi-dataset/> (accessed on 20 February 2021).
125. NITRC: MIRIAD (Minimal Interval Resonance Imaging in Alzheimer’s Disease): Tool/Resource Info’. Available online: <https://www.nitrc.org/projects/miriad/> (accessed on 20 February 2021).
126. E9 Statistical Principles for Clinical Trials, European Medicines Agency, Notes for Guidance on Statistical Principles for Clinical Trials (CPMP/ICH/363/96,2006). Available online: <https://www.ema.europa.eu/en/ich-e9-statistical-principles-clinical-trials> (accessed on 20 February 2021).
127. FSL—FslWiki. Available online: <https://fsl.fmrib.ox.ac.uk/fsl/fslwiki> (accessed on 22 February 2021).
128. SPM12 Software—Statistical Parametric Mapping. Available online: <https://www.fil.ion.ucl.ac.uk/spm/software/spm12/> (accessed on 22 February 2021).
129. Nipype Developers, Nipype: Neuroimaging in Python: Pipelines and Interfaces. Available online: <http://nipype.org/nipype> (accessed on 22 February 2021).
130. Community of practice devoted to the use of the Python programming language in the analysis of neuroimaging data. Available online: <http://nipype.org> (accessed on 22 February 2021).
131. AFNI (Analysis of Functional Neuro Images) Is a Leading Software Suite of C, Python, Rprograms and Shell Scripts Primarily Developed for the Analysis and Display of Multiple MRI Modalities. Available online: <https://afni.nimh.nih.gov/> (accessed on 2 March 2021).
132. ANTs by Stnava. Available online: <http://stnava.github.io/ANTs/> (accessed on 2 March 2021).
133. UCL Centre for Medical Image Computing—UCL—University College London’. Available online: <https://www.ucl.ac.uk/medical-image-computing/> (accessed on 2 March 2021).
134. FreeSurfer. Available online: <http://surfer.nmr.mgh.harvard.edu/> (accessed on 2 March 2021).
135. MNE—MNE 0.22.0 Documentation. Available online: <https://mne.tools/stable/index.html> (accessed on 2 March 2021).
136. 3D Slicer. Available online: <https://www.slicer.org/> (accessed on 2 March 2021).
137. SPM—Statistical Parametric Mapping. Available online: <https://www.fil.ion.ucl.ac.uk/spm/> (accessed on 2 March 2021).
138. Image Processing Toolbox. Available online: <https://www.mathworks.com/products/image.html> (accessed on 2 March 2021).
139. Tools for NIFTI and ANALYZE Image. Available online: <https://www.mathworks.com/matlabcentral/fileexchange/8797-tools-for-nifti-and-analyze-image> (accessed on 3 March 2021).
140. MRICro Software Guide. Available online: <https://people.cas.sc.edu/rorden/mricro/mricro.html#Introduction> (accessed on 3 March 2021).
141. R2015b—MATLAB & Simulink. Available online: <https://www.mathworks.com/help/matlab/release-notes-R2015b.html> (accessed on 2 March 2021).
142. Keras: The Python Deep Learning API. Available online: <https://keras.io/> (accessed on 2 March 2021).
143. TensorFlow. Available online: <https://www.tensorflow.org/> (accessed on 4 March 2021).
144. Laboratory Montreal University of Theano: Optimizing Compiler for Evaluating Mathematical Expressions on CPUs and GPUs. Available online: <http://deeplearning.net/software/theano/> (accessed on 24 February 2021).
145. PyTorch. Available online: <https://www.pytorch.org/> (accessed on 23 February 2021).
146. Home—MatConvNet. Available online: <https://www.vlfeat.org/matconvnet/> (accessed on 23 February 2021).
147. Deep Learning Toolbox. Available online: <https://www.mathworks.com/products/deep-learning.html> (accessed on 23 February 2021).

148. Martínez-Murcia, F.J.; Górriz, J.M.; Ramírez, J.; Illán, I.A.; Segovia, F.; Castillo-Barnes, D.; Salas-Gonzalez, D. Functional Brain Imaging Synthesis Based on Image Decomposition and Kernel Modeling: Application to Neurodegenerative Diseases. *Front. Neuroinform.* **2017**, *11*, 65. [[CrossRef](#)] [[PubMed](#)]
149. Sheikh, T.S.; Khan, A.; Fahim, M.; Ahmad, M. Synthesizing Data Using Variational Autoencoders for Handling Class Imbalanced Deep Learning. In *Analysis of Images, Social Networks and Texts*; Springer International Publishing: Cham, Switzerland, 2020; pp. 270–281. [[CrossRef](#)]
150. Yang, G.; Yu, S.; Dong, H.; Slabaugh, G.; Dragotti, P.L.; Ye, X.; Liu, F.; Arridge, S.; Keegan, J.; Guo, Y.; et al. Dagan: Deep De-Aliasing Generative Adversarial Networks for Fast Compressed Sensing MRI Reconstruction. *IEEE Trans. Med Imaging* **2018**, *37*, 1310–1321. [[CrossRef](#)]
151. Mardani, M.; Gong, E.; Cheng, J.Y.; Vasanawala, S.; Zaharchuk, G.; Alley, M.; Thakur, N.; Han, S.; Dally, W.; Pauly, J.M.; et al. Deep Generative Adversarial Networks for Compressed Sensing Automates MRI. *arXiv* **2017**, arXiv:1706.00051.
152. Shin, H.-C.; Roth, H.R.; Gao, M.; Lu, L.; Xu, Z.; Nogues, I.; Yao, J.; Mollura, D.; Summers, R.M. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Trans. Med. Imaging* **2016**, *35*, 1285–1298. [[CrossRef](#)]
153. Valliani, A.; Soni, A. Deep Residual Nets for Improved Alzheimer’s Diagnosis. In Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, New York, NY, USA, 20–23 August 2017; p. 615. [[CrossRef](#)]
154. Bäckström, K.; Nazari, M.; Gu, I.Y.; Jakola, A.S. An efficient 3D deep convolutional network for Alzheimer’s disease diagnosis using MR images. In Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018; pp. 149–153. [[CrossRef](#)]
155. Deep Learning. Available online: <https://www.deeplearningbook.org/> (accessed on 27 February 2021).
156. Brownlee, J. How to Visualize Filters and Feature Maps in Convolutional Neural Networks. In *Machine Learning Mastery*. 2019. Available online: <https://machinelearningmastery.com/how-to-visualize-filters-and-feature-maps-in-convolutional-neural-networks/> (accessed on 2 March 2021).