

## Article

# Diagnosis of Alzheimer's Disease Based on Accelerated Mirror Descent Optimization and a Three-Dimensional Aggregated Residual Network

Yue Tu, Shukuan Lin \*, Jianzhong Qiao, Peng Zhang and Kuankuan Hao

Department of Computer Science and Engineering, Northeastern University, Shenyang 110819, China

\* Correspondence: linshukuan@mail.neu.edu.cn

**Abstract:** Alzheimer's disease (AD), a neuropsychiatric disorder, continually arises in the elderly. To date, no targeted medications have been developed for AD. Early and fast diagnosis of AD plays a pivotal role in identifying potential AD patients, enabling timely medical interventions, and mitigating disease progression. Computer-aided diagnosis (CAD) becomes possible with the burgeoning of deep learning. However, the existing CAD models for processing 3D Alzheimer's disease images usually have the problems of slow convergence, disappearance of gradient, and falling into local optimum. This makes the training of 3D diagnosis models need a lot of time, and the accuracy is often poor. In this paper, a novel 3D aggregated residual network with accelerated mirror descent optimization is proposed for diagnosing AD. First, a novel unbiased subgradient accelerated mirror descent (SAMD) optimization algorithm is proposed to speed up diagnosis network training. By optimizing the nonlinear projection process, our proposed algorithm can avoid the occurrence of the local optimum in the non-Euclidean distance metric. The most notable aspect is that, to the best of our knowledge, this is the pioneering attempt to optimize the AD diagnosis training process by improving the optimization algorithm. Then, we provide a rigorous proof of the SAMD's convergence, and the convergence of SAMD is better than any existing gradient descent algorithms. Finally, we use our proposed SAMD algorithm to train our proposed 3D aggregated residual network architecture (ARCNN). We employed the ADNI dataset to train ARCNN diagnostic models separately for the AD vs. NC task and the sMCI vs. pMCI task, followed by testing to evaluate the disease diagnostic outcomes. The results reveal that the accuracy can be improved in diagnosing AD, and the training speed can be accelerated. Our proposed method achieves 95.4% accuracy in AD diagnosis and 79.9% accuracy in MCI diagnosis; the best results contrasted with several state-of-the-art diagnosis methods. In addition, our proposed SAMD algorithm can save about 19% of the convergence time on average in the AD diagnosis model compared with the gradient descent algorithms, which is very momentous in clinic.

**Keywords:** Alzheimer's disease; optimization algorithm; mirror descent; CNN; MRI



**Citation:** Tu, Y.; Lin, S.; Qiao, J.; Zhang, P.; Hao, K. Diagnosis of Alzheimer's Disease Based on Accelerated Mirror Descent Optimization and a Three-Dimensional Aggregated Residual Network. *Sensors* **2023**, *23*, 8708. <https://doi.org/10.3390/s23218708>

Academic Editor: Fabio Baselice

Received: 13 September 2023

Revised: 23 October 2023

Accepted: 24 October 2023

Published: 25 October 2023

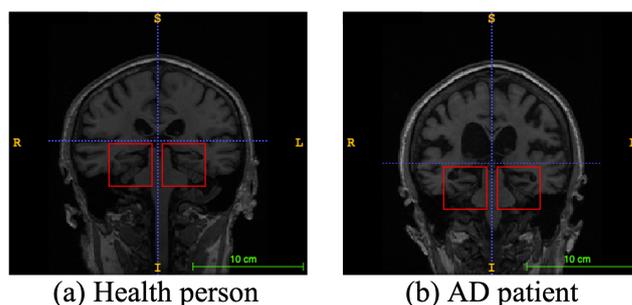


**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Alzheimer's disease (AD), a progressive neurodegenerative disease, is the most typical condition of dementia, mainly in the elderly. It affects people's memory function, thinking ability, and behavior capability. With the growth of modern life expectancy, the aging phenomenon is aggravating [1]. So, the prevalence of AD is further rising. According to the latest research, patients suffering from AD will reach 74.7 million by 2030 [2], with an average of one person who has AD every three seconds. At present, the etiology of AD is still unclear in medicine, and there is no effective therapeutic method to prevent the incidence and development of AD. Most of the new cases are in low- and middle-income countries. The substantial medical and nursing costs will cause a significant burden to society [3]. Patients with AD usually have organic brain changes, such as narrowing of

the gyrus and widening of the sulcus, and especially atrophy of brain tissue, such as the hippocampus and many lobes [4]. Early diagnosis of AD usually used neuroimaging data [5]. Many imaging studies [6] have demonstrated that Alzheimer's patients have different degrees of hippocampal volume atrophy, the corresponding lateral ventricle enlargement, and temporal angle enlargement phenomena, which gradually extend to the cingulate gyrus and neocortical junction with time. Figure 1 shows an example of different types of magnetic resonance images. In this figure, the left side is the hippocampus of a normal human brain, and the right side is the pathological hippocampus of an AD patient. We can see the atrophy of the hippocampus clearly. These abnormalities, easily obtained by craniocerebral MRI (magnetic resonance imaging), supply essential information for clinicians to diagnose AD.



**Figure 1.** An example of different types of magnetic resonance images. The left image is a healthy person, and the right one is an Alzheimer's disease patient. The red boxes represent hippocampi.

Moreover, MCI (mild cognitive impairment), a cognitive impairment state between the change in average random cognitive ability and dementia state, has a high probability of eventually becoming AD. MCI is split into two subtypes: sMCI (stable MCI) and pMCI (progress MCI). Patients with sMCI can prevent the development of the disease through drugs and medical intervention. However, patients with pMCI have a high probability of it turning into AD within three years [7]. If AD can be diagnosed early, clinicians can find the probable dementia of patients as soon as possible, carry out drug and medical intervention, and prevent the deterioration of the disease.

The rapid development of neural network technology makes computer-aided AD diagnosis possible. More and more researchers [8–10] think about how to use the MRI data of patients for AD diagnosis. The human brain has many functional areas, so the structure of MRI is very complex. When neural networks extract features from complex images, the optimization process is prone to the problems of vanishing gradient [11] and local optimization [12]. Moreover, the craniocerebral MRI data are three-dimensional matrix data. In the process of feature extraction of 3D MRI using neural network models, the training speed of the models is slow, and the accuracy of the diagnosis models is low.

When constructing the 3D MRI CAD model of AD, because data are highly complex, the neural network model often has slow speed in the feature extraction process. At the same time, due to the problems of vanishing gradient [11] and local optimal solutions [12] in the existing optimization algorithms when dealing with complex medical image data, the accuracy of prediction is often not satisfactory. Thus, accelerating the training speed, avoiding the disappearance of partial derivatives in the network training process, and improving accuracy have become challenges of diagnosing AD.

A novel 3D aggregated residual network with subgradient accelerated mirror descent optimization is proposed to tackle the aforementioned challenges. The main contributions are as follows:

- We propose SAMD, an unbiased subgradient accelerated mirror descent optimization algorithm, designed to expedite the training of diagnostic networks. By optimizing and improving the mirror descent algorithm [13], SAMD avoids the optimization problem falling into local optimum under the non-Euclidean distance metric. SAMD

achieves accelerated convergence by introducing the step factor and the deviation correction factor to the mirror descent algorithm. Furthermore, the incorporation of a subgradient unbiased estimation mechanism effectively mitigates the issue of vanishing gradient.

- We propose a 3D aggregated residual network (ARCNN) for feature extraction from craniocerebral MRI scans to diagnose AD. The ARCNN leverages our proposed aggregation residual blocks (ARBs) to capture global information within the input data, enhancing its capability to recognize complex patterns and features while improving model stability. Experimental results on the ADNI dataset highlight the superior performance of the diagnostic model trained using the SAMD algorithm in terms of accuracy and efficiency compared to other state-of-the-art methods.
- Our proposed SAMD algorithm achieves a convergence rate that is a square order faster than gradient descent algorithms such as Adam and SGD. Experimental results further validate the efficiency of the SAMD algorithm, showcasing an average time-saving of approximately 19% compared to gradient descent algorithms in the training phase. The SAMD algorithm accelerates AD diagnosis model training, overcoming real-world healthcare computational constraints, ensuring local data security, enabling continuous updates, and enhancing diagnosis efficiency while complying with privacy regulations and evolving medical knowledge.

## 2. Related Work

The vital technology of computer-aided AD diagnosis is to extract neuroimaging features from patients' MRIs. Generally, the existing AD diagnosis models are separated into four types of feature extraction techniques (i.e., voxel, patch, ROI, and whole-image).

Methods based on voxels [14] attempt to recognize AD-related classification microstructure. Usually, the feature dimension can reach the order of millions, and the number of images used for model training is relatively small (such as dozens). Those make the method face the risk of overfitting.

Comparatively, methods based on the region [15] obtained quantitative features by region segmentation algorithm and then constructed a classifier used to identify patients in a normal control group. Intuitively, these methods only attended to craniocerebral regions defined by experience. Therefore, they may not contain all the possible pathological parts in the cerebrum. Ju et al. [16] used fMRI data to diagnose AD. A brain was divided into 90 regions. According to the Pearson correlation coefficient, the connection of 90 regions is calculated to form a  $90 \times 90$  relation matrix, which is input to the network for diagnosis.

In order to obtain the changes of local brain regions, a patch-based method [17] used the intermediate scale of feature representation (between the voxel level and region level) to construct a classifier. However, a key question is determining the number, place, and proportions of patches in MR images. Therefore, Lian et al. [10] proposed a CNN (convolutional neural network) to find the voxels with high contrast through statistical analysis and comparison of ordinary images and AD images, and then extracted patches to improve the diagnosis effect.

Wang et al. [18] firstly input 3D MRI into DenseNet and then applied ensemble learning for diagnosis. Cui et al. [19] proposed a diagnostic framework based on CNN and RNN. This method can use the imaging data of different time points. Fang et al. [20] designed a neural network framework, taking advantage of 3D CNN and FSBI-LSTM. It was regarded as a feature extraction to extract the hidden spatial features to promote performance further. Ordinary diagnosis networks used convolutional layers to extract features (such as Resnet [21], Unet [22], etc.), followed by fully connected layers for diagnosis. Because the fully connected layer only processes the one-dimensional information, it resulted in information loss. Fang et al. [20] used bidirectional LSTM instead of fully connected layers to capture spatial information and improve the diagnosis effect. Pan et al. [23] used GAN to generate missing multimodal data for diagnosis. Precisely, the traditional multimodal diagnostic methods discard the subjects without PET (position emission computed

tomography) data to solve the problem of modal missing. However, this strategy may significantly reduce training data, thus diminishing the diagnostic performance.

Ji et al. [24] used a deep CNN to diagnose AD, and an effective method was used to train the model using an unbalanced small dataset. Recently, Zhang et al. [25] proposed a diagnostic model with embedding feature selection and feature fusion. They proposed a norm regularization-based optimization algorithm and theoretically demonstrated that it can converge to a global optimum. Their proposed method achieved 84% accuracy in diagnosing AD.

In the existing Alzheimer's diagnosis models, no matter which feature extraction methods are used for modeling, the optimization algorithms based on gradient descent are used for model training. These training algorithms have some problems, such as local optimization, vanishing gradient, and slow training speed. Therefore, we propose a novel unbiased subgradient accelerated mirror descent optimization algorithm and apply it to our 3D aggregated residual network diagnosis model. The best we can tell is that it is the first time to improve the optimization algorithm of the model in Alzheimer's diagnosis, which improves the training speed and diagnostic accuracy of the model.

### 3. Materials and Methods

#### 3.1. Datasets

The data used in our paper were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI). Parameters of these MR images are listed in Table 1. 348 subjects are separated into 4 categories, including AD, NC (normal control), sMCI, and pMCI. In total, the ADNI dataset contains 78 AD, 100 NC, 117 sMCI, and 53 pMCI subjects.

**Table 1.** Parameters of our acquired dataset.

Parameters	ADNI
Magnetic field intensity	1.5T
Field of view	160
Acquisition matrix	192 × 192
Thick of scans	1.2 mm
Manufacturer	SIMENS
TE (Echo time)	3.5–3.7 ms
TI (Inversion time)	1000 ms
TR (repetition Time)	3000 ms
Weight	T1

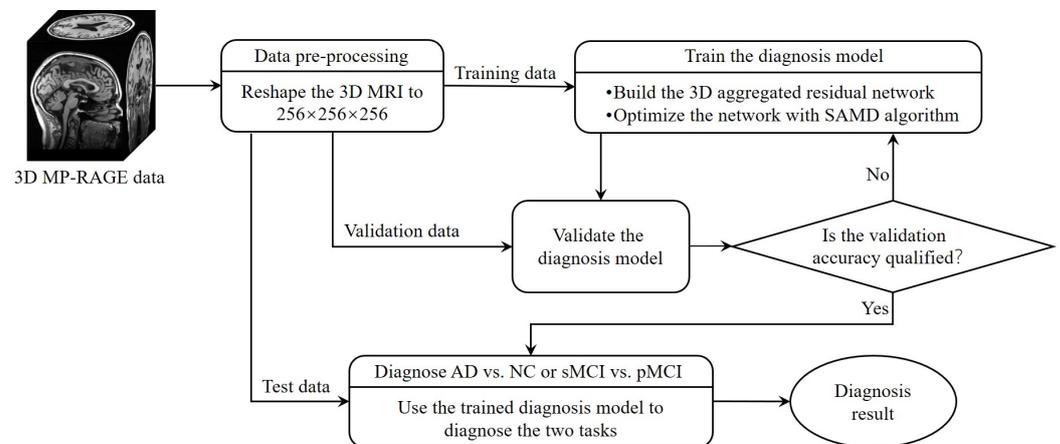
#### 3.2. Proposed Method

The whole framework of our proposed diagnostic model is shown in Figure 2. Before training the diagnosis network, data preprocessing (Section 3.2.1) was performed before training. Then, we proposed a 3D aggregated residual diagnosis network (Section 3.2.2) for feature extracting. Our proposed unbiased subgradient accelerated mirror descent optimization algorithm was used in the training process to obtain a faster training speed (Section 3.2.3). Our method addresses two binary classification tasks: diagnosing AD vs. NC and distinguishing sMCI from pMCI. We chose this approach due to clinical relevance; the AD vs. NC task aids early AD diagnosis, while sMCI vs. pMCI classifies cognitive impairment degree. It aligns with disease progression trends, avoiding inconsistencies in a combined model and ensuring meaningful diagnostic insights.

##### 3.2.1. Data Preprocessing

Our dataset comprises numerous subjects, each of which encompasses multiple MR images captured at various time points. We divide the data classified by subjects into MR images one by one and complete all the information of MR images at the same time. In this way, our dataset becomes image-based. The ADNI dataset contains 220 AD, 478 NC, 448 sMCI, and 328 pMCI. By interpolating the original data, every 3D MR image is re-

shaped into a  $256 \times 256 \times 256$  data cube, which is suitable as an input of our proposed 3D aggregated residual network.

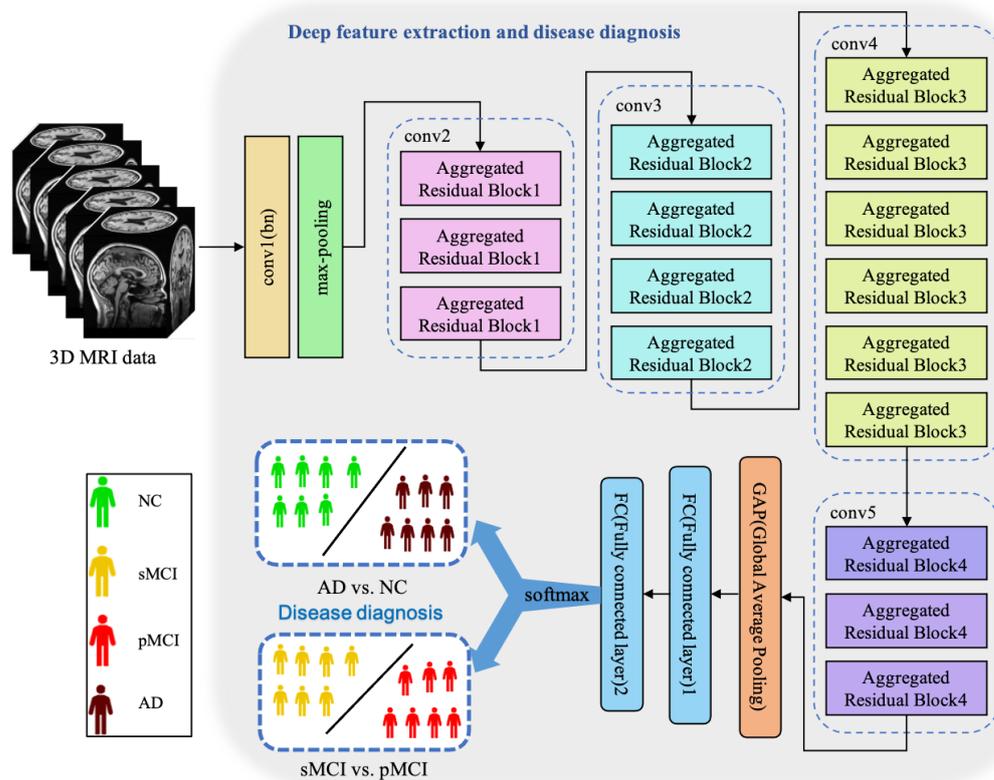


**Figure 2.** The framework of our proposed diagnostic model.

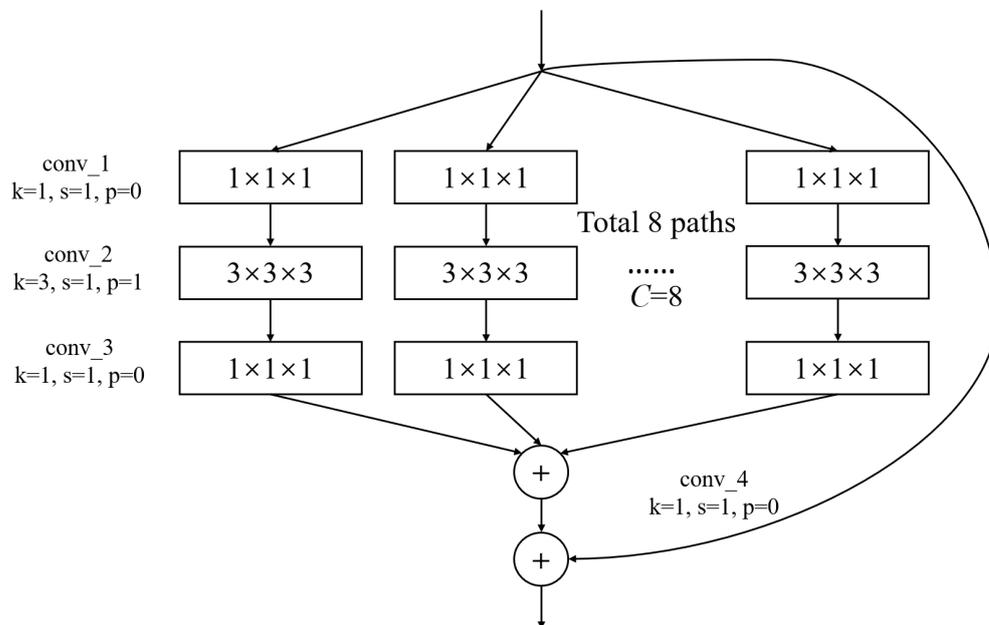
### 3.2.2. Three-Dimensional Aggregated Residual Diagnosis Network

Essentially, CNN is a feature extractor of layer-by-layer learning mode, which can mine different features from input images. Due to the specific structure of MRI data, specific neural network architectures are required to adequately capture spatial dependencies and features in healthy and ill brains. Because the 3D MRI data of AD patients have large amounts of information, they are often used to deepen or widen the network to enhance the accuracy of feature extraction. The design is conducive to extracting deeper features from MRI, to obtain higher accuracy. However, with the increase in hyperparameters (such as channels, filter size, etc.), the network training cost and design difficulty will greatly increase. We propose a 3D aggregated residual network architecture to extract features from MR images, shown in Figure 3.

In Figure 3, conv1(bn) denotes a convolutional layer with batch normalization applied. Conv2-conv5 are aggregated residual blocks (ARBs) with a unique cardinality mechanism to realize the sharing of parameters in the network model, shown in Figure 4. ARBs are designed to aggregate features from multiple preceding layers. This helps the network capture both low-level and high-level information effectively, enabling the model to learn intricate patterns and representations. Using ARB structures can increase the model's accuracy without significantly increasing the magnitude of parameters. The hyperparameter  $C$  represents the number of branches in the aggregated feature extraction convolutional layers. In accordance with the data characteristics and experimental conditions, we set  $C$  to 8. GAP stands for global average pooling, which calculates the average of feature maps, reducing them to scalar values. This reduces data dimensionality, leading to fewer parameters in subsequent fully connected layers and lower computational overhead. As GAP computes the average over all positions, it is spatially invariant, enhancing the model's ability to handle features at different positions. FC represents the fully connected layer. Table 2 displays details of the diagnostic network structure, including kernels, number of neurons, size of input, and output vectors at each layer. ARCNN is designed to extract neuroimaging features from MRIs, and the network can acquire subtle changes of AD patients' brain organs to give better diagnosis results.



**Figure 3.** The architecture of our proposed ARCNN. This model achieves two diagnosis tasks so that it needs to be trained respectively.



**Figure 4.** The architecture of ARB.  $k$  represents the convolutional kernel size, which is a small matrix used for feature extraction in convolutional operations.  $p$  represents the padding size, which is the technique of adding extra pixels around the input image before convolution to control the output size.  $s$  represents the stride, which determines the step size at which the kernel moves across the input data.

**Table 2.** The topology of our diagnosis network.

Layer	Stage	Input	Size	Output
1	conv1	$256 \times 256 \times 256 \times 1$	$7 \times 7 \times 7 \times 64$	$128 \times 128 \times 128 \times 64$
2	max pool	$128 \times 128 \times 128 \times 64$	—	$64 \times 64 \times 64 \times 64$
3	conv2	$64 \times 64 \times 64 \times 64$	$\begin{bmatrix} 1 \times 1 \times 1 \times 4 \\ 3 \times 3 \times 3 \times 4, C = 8 \\ 1 \times 1 \times 1 \times 256 \end{bmatrix} \times 3$	$64 \times 64 \times 64 \times 256$
4	conv3	$64 \times 64 \times 64 \times 256$	$\begin{bmatrix} 1 \times 1 \times 1 \times 8 \\ 3 \times 3 \times 3 \times 8, C = 8 \\ 1 \times 1 \times 1 \times 512 \end{bmatrix} \times 4$	$32 \times 32 \times 32 \times 512$
5	conv4	$32 \times 32 \times 32 \times 512$	$\begin{bmatrix} 1 \times 1 \times 1 \times 16 \\ 3 \times 3 \times 3 \times 16, C = 8 \\ 1 \times 1 \times 1 \times 1024 \end{bmatrix} \times 6$	$16 \times 16 \times 16 \times 1024$
6	conv5	$16 \times 16 \times 16 \times 1024$	$\begin{bmatrix} 1 \times 1 \times 1 \times 32 \\ 3 \times 3 \times 3 \times 32, C = 8 \\ 1 \times 1 \times 1 \times 2048 \end{bmatrix} \times 3$	$8 \times 8 \times 8 \times 2048$
7	GAP	$8 \times 8 \times 8 \times 2048$	—	$1 \times 1 \times 1 \times 2048$
8	FC1	$1 \times 2048$	—	$1 \times 512$
9	FC2	$1 \times 512$	—	$1 \times 2$

### 3.2.3. Accelerated Mirror Descent Optimization

Most diagnostic models of AD are usually 2D, and the extracted brain spatial structure information is insufficient. Therefore, researchers consider using a 3D model for feature extraction. Three-dimensional MRI data contain more features of brain organs, so the training speed of the 3D neural network model is slow. Moreover, training time is still a problem that confines the development of neural network models, and more efficient methods are always sought after. Aiming to accelerate the training speed of the 3D Alzheimer's disease diagnostic model, a novel accelerated mirror descent algorithm is proposed to decrease training time. This subsection will introduce the proposed unbiased subgradient accelerated mirror descent algorithm (SAMD). Before introducing our proposed algorithm, a table is listed to contain the symbolic representations and definitions used in this subsection, as shown in Table 3.

**Table 3.** A notation of symbols and definitions.

Name	Meanings
$B_m(a, b)$	Bregman divergence of $a$ and $b$
$m(\cdot)$	Bregman divergence generating function
$\theta_m$	Strong convex coefficient
$L$	Lipschitz constant
$F(x)$	Cost function
$\langle x, y \rangle$	Scalar product of $x, y$
$\nabla F(x)$	Gradient of $F(x)$
$\partial F(x)$	Subgradient of $F(x)$
$\alpha_t$	Step factor
$\beta_t$	Deviation correction factor
$\gamma_t$	Dynamically adjusted factor

### Preliminary

Mathematically, Bregman divergence [26] is defined as a method similar to distance measurement, which extends the square of Euclidean distance to a kind of distance. The

distance has no trigonometric inequality and symmetry. Bregman divergence  $B_m(a, b)$ , with regard to the function  $m(a)$ , is defined as follows:

$$B_m(a, b) = m(a) - m(b) - \langle \nabla m(b), a - b \rangle \quad (1)$$

where  $\langle a, b \rangle$  is the scalar product of  $a$  and  $b$ .  $m(\cdot)$  is a distance generation function.

Assume that  $m(\cdot)$  is a strong convex function. Then we have the following inequality [27]:

$$m(a) \geq m(b) + \langle \nabla m(b), a - b \rangle + \frac{1}{2}\theta_m \|a - b\|^2 \quad (2)$$

where  $\theta_m$  is a strong convex coefficient.

According to the knowledge of convex optimization, the Bregman divergence  $B_m(a, b)$  has the following property:

$$B_m(a, b) \geq \frac{1}{2}\theta_m \cdot \|a - b\|^2 \quad (3)$$

### SAMD Optimization Algorithm

The traditional gradient descent algorithm is slow when training data are complex. In the process of optimizing an objective function, the gradient descent algorithm is apt to converge to the local minimum [28]. Moreover, the gradient depends on derivatives, which is imprecise due to the nonsmoothness of the objective function. In addition, numerical schemes need to be used to solve them due to nonsmooth and unknown gradient flows. These gradient flows must have higher-order derivatives than the second-order derivatives. Meanwhile, implementing gradient flow is not easy because numerical artifacts will occur when the derivative is higher than second-order, but the gradient is usually like this.

The optimization objective of the mirror descent algorithm using Bregman divergence as the mirror map is [13]:

$$x_{t+1} = \operatorname{argmin}_{x \in \Omega} \{ \langle \nabla F(x_t), x - x_t \rangle + L \cdot B_m(x, x_t) \} \quad (4)$$

where  $L$  is a Lipschitz constant.  $x$  indicates the parameters to be trained and  $\Omega$  is the search space of  $x$ .

However, for the two items  $\langle \nabla F(x_t), x - x_t \rangle$  and  $B_m(x, x_t)$ , there is only one parameter  $L$  in the projection calculation. In the original mirror descent algorithm (see Equation (4)), constant  $L$  (called Lipschitz constant) is put forward to regulate the proportion of  $\langle \nabla F(x_t), x - x_t \rangle$  and  $B_m(x, x_t)$ .  $L$  is preset as a hyperparameter when the mirror descent iteration is performed. In the iterative change process, the algorithm cannot dynamically adjust the influence of  $\langle \nabla F(x_t), x - x_t \rangle$  and  $B_m(x, x_t)$  on the iterative results according to the iterative steps. This makes the mirror descent algorithm look a little inflexible. Moreover, the single parameter will make the algorithm oscillate in the iterative process, resulting in the poor effect of the optimization algorithm. So, we use the idea of momentum, which introduces the step factor  $\alpha_t$  and the deviation correction factor  $\beta_t$  to dynamically adjust the change speed in the iterative process of the algorithm so that the change of the process tends to be more stable. In this paper, the optimization objective is changed as:

$$x_{t+1} = \operatorname{argmin}_{x \in \Omega} \{ \alpha_t \cdot \langle h(x_t), x - x_t \rangle + \beta_t \cdot B_m(x, x_t) \} \quad (5)$$

Among them,  $\alpha_t$  is used to control the change of  $\langle \nabla F(x_t), x - x_t \rangle$ , and  $\beta_t$  is used to control the change of  $B_m(x, x_t)$ . According to the projection calculation process of the mirror descent algorithm, with iterations, the weights of  $\beta_t$  and  $\alpha_t$  change, which can accelerate the speed of the optimization algorithm to the optimal solution to a certain extent. In consideration of the above trends and related knowledge, we set  $\alpha_t = \frac{1}{t+1}$  and  $\beta_t = \frac{1}{1-\lambda^t}$ .  $\beta_t$  changes with the number of iterations, and the change rate is greater than  $\alpha_t$ .

The mirror descent algorithm is effective when  $\nabla F(x_t)$  can be calculated by a numerical calculation algorithm. However, when the objective function is nonsmooth or the derivative does not exist,  $\nabla F(x_t)$  can not be calculated. At this point, the gradient descent algorithms will not work. Therefore, we use the subgradient instead of traditional gradient in this paper. The subgradient has good performance in solving nondifferentiable convex optimization problems. The subgradients of function  $F(x)$  can be regarded as a cluster of supported hyperplanes, represented by  $\partial F(x)$ . Furthermore, the unbiased estimation of the subgradients  $E[\partial F(x)]$  is used to replace the gradient so that the results of the numerical analysis are unique, unbiased, and bounded. To prove this point, we give Theorem 1.

**Theorem 1.** *Given an objective function  $F(x)$ , for any  $x \in \Omega$ , there exists a real number  $N > 0$  subject to  $E[\partial F(x)] \leq N$ .*

**Proof.**  $F(x)$  is a finite function in a finite set. That is, it is uniformly bounded. Therefore, the unbiased estimation of subgradients is also bounded. The mathematical expectation of a bounded function is bounded. Obviously,  $\partial F(x)$  is a bounded function. So,  $\forall x \in \Omega$ , and there exists a real number  $N > 0$  subject to  $E[\partial F(x)] \leq N$ .  $\square$

Based on some knowledge of mathematical analysis, we can prove Theorem 1. Theorem 1 shows that when we calculate the unbiased estimate of the subgradient of  $F(x)$ , the result is bounded. Hence,  $E[\partial F(x)]$  can avoid the vanishing gradient problem.  $E[\partial F(x)]$  can be calculated by the following equation:

$$E[\partial F(x_t)] = \begin{cases} E(\nabla F(x_t)) & \text{if } F(x) \text{ is differentiable} \\ E\left(\frac{1}{n} \sum_{j=1}^n f_j(x_t)\right) & \text{if } F(x) \text{ is nondifferentiable} \end{cases} \quad (6)$$

where  $f_j(x_t)$  is the sampling of the subgradient hyperplane by  $F(x)$  at  $x = x_t$ . By the following equation, the gradient of  $F(x)$  can be replaced by  $h(x) = E[\partial F(x)]$ . This can avoid the vanishing gradient problem.

Finally, the iterative weighted average method is used to control the objective function change further so that the algorithm can achieve the optimal solution faster. We define  $\bar{x}_{t+1}$  as the weighted average value in the  $t + 1$  iterative step. It can be calculated by the following equation:

$$\bar{x}_{t+1} = \gamma_{t+1} \cdot \bar{x}_t + (1 - \gamma_{t+1}) \cdot x_{t+1} \quad (7)$$

where  $\bar{x}_t$  is the weighted average value in the previous step and  $x_{t+1}$  is calculated by Equation (5).  $\gamma_{t+1}$  is a dynamically adjusted parameter, which used to adjust the weight between  $\bar{x}_t$  and  $x_{t+1}$  in different iteration steps.  $\gamma_{t+1}$  is determined by accumulating the reciprocal of inner product term weight of the current step and the previous step. The accumulation of the reciprocal of inner product term weight at step  $t$  is represented by  $S_t$ , and the calculation method [29] is:

$$S_t = \sum_{i=0}^t \frac{1}{\alpha_i} \quad (8)$$

This setting can make the iteration step span of each step of the algorithm less large, and can achieve the optimal solution more accurately.

The approach is summarized in Algorithm 1.

Algorithm 1 shows our proposed stochastic unbiased subgradient accelerated mirror descent algorithm. Line 1 gives the initial of the whole algorithm. Lines 3–4 calculate  $\alpha_t$  and  $B_m(x, x_t)$ , which are put forward to dynamically regulate the impact of  $\langle \nabla F(x_t), x - x_t \rangle$  and  $B_m(x, x_t)$  on the iterative equation (see Line 6). Line 5 calculates our proposed unbiased subgradient method. Lines 7–8 show our proposed iterative weighted average method. By using this method, our proposed SAMD algorithm can converge faster and avoid the error caused by too large a step size. Lines 9–12 are used to evaluate whether the SAMD

algorithm fulfills the convergence situation. Through iterations, our proposed SAMD algorithm can train the diagnostic network until it reaches the optimal condition. Then, the convergence of our proposed SAMD will be analyzed below.

---

**Algorithm 1** A stochastic unbiased subgradient accelerated mirror descent algorithm

---

**Input:** Loss function  $F(x)$

**Output:** The optimization result  $x^*$

- 1: Initial:  $t = 0$ ,  $x_0 = \operatorname{argmin}_{x \in \Omega} m(x)$ , obtain  $\lambda$  by grid search
  - 2: **while**  $t \leq n$  **do**
  - 3:   Calculate  $\alpha_t = \frac{1}{t+1}$
  - 4:   Calculate  $\beta_t = \frac{1}{1-\lambda^t}$
  - 5:   Calculate  $h(x_t) = E[\partial F(x_t)]$
  - 6:   Set  $x_{t+1}$  by  $x_{t+1} = \operatorname{argmin}_{x \in \Omega} \{\alpha_t \cdot \langle h(x_t), x - x_t \rangle + \beta_t \cdot B_m(x, x_t)\}$
  - 7:   Calculate  $S_t = \sum_{i=0}^t \frac{1}{\alpha_i}$ ,  $S_{t+1} = \sum_{i=0}^{t+1} \frac{1}{\alpha_i}$  and  $\gamma_{t+1} = \frac{S_t}{S_{t+1}}$
  - 8:   Set  $\bar{x}_{t+1}$  by  $\bar{x}_{t+1} = \gamma_{t+1} \cdot \bar{x}_t + (1 - \gamma_{t+1}) \cdot x_{t+1}$
  - 9:   **if**  $\|\bar{x}_{t+1} - \bar{x}_t\| < \varepsilon$  **then**
  - 10:     Let  $x^* = \bar{x}_{t+1}$
  - 11:     **break**
  - 12:   **end if**
  - 13: **end while**
  - 14: return  $x^*$ .
- 

### Convergence of SAMD

The following lemma can be obtained according to the properties of strongly convex functions and Bregman divergence constructed by Bregman function.

**Lemma 1.**  $\forall p, q, r \in \Omega$ , we have:

$$B_m(p, r) - B_m(q, r) = B_m(p, q) + \langle \nabla m(q) - \nabla m(r), p - q \rangle \quad (9)$$

where constrained set  $\Omega$  is a closed convex set.

**Proof.** According to the definition of Bregman divergence (see in Equation (1)), the following three equations can be deduced:

$$B_m(p, q) = m(p) - m(q) - \langle \nabla m(q), p - q \rangle \quad (10)$$

$$B_m(p, r) = m(p) - m(r) - \langle \nabla m(r), p - r \rangle \quad (11)$$

$$B_m(q, r) = m(q) - m(r) - \langle \nabla m(r), q - r \rangle \quad (12)$$

Then, we have the following process by Equation (11) minus Equation (12):

$$\begin{aligned} B_m(p, r) - B_m(q, r) &= m(p) - m(q) - \langle \nabla m(r), p \rangle + \langle \nabla m(r), q \rangle \\ &= m(p) - m(q) - \langle \nabla m(r), p - q \rangle \end{aligned} \quad (13)$$

By Equation (13) minus Equation (10), we then have

$$\begin{aligned} B_m(p, r) - B_m(q, r) - B_m(p, q) &= \langle \nabla m(q), p - q \rangle - \langle \nabla m(r), p - q \rangle \\ &= \langle \nabla m(q) - \nabla m(r), p - q \rangle. \end{aligned} \quad (14)$$

So, we have  $B_m(p, r) - B_m(q, r) = B_m(p, q) + \langle \nabla m(q) - \nabla m(r), p - q \rangle$ .  $\square$

This lemma is proved by the property of Bregman divergence and is used to obtain the convergence of the SAMD algorithm. Based on some knowledge of mathematical analysis, we can obtain Lemma 2.

To prove the convergence of the unbiased subgradient accelerated mirror descent algorithm, the following lemma can be deduced:

**Lemma 2.** When  $t \geq 0$ , for any  $q \in \Omega$ , we have the following inequality:

$$\alpha_t \langle h(p_t), p_t - q \rangle \leq \beta_t [B_m(p_t, q) - E[B_m(p_{t+1}, q) | \delta_t]] + \frac{(\alpha_t N)^2}{2\theta_m} \tag{15}$$

where  $\delta_t$  is a sigma-field, i.e.,  $\delta_t = \{h(p_0), h(p_1), \dots, h(p_t)\}$ ,  $h(p) = E[\partial F(p)]$ . Especially,  $\delta_0 = \delta\{p_0\}$ .

**Proof.** From the first-order necessary condition, we can obtain

$$\langle \alpha_t h(p_t) + \beta_t (\nabla m(p_{t+1}) - \nabla m(p_t)), q - p_{t+1} \rangle \geq 0 \tag{16}$$

Then, it can be acquired that

$$\alpha_t \langle h(p_t), p_{t+1} - q \rangle \leq \beta_t \langle \nabla m(p_{t+1}) - \nabla m(p_t), q - p_{t+1} \rangle \tag{17}$$

From Lemma 1, it can be concluded that

$$\langle \nabla m(p_{t+1}) - \nabla m(p_t), q - p_{t+1} \rangle = B_m(p_t, q) - B_m(p_{t+1}, q) - B_m(p_t, q_{t+1}) \tag{18}$$

Take Equation (18) into Equation (17), and we can obtain

$$\alpha_t \langle h(p_t), p_{t+1} - q \rangle \leq \beta_t [B_m(p_t, q) - B_m(p_{t+1}, q) - B_m(p_t, p_{t+1})] \tag{19}$$

Bregman divergence has the following property (see in Equations (2) and (3)):

$$B_m(p_t, p_{t+1}) \geq \frac{1}{2} \theta_m \cdot \|p_t - p_{t+1}\|^2 \tag{20}$$

Then, Equation (19) can be transformed to

$$\alpha_t \langle h(p_t), p_{t+1} - q \rangle \leq \beta_t \left[ B_m(p_t, q) - B_m(p_{t+1}, q) - \frac{1}{2} \theta_m \|p_t - p_{t+1}\|^2 \right] \tag{21}$$

$$\begin{aligned} \alpha_t \langle h(p_t), p_{t+1} - q \rangle &\geq \alpha_t (\langle h(p_t), p_{t+1} - p_t \rangle + \langle h(p_t), p_t - q \rangle) \\ &\geq - \left| \left\langle \frac{\alpha_t \cdot h(p_t)}{\sqrt{\theta_m}}, \sqrt{\theta_m} (p_{t+1} - p_t) \right\rangle \right| \end{aligned} \tag{22}$$

The above inequalities are sorted out:

$$\alpha_t \langle h(p_t), p_{t+1} - q \rangle \geq \alpha_t \langle h(p_t), p_t - q \rangle - \left| \left\langle \frac{\alpha_t h(p_t)}{\sqrt{\theta_m}}, \sqrt{\theta_m} (p_{t+1} - p_t) \right\rangle \right| \tag{23}$$

By using the Fenchel inequality, Equation (23) is transformed to

$$\alpha_t \langle h(p_t), p_{t+1} - q \rangle \geq \alpha_t \langle h(p_t), p_t - q \rangle - \frac{1}{2} \left( \theta_m \|p_{t+1} - p_t\|^2 + \frac{\alpha_t^2}{\theta_m} \|h(p_t)\|_*^2 \right) \tag{24}$$

The above equation is transformed:

$$\alpha_t \langle h(p_t), p_t - q \rangle \leq \alpha_t \langle h(p_t), p_{t+1} - q \rangle + \frac{\alpha_t^2}{2\theta_m} \|h(p_t)\|_*^2 + \frac{\theta_m}{2} \|p_{t+1} - p_t\|^2 \tag{25}$$

Taking Equation (21) into Equation (25), we have

$$\alpha_t < h(p_t), p_t - q > \leq \beta_t \left[ B_m(p_t, q) - B_m(p_{t+1}, q) - \frac{\theta_m}{2} \|p_{t+1} - p_t\|^2 \right] + \frac{\alpha_t^2}{2\theta_m} \|h(p_t)\|_*^2 + \frac{1}{2} \|p_{t+1} - p_t\|^2 \quad (26)$$

According to Algorithm 1, there has  $\beta_t = \frac{1}{1-\lambda^t}$  ( $\lambda \in (0, 1)$ ), so the value range of  $\beta_t$  is  $\beta_t \in (1, +\infty)$ . Then, we take conditional expectations on  $\delta_t$ :

$$\alpha_t < h(p_t), p_t - q > \leq \beta_t [B_m(p_t, q) - E[B_m(p_{t+1}, q)|\delta_t]] + \frac{(\alpha_t N)^2}{2\theta_m} \quad (27)$$

□

From Lemma 2, we can deduce the following theorem to converge our proposed unbiased subgradient accelerated mirror descent algorithm.

**Theorem 2.** For an optimization question  $\min_{p \in \Omega} F(p)$ , where  $F(p)$  is a strong convex function,  $\bar{p}_t$  is the optimal solution obtained by the iterative algorithm, and  $p^*$  is the optimal solution, we have

$$F(\bar{p}_t) - F(p^*) \leq \frac{N^2}{2\theta_m(t+1)} \quad (28)$$

**Proof.** In Lemma 2, let  $q = p^*$ , and we can obtain

$$\alpha_t < h(p_t), p_t - p^* > + \beta_t E[B_m(p_{t+1}, p^*)|\delta_t] \leq \beta_t B_m(p_t, p^*) + \frac{\alpha_t^2 N^2}{2\theta_m} \quad (29)$$

By the strong convexity of  $F(p)$ , we can obtain

$$\alpha_t < h(p_t), p_t - p^* > \geq \alpha_t [F(p_t) - F(p^*)] + \frac{\alpha_t \theta_F}{2} \|p_t - p^*\|^2 \geq \alpha_t [F(p_t) - F(p^*)] \quad (30)$$

Take Equation (30) into Equation (29). Divide both sides of the inequality by  $\alpha_t$  and calculate the total expectation:

$$\frac{\beta_t}{\alpha_t} E[B_m(p_{t+1}, p^*)] + E[F(p_t) - F(p^*)] \leq \frac{\beta_t}{\alpha_t} E[B_m(p_t, p^*)] + \frac{\alpha_t N^2}{2\theta_m} \quad (31)$$

Then, we have

$$E[F(p_t)] - F(p^*) \leq \frac{N^2}{2\theta_m(t+1)} + \frac{\beta_t}{\alpha_t} [B_m(p_t, p^*) - B_m(p_{t+1}, p^*)] \quad (32)$$

When the algorithm iterates to  $\bar{p}_t$ , we can come to the following conclusion:

$$F(\bar{p}_t) - F(p^*) \leq \frac{N^2}{2\theta_m(t+1)} \quad (33)$$

□

According to Theorem 2, our proposed SAMD algorithm can converge at  $O(1/t)$ , a square order faster than the gradient descent algorithms (such as Adam and SGD (stochastic gradient descent), which have  $O(1/\sqrt{t})$  convergence). It is proved theoretically that the SAMD algorithm proposed in our paper can accelerate the convergence of the training process of the diagnostic network. At the same time, through the relevant description and theorem-proof of the algorithm, it can be obtained that the SAMD algorithm can avoid the problems of vanishing gradient and local optimal solution.

## 4. Experiments and Discussion

We validate our proposed model on the ADNI dataset. In Section 4.1, we describe some settings and performance measures. Sections 4.2–4.5 verify the performance of our proposed model.

### 4.1. Experimental Settings

Our proposed model is verified on both AD vs. NC task and sMCI vs. pMCI task. The hardware we utilized consisted of an Intel(R) Core(R) i9-9900K CPU, operating at 3.60 GHz with 16 cores, 16 GB of RAM, a storage setup comprising a 256 GB SSD and a 2 TB SATA drive, along with an NVIDIA RTX 2080 Ti GPU equipped with 11 GB of dedicated memory. This configuration provided the essential computational resources for our experiments.

To assess the performance of our proposed model, we divided the dataset into training, validation, and testing sets. For each class of data, we employed a split ratio of 7:2:1 for these three sets. The training set, comprising 70% of the data, was used for model training, during which the model learned to diagnose diseases by recognizing the features and labels of the data. The validation set, consisting of 20% of the data, played a crucial role in monitoring the model's performance during training and tuning hyperparameters, aiding in preventing overfitting. The remaining 10% of the data served as the testing set, used to evaluate the model's ultimate performance. Importantly, the testing set contained data that the model had not previously encountered, ensuring the objectivity of the evaluation. Furthermore, to enhance the evaluation of the model's performance and generalization capabilities, we ensured that images of the same subject do not appear in more than one of the training, validation, and test sets during the dataset partitioning process.

More specifically, four metrics, namely accuracy (ACC), sensitivity (SEN), specificity (SPE), and area under the curve (AUC, obtained by summing the area under the ROC curve), are used to evaluate our proposed model's diagnostic performance, which are the most commonly occurring in bioinformatics literature [30,31].

### 4.2. Compared with Other Optimizers

To evaluate the performance of the SAMD optimizer, we train our model and then test on ADNI in the two tasks of diagnosis. The results of SAMD are compared with other mainstream optimization methods by using the same diagnostic networks. Hence, Adam and SGD are chosen as two baselines. Adam, a widely used method in training neural networks, has proved to be better than many gradient methods. SGD is a well-known convex optimization method. Table 4 presents the diagnostic performance of our algorithm in two diagnostic tasks, as compared to other methods, on the testing datasets.

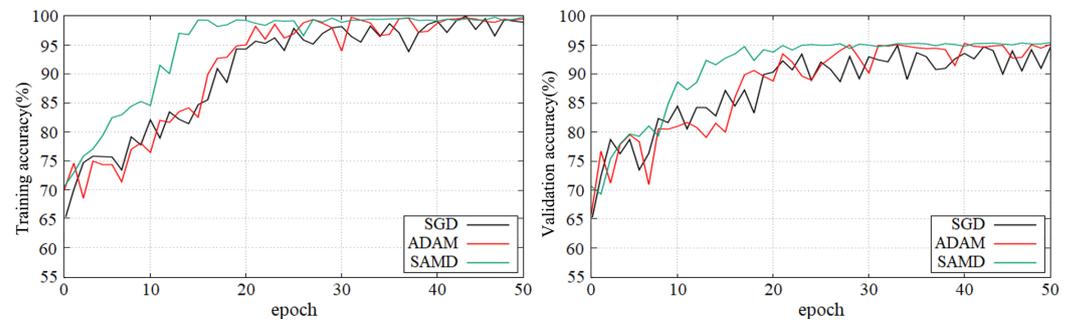
**Table 4.** Diagnostic performance of two tasks in ADNI dataset.

Methods	AD vs. NC				sMCI vs. pMCI			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
SGD	0.947	<b>0.972</b>	0.876	0.969	0.779	0.733	0.762	0.787
ADAM	0.951	0.960	0.929	0.967	0.788	0.526	<b>0.854</b>	0.781
<b>SAMD (ours)</b>	<b>0.954</b>	0.901	<b>0.952</b>	<b>0.970</b>	<b>0.799</b>	<b>0.737</b>	0.833	<b>0.789</b>

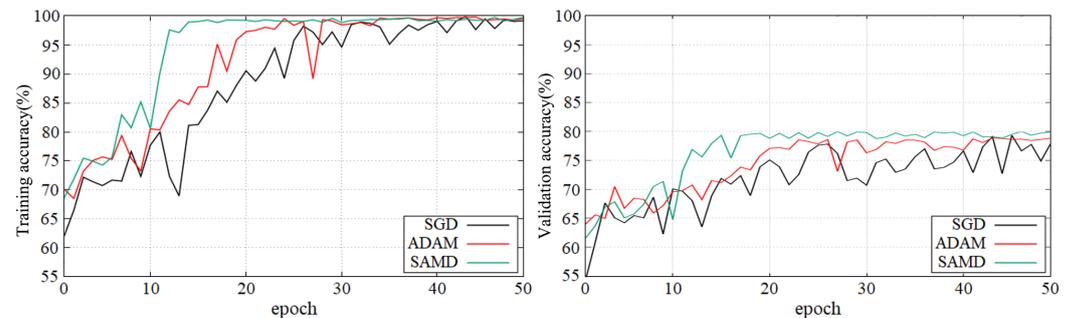
Note: The bolded values in each column represent the best-performing values for that specific column.

Table 4 shows that the proposed SAMD method has achieved excellent performance in most evaluation indicators. Primarily, our model can diagnose AD with 95.4% accuracy. This better performance is due to the proposed SAMD method being an optimizer based on mirror descent. The proposed method avoids the local optimization and the smoothness limitation of the cost function by constructing the optimization problem in the dual space. Therefore, it can reach the optimal value of the model more accurately. These experimental results also indicate that this model is still effective when the MRI to be diagnosed has different parameters (such as TE, TR) or the distribution of patients is inhomogeneous.

Figures 5 and 6 show the training and validation accuracy of different epochs in two diagnostic tasks. We can obtain that all models are trained to the optimal state from these figures. It is worth noting that the network architecture which used our proposed SAMD optimizer can achieve the optimal state in fewer epochs. That is, our proposed method has a faster convergence speed.



**Figure 5.** Training and validation accuracy with epochs in AD vs. NC diagnostic task by using ADNI dataset.



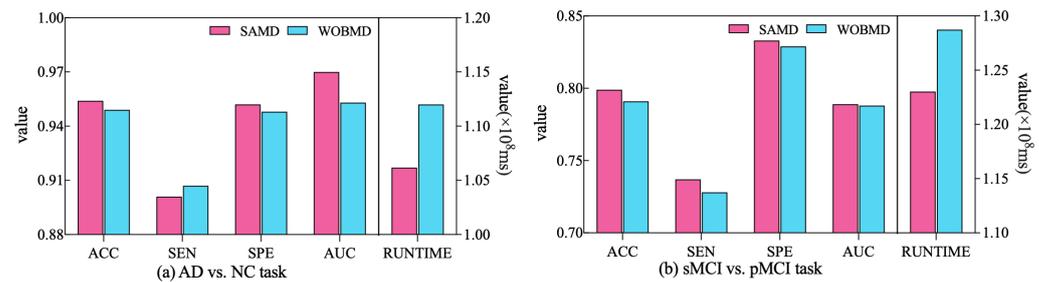
**Figure 6.** Training and validation accuracy with epochs in sMCI vs. pMCI diagnostic task by using ADNI dataset.

As shown in these two plots, our proposed SAMD method can achieve higher accuracy in a shorter time. Our proposed method, SAMD, can converge faster than other algorithms. Moreover, the model trained by SAMD can achieve higher accuracy eventually. According to the experimental results, we can believe that our proposed SAMD algorithm is effective and efficient in optimizing our proposed AD diagnostic network.

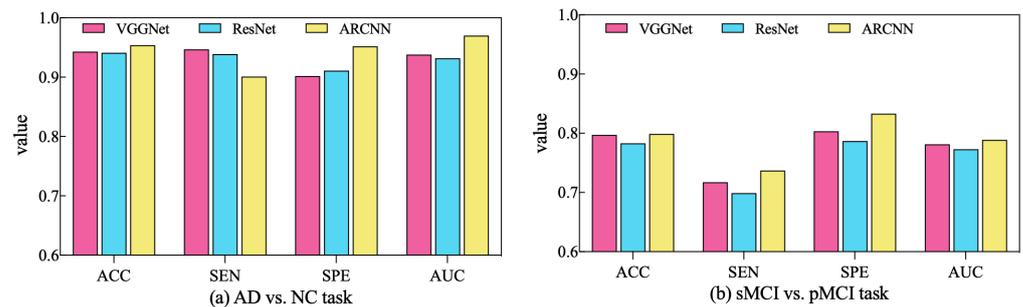
In order to better demonstrate the efficacy of the proposed adjustment factors in our study, we conducted a comparative experiment. Specifically, we established a control group (referred to as WOBMD) where we set  $\beta_t$  in SAMD to 1 (i.e., without considering  $\beta_t$ ). We applied WOBMD to train and optimize our diagnostic model ARCNN, and subsequently compared the diagnostic outcomes and runtime on both tasks with the results obtained using SAMD, as illustrated in Figure 7. The experimental results reveal that our SAMD algorithm, featuring two adjustment factors, outperforms the WOBMD algorithm, which utilizes only a single adjustment factor. This observation underscores the superiority and innovativeness of our proposed algorithm.

#### 4.3. Performance of Diagnostic Network

In order to assess the performance of the SAMD algorithm when optimizing different diagnostic networks, a series of comparative experiments were conducted. Specifically, two widely employed CNN-based networks in the AD diagnosis field, VGGNet [32] and ResNet [33], were selected for evaluation. SAMD was employed as the optimizer for training and optimizing these two networks on two different diagnostic tasks. The resulting diagnostic outcomes are presented in Figure 8.



**Figure 7.** Comparative performance of SAMD and WOBMD.



**Figure 8.** Comparison of diagnostic performance across different diagnostic networks using SAMD as the optimizer in two diagnostic tasks.

As illustrated in Figure 8, SAMD consistently demonstrated good performance across three diagnostic models (including VGGNet, ResNet, and our proposed ARCNN), with particularly superior results observed in the ARCNN model. This observation underscores the adaptability and effectiveness of the SAMD algorithm in optimizing AD diagnostic networks. Furthermore, the experimental results revealed that, in comparison to the two commonly used CNN network architectures, our proposed ARCNN architecture consistently yielded superior diagnostic performance. This demonstrates the advanced and superior characteristics of our model. The experiments confirm the efficacy of the SAMD algorithm in optimizing different diagnostic networks, highlighting its adaptability and effectiveness in the field of medical image analysis, especially for AD diagnosis. Additionally, it is evident that our proposed ARCNN model, with its unique ARB structure, outperforms both ResNet with residual blocks and VGGNet without any special network structure in both diagnostic tasks. This indicates that our proposed ARCNN diagnostic network, featuring ARBs, exhibits advanced performance.

To further validate the diagnostic performance of our ARCNN model optimized using SAMD, we gathered an additional dataset (represented as MRND), including 67 AD, 89 sMCI, 48 pMCI, and 75 NC. Using the model previously trained on ADNI data, we conducted a performance evaluation with this new MRI dataset, and the results are depicted in Figure 9.

As observed in the figure, our model continues to exhibit superior performance on the new MRI data, indicating the robustness and generalizability of our proposed approach. This outcome underscores the versatility of our method in effectively diagnosing AD across different datasets and highlights its potential for broader clinical applications. The results obtained from the evaluation on the new dataset reaffirm the effectiveness of our SAMD-optimized ARCNN model and its potential for widespread application in AD diagnosis.

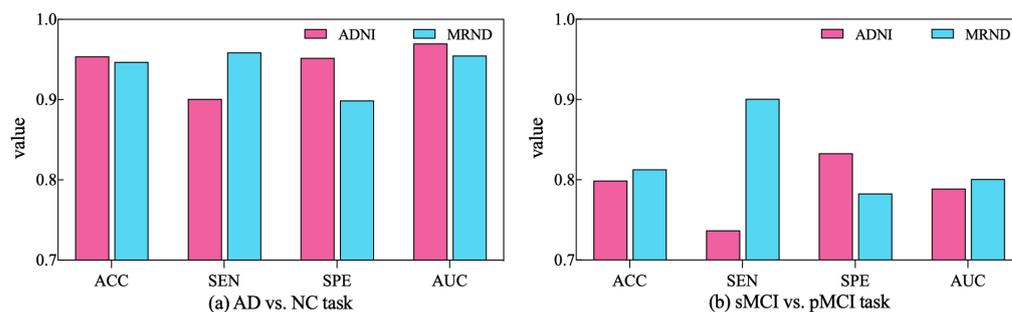


Figure 9. The performance of ARCNN trained with SAMD on MRND.

4.4. Performance of Training Time

The SAMD algorithm is proposed to accelerate the training speed of neural networks. The training time of two diagnosis tasks is recorded to verify this point. We applied SAMD, SGD, and Adam optimization to the ARCNN diagnostic model separately and recorded the training times for both diagnostic tasks, as shown in Figure 10. To ensure experimental fairness, we maintained identical datasets and experimental settings across the three different optimization algorithms for ARCNN. In a holistic perspective, the ARCNN model optimized using our proposed SAMD algorithm exhibited an average reduction of approximately 19% in training time compared to models optimized using SGD and Adam when training was completed. The experimental results demonstrate that our proposed SAMD optimization algorithm is capable of accelerating the training process of the ARCNN diagnostic network in different diagnostic tasks.

Furthermore, we present the average inference times for two diagnostic tasks conducted using our proposed diagnostic network. For the AD vs. NC task, the average inference time is 79.5216 ms, while for the sMCI vs. pMCI task, it is 81.3423 ms. It is worth noting that for a well-trained diagnostic model, the process of disease diagnosis is typically instantaneous, requiring minimal waiting time. Consequently, when deploying this system in clinical settings, real-time disease diagnosis can be achieved, facilitating prompt medical interventions and decisions.

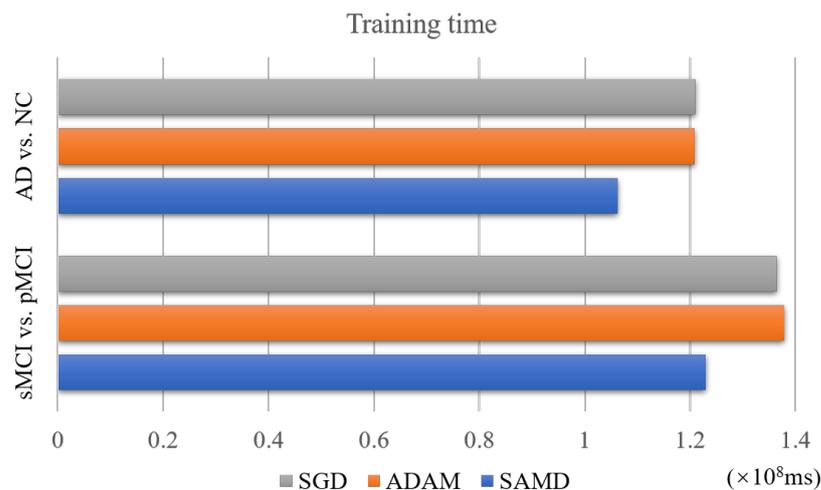


Figure 10. Training time of different optimization algorithms on various tasks.

4.5. Comparison with Previous Work

In Table 5, several SOTA results reported in the literature for AD vs. NC diagnosis and sMCI vs. pMCI diagnosis using structure MRI data of ADNI have been summarized. The underlined values are the optimal values in the baseline methods. From this table, we can see that our proposed method has a more powerful diagnosis capability than the

existing machine learning-based AD diagnostic methods. It is worth noting that although many baselines use more data for training than us, our proposed model still achieves the highest accuracy in the two classification tasks. These results demonstrate that our model effectively diagnoses different parametric nuclear magnetic scanings. The 3D aggregated residual diagnosis network has excellent feature extraction ability. The SAMD algorithm can train the neural network model accurately and efficiently, so our proposed architecture can obtain great diagnosis results.

**Table 5.** Comparison with previous work.

Reference	Subject	AD vs. NC				sMCI vs. pMCI			
		ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
Janousova et al. [34]	231 NC + 63 sMCI + 168 pMCI +198 AD	0.880	0.850	0.910	-	0.700	0.640	0.750	-
Liu et al. [35]	77 NC + 85 AD	0.914	0.923	0.904	-	-	-	-	-
Korolev et al. [36]	61 NC + 77 sMCI + 43 pMCI + 50 AD	0.800	-	-	0.870	0.520	-	-	0.520
Karasawa et al. [37]	574 NC + 346 AD	<u>0.940</u>	-	-	-	-	-	-	-
Khvostikov et al. [38]	58 NC + 48 AD	0.854	0.883	0.900	-	-	-	-	-
Lin et al. [39]	229 NC + 188 AD	0.799	0.840	0.748	0.861	-	-	-	-
Xu et al. [30]	165 NC + 95 sMCI +126 pMCI + 142 AD	0.904	<b>0.924</b>	0.887	0.954	0.637	<b>0.786</b>	0.454	0.679
Cui et al. [31]	223NC + 231 sMCI + 165 pMCI + 192 AD	0.923	0.906	0.937	<u>0.970</u>	0.750	0.733	0.762	0.777
Zhu et al. [40]	419 NC + 345 AD	0.916	0.874	<u>0.948</u>	0.958	-	-	-	-
Poloni et al. [8]	302 NC + 209 AD	0.826	-	-	0.900	-	-	-	-
Alinsaif et al. [9]	50 sMCI +50 pMCI	-	-	-	-	0.700	0.600	<u>0.800</u>	-
Lin et al. [41]	308 NC + 233 sMCI +183 pMCI + 362 AD	0.923	0.904	0.944	0.928	0.741	0.750	0.731	0.766
Gao et al. [42]	427 NC + 342 sMCI +234 pMCI + 352 AD	0.920	0.891	0.940	0.956	<u>0.753</u>	0.773	0.741	<u>0.786</u>
Li et al. [43]	153 NC + 151 AD	0.930	<b>0.924</b>	<u>0.948</u>	-	-	-	-	-
<b>proposed</b>	100 NC + 117 sMCI + 53 pMCI + 78 AD	<b>0.954</b>	0.901	<b>0.952</b>	<b>0.970</b>	<b>0.799</b>	0.737	<b>0.833</b>	<b>0.789</b>

Note: The bolded values in each column represent the best-performing values for that specific column. The underlined values indicate the best-performing values in the current column among the baselines.

#### 4.6. Future Research Directions

In this subsection, we discuss future research directions for AD diagnosis models. Emerging avenues encompass the integration of cutting-edge imaging modalities, such as the fusion of fMRI and PET with MRI images, to bolster early detection capabilities. Furthermore, the development of interpretable AI models is poised to enhance diagnostic transparency, fostering trust within clinical contexts. Additionally, the adoption of early detection strategies, including biomarker discovery and digital health technologies, holds the potential to identify AD in its incipient stages.

The future of AD diagnosis hinges on the amalgamation of advanced imaging, the advancement of interpretable AI models, the realization of early detection, and the prediction of disease progression. Through continuous research endeavors, achieving more efficient and accurate AD diagnosis becomes attainable, enabling timely clinical interventions and treatments, ultimately alleviating societal burdens and enhancing the overall wellbeing of patients.

## 5. Conclusions

Computer-aided diagnosis of AD plays a vital role in helping doctors diagnose disease and find potential patients. We present a novel 3D aggregated residual network with accelerated mirror descent optimization for diagnosing AD. An unbiased subgradient accelerated mirror descent optimization algorithm is proposed to accelerate the training speed of our diagnostic model. The best we can tell is that it is the first work to accelerate the Alzheimer's diagnosis process by improving the optimization algorithm of the diagnostic model. Theoretically, our proposed SAMD optimization algorithm can converge at an  $O(1/t)$  rate. A 3D aggregated residual network is proposed to extract features from 3D MRI data. The experimental results on the ADNI dataset indicate that our proposed method can improve diagnosing AD and achieve 95.4% accuracy in AD diagnosis and 79.9% accuracy in MCI diagnosis. Compared with several SOTA diagnosis models, our proposed model shows excellent performance. In addition, our proposed SAMD algorithm

can save about 19% of the convergence time on average compared with several gradient descent algorithms.

We will explore further to improve the disease diagnosis's efficiency in the future. Moreover, MRI scans are not performed in many hospitals during AD diagnosis. We may consider different kinds of medical images for the diagnosis of AD and make the diagnosis network adaptable for multimodal data.

**Author Contributions:** All authors contributed to the study conception and design. Y.T.: conceptualization, methodology, software, investigation, formal analysis, writing—original draft. S.L.: resources, supervision, writing—review and Editing. J.Q.: resources, supervision. P.Z.: software, validation. K.H.: software. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Science and Technology Development Project of Liaoning Province of China under Grant number 2021JH6/10500127 and Science Research Project of Liaoning Department of Education of China under Grant number LJKZ0008.

**Institutional Review Board Statement:** A benchmark dataset, Alzheimer's Disease Neuroimaging Initiative (ADNI), was used in our proposed work where the terms of use are declared in the following link: <http://adni.loni.usc.edu/terms-of-use/> (accessed on 19 March 2022).

**Informed Consent Statement:** A benchmark dataset, Alzheimer's Disease Neuroimaging Initiative (ADNI), which was used in our work, has obtained informed consent from the participants. More information can be found in the following link: <http://adni.loni.usc.edu/study-design/> (accessed on 19 March 2022).

**Data Availability Statement:** The dataset used in this study was obtained from Alzheimer's Disease Neuroimaging Initiative (ADNI). More information regarding ADNI can be obtained from the following link: <http://adni.loni.usc.edu/> (accessed on 19 March 2022).

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

## References

- Holtzman, D.M.; Morris, J.C.; Goate, A.M. Alzheimer's disease: The challenge of the second century. *Sci. Transl. Med.* **2011**, *3*, 77sr1. [CrossRef]
- Alzheimer's Disease International. World Alzheimer Report 2022. Available online: <https://www.alzint.org/resource/world-alzheimer-report-2022/> (accessed on 10 July 2023).
- Jia, J.; Wei, C.; Chen, S.; Li, F.; Tang, Y.; Qin, W.; Zhao, L.; Jin, H.; Xu, H.; Wang, F.; et al. The cost of Alzheimer's disease in China and re-estimation of costs worldwide. *Alzheimer's Dement.* **2018**, *14*, 483–491. [CrossRef]
- Smith, J.C.; Nielson, K.A.; Antuono, P.; Lyons, J.A.; Hanson, R.J.; Butts, A.M.; Hantke, N.C.; Verber, M.D. Semantic memory functional MRI and cognitive function after exercise intervention in mild cognitive impairment. *J. Alzheimer's Dis.* **2013**, *37*, 197–215. [CrossRef] [PubMed]
- Altaf, F.; Islam, S.M.; Akhtar, N.; Janjua, N.K. Going deep in medical image analysis: Concepts, methods, challenges, and future directions. *IEEE Access* **2019**, *7*, 99540–99572. [CrossRef]
- Haller, S.; Zanchi, D.; Rodriguez, C.; Giannakopoulos, P. Brain structural imaging in Alzheimer's disease. In *Biomarkers for Preclinical Alzheimer's Disease*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 107–117.
- Albert, M.S.; DeKosky, S.T.; Dickson, D.; Dubois, B.; Feldman, H.H.; Fox, N.C.; Gamst, A.; Holtzman, D.M.; Jagust, W.J.; Petersen, R.C.; et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's Dement.* **2011**, *7*, 270–279. [CrossRef]
- Poloni, K.M.; de Oliveira, I.A.D.; Tam, R.; Ferrari, R.J. Brain MR image classification for Alzheimer's disease diagnosis using structural hippocampal asymmetrical attributes from directional 3-D log-Gabor filter responses. *Neurocomputing* **2021**, *419*, 126–135. [CrossRef]
- Alinsaf, S.; Lang, J.; Alzheimer's Disease Neuroimaging Initiative. 3D Shearlet-Based Descriptors Combined with Deep Features for the Classification of Alzheimer's Disease Based on MRI Data. *Comput. Biol. Med.* **2021**, *138*, 104879. [CrossRef] [PubMed]
- Lian, C.; Liu, M.; Zhang, J.; Shen, D. Hierarchical Fully Convolutional Network for Joint Atrophy Localization and Alzheimer's Disease Diagnosis Using Structural MRI. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *42*, 880–893. [CrossRef]
- Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
- Tieleman, T.; Hinton, G. Rmsprop: Divide the gradient by a running average of its recent magnitude. coursera: Neural networks for machine learning. *COURSERA Neural Netw. Mach. Learn.* **2012**, *4*, 26–31.
- Ben-Tal, A.; Margalit, T.; Nemirovski, A. The ordered subsets mirror descent optimization method with applications to tomography. *SIAM J. Optim.* **2001**, *12*, 79–108. [CrossRef]

14. Baron, J.; Chetelat, G.; Desgranges, B.; Perchey, G.; Landeau, B.; De La Sayette, V.; Eustache, F. In vivo mapping of gray matter loss with voxel-based morphometry in mild Alzheimer's disease. *Neuroimage* **2001**, *14*, 298–309. [[CrossRef](#)]
15. Fan, Y.; Shen, D.; Gur, R.C.; Gur, R.E.; Davatzikos, C. COMPARE: Classification of morphological patterns using adaptive regional elements. *IEEE Trans. Med. Imaging* **2006**, *26*, 93–105. [[CrossRef](#)]
16. Ju, R.; Hu, C.; Zhou, P.; Li, Q. Early diagnosis of Alzheimer's disease based on resting-state brain networks and deep learning. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2017**, *16*, 244–257. [[CrossRef](#)]
17. Coupé, P.; Eskildsen, S.F.; Manjón, J.V.; Fonov, V.S.; Pruessner, J.C.; Allard, M.; Collins, D.L.; Alzheimer's Disease Neuroimaging Initiative. Scoring by nonlocal image patch estimator for early detection of Alzheimer's disease. *NeuroImage Clin.* **2012**, *1*, 141–152. [[CrossRef](#)]
18. 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](#)]
19. Cui, R.; Liu, M.; Alzheimer's Disease Neuroimaging Initiative. RNN-based longitudinal analysis for diagnosis of Alzheimer's disease. *Comput. Med. Imaging Graph.* **2019**, *73*, 1–10. [[CrossRef](#)] [[PubMed](#)]
20. Feng, C.; Elazab, A.; Yang, P.; Wang, T.; Zhou, F.; Hu, H.; Xiao, X.; Lei, B. Deep learning framework for Alzheimer's disease diagnosis via 3D-CNN and FSBi-LSTM. *IEEE Access* **2019**, *7*, 63605–63618. [[CrossRef](#)]
21. Puente-Castro, A.; Fernandez-Blanco, E.; Pazos, A.; Munteanu, C.R. Automatic assessment of Alzheimer's disease diagnosis based on deep learning techniques. *Comput. Biol. Med.* **2020**, *120*, 103764. [[CrossRef](#)]
22. Alloui, H.; Sadgal, M.; Elfazziki, A. Utilization of a convolutional method for Alzheimer disease diagnosis. *Mach. Vis. Appl.* **2020**, *31*, 1–19. [[CrossRef](#)]
23. Pan, Y.; Liu, M.; Lian, C.; Zhou, T.; Xia, Y.; Shen, D. Synthesizing missing PET from MRI with cycle-consistent generative adversarial networks for Alzheimer's disease diagnosis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention, Proceedings of the 21st International Conference, Granada, Spain, 16–20 September 2018*; Springer: Cham, Switzerland, 2018; pp. 455–463.
24. 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, Jeju Island, Republic of Korea, 15–18 June 2019*; pp. 87–91.
25. Zhang, Y.; Wang, S.; Xia, K.; Jiang, Y.; Qian, P.; Alzheimer's Disease Neuroimaging Initiative. Alzheimer's disease multiclass diagnosis via multimodal neuroimaging embedding feature selection and fusion. *Inf. Fusion* **2021**, *66*, 170–183. [[CrossRef](#)]
26. Santos-Rodríguez, R.; García-García, D.; Cid-Sueiro, J. Cost-sensitive classification based on Bregman divergences for medical diagnosis. In *Proceedings of the 2009 International Conference on Machine Learning and Applications, Miami, FL, USA, 13–15 December 2009*; pp. 551–556.
27. Angulo, H.; Giménez, J.; Moros, A.M.; Nikodem, K. On strongly  $h$ -convex functions. *Ann. Funct. Anal.* **2011**, *2*, 85–91. [[CrossRef](#)]
28. Hardt, M.; Ma, T.; Recht, B. Gradient Descent Learns Linear Dynamical Systems. *J. Mach. Learn. Res.* **2018**, *19*, 1–44.
29. Nedic, A.; Lee, S. On stochastic subgradient mirror-descent algorithm with weighted averaging. *SIAM J. Optim.* **2014**, *24*, 84–107. [[CrossRef](#)]
30. Xu, L.; Yao, Z.; Li, J.; Lv, C.; Zhang, H.; Hu, B. Sparse feature learning with label information for Alzheimer's disease classification based on magnetic resonance imaging. *IEEE Access* **2019**, *7*, 26157–26167. [[CrossRef](#)]
31. Cui, R.; Liu, M. Hippocampus analysis by combination of 3-D DenseNet and shapes for Alzheimer's disease diagnosis. *IEEE J. Biomed. Health Inform.* **2019**, *23*, 2099–2107. [[CrossRef](#)] [[PubMed](#)]
32. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
33. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016*; pp. 770–778.
34. Janoušová, E.; Vounou, M.; Wolz, R.; Gray, K.R.; Rueckert, D.; Montana, G. Biomarker discovery for sparse classification of brain images in Alzheimer's disease. *Ann. BMVA* **2012**, *2012*, 1–11.
35. Liu, S.; Liu, S.; Cai, W.; Che, H.; Pujol, S.; Kikinis, R.; Feng, D.; Fulham, M.J.; ADNI. Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease. *IEEE Trans. Biomed. Eng.* **2015**, *62*, 1132–1140. [[CrossRef](#)]
36. 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, VIC, Australia, 18–21 April 2017*; pp. 835–838.
37. Karasawa, H.; Liu, C.L.; Ohwada, H. Deep 3D convolutional neural network architectures for alzheimer's disease diagnosis. In *Asian Conference on Intelligent Information and Database Systems, Proceedings of the 10th Asian Conference, ACIIDS 2018, Dong Hoi City, Vietnam, 19–21 March 2018*; Springer: Cham, Switzerland, 2018; pp. 287–296.
38. Khvostikov, A.; Aderghal, K.; Benois-Pineau, J.; Krylov, A.; Catheline, G. 3D CNN-based classification using sMRI and MD-DTI images for Alzheimer disease studies. *arXiv* **2018**, arXiv:1801.05968.
39. Lin, W.; Tong, T.; Gao, Q.; Guo, D.; Du, X.; Yang, Y.; Guo, G.; Xiao, M.; Du, M.; Qu, X.; et al. Convolutional neural networks-based MRI image analysis for the Alzheimer's disease prediction from mild cognitive impairment. *Front. Neurosci.* **2018**, *12*, 777. [[CrossRef](#)]
40. Zhu, T.; Cao, C.; Wang, Z.; Xu, G.; Qiao, J. Anatomical Landmarks and DAG Network Learning for Alzheimer's Disease Diagnosis. *IEEE Access* **2020**, *8*, 206063–206073. [[CrossRef](#)]

41. Lin, W.; Lin, W.; Chen, G.; Zhang, H.; Gao, Q.; Huang, Y.; Tong, T.; Du, M.; Alzheimer's Disease Neuroimaging Initiative. Bidirectional mapping of brain MRI and PET with 3D reversible GAN for the diagnosis of Alzheimer's disease. *Front. Neurosci.* **2021**, *15*, 646013. [[CrossRef](#)]
42. Gao, X.; Shi, F.; Shen, D.; Liu, M. Task-induced pyramid and attention GAN for multimodal brain image imputation and classification in alzheimer's disease. *IEEE J. Biomed. Health Inform.* **2021**, *26*, 36–43. [[CrossRef](#)] [[PubMed](#)]
43. Li, H.; Tan, Y.; Miao, J.; Liang, P.; Gong, J.; He, H.; Jiao, Y.; Zhang, F.; Xing, Y.; Wu, D. Attention-based and micro designed EfficientNetB2 for diagnosis of Alzheimer's disease. *Biomed. Signal Process. Control* **2023**, *82*, 104571. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.