Article

On Disharmony in Batch Normalization and Dropout Methods for Early Categorization of Alzheimer's Disease

Ahsan Bin Tufail [1,2,†], Inam Ullah [3,†], Ateeq Ur Rehman [4], Rehan Ali Khan [5], Muhammad Abbas Khan [6], Yong-Kui Ma [2], Nadar Hussain Khokhar [7], Muhammad Tariq Sadiq [8], Rahim Khan [2], Muhammad Shafiq [9,*], Elsayed Tag Eldin [10,*] and Nivin A. Ghamry [11]

1   Department of Computer Science, National University of Sciences and Technology, Balochistan Campus, Quetta 87300, Pakistan
2   School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150001, China
3   BK21 Chungbuk Information Technology Education and Research Center, Chungbuk National University, Cheongju 28644, Korea
4   Department of Electrical Engineering, Government College University, Lahore 54000, Pakistan
5   Department of Electrical Engineering, University of Science and Technology Bannu, Bannu 28100, Pakistan
6   Department of Electrical Engineering, FICT, Balochistan University of Information Technology, Engineering and Management Sciences, Quetta 87300, Pakistan
7   Department of Civil Engineering, National University of Sciences and Technology, Balochistan Campus, Quetta 87300, Pakistan
8   School of Architecture, Technology and Engineering, University of Brighton, Brighton BN2 4AT, UK
9   Department of Information and Communication Engineering, Yeungnam University, Gyeongsan 38541, Korea
10   Faculty of Engineering and Technology, Future University in Egypt, New Cairo 11835, Egypt
11   Faculty of Computers and Artificial Intelligence, Cairo University, Giza 3750010, Egypt
*   Correspondence: shafiq@ynu.ac.kr (M.S.); elsayed.tageldin@fue.edu.eg (E.T.E.)
†   These authors contributed equally to this work.

Abstract: Alzheimer's disease (AD) is a global health issue that predominantly affects older people. It affects one's daily activities by modifying neural networks in the brain. AD is categorized by the death of neurons, the creation of amyloid plaques, and the development of neurofibrillary tangles. In clinical settings, an early diagnosis of AD is critical to limit the problems associated with it and can be accomplished using neuroimaging modalities, such as magnetic resonance imaging (MRI) and positron emission tomography (PET). Deep learning (DL) techniques are widely used in computer vision and related disciplines for various tasks such as classification, segmentation, detection, etc. CNN is a sort of DL architecture, which is normally useful to categorize and extract data in the spatial and frequency domains for image-based applications. Batch normalization and dropout are commonly deployed elements of modern CNN architectures. Due to the internal covariance shift between batch normalization and dropout, the models perform sub-optimally under diverse scenarios. This study looks at the influence of disharmony between batch normalization and dropout techniques on the early diagnosis of AD. We looked at three different scenarios: (1) no dropout but batch normalization, (2) a single dropout layer in the network right before the softmax layer, and (3) a convolutional layer between a dropout layer and a batch normalization layer. We investigated three binaries: mild cognitive impairment (MCI) vs. normal control (NC), AD vs. NC, AD vs. MCI, one multiclass AD vs. NC vs. MCI classification problem using PET modality, as well as one binary AD vs. NC classification problem using MRI modality. In comparison to using a large value of dropout, our findings suggest that using little or none at all leads to better-performing designs.

Keywords: neuroimaging; classification; augmentation; statistical comparison; batch normalization; dropout

1. Introduction

Alzheimer's disease (AD) is a brain illness that is believed to be a continuum; it is preferable to detect it in the preclinical stage so that treatment may begin as soon as

possible. Changes are introduced at three levels of hierarchy in AD: neuronal, regional, and clinical, all of which have substantial influences on the daily routines of AD patients [1]. The formation and development of tangles, the deregulation and degradation of protein pathways, inflammation of neurons, and other factors, play roles in the evolution of AD [2]. Subjective cognitive decline, which may be the first indication of preclinical AD, can occur before AD affects people in clinical settings and communities [3–7]. In fact, AD is so closely linked to age that some researchers believe it may be a natural component of the aging process [8]. Late-onset AD varies significantly from early-onset AD with diversity in phenotypic presentation and involves more distinct brain pathways than usual AD [9], making it difficult to divide participants into coherent groupings. Furthermore, several alleles related to AD, such as APOE e4 and APOE e2, have been linked to increased or decreased risks of AD [10]. Furthermore, sex and genetic factors, such as T2D, are modifiable risk factors in the evolution of AD, which may contribute to the prevention of this neurodegenerative disorder [11]. Plasma metabolites, lipokines, and amino acids have been linked to AD and are helpful in providing information about changes in the structure of the brain [12], particularly the link between brain arterial dilation and the pathophysiology of AD [13], regardless of demographics, education, or other related factors.

Mild cognitive impairment (MCI), both amnestic and non-amnestic, is a type of cognitive impairment marked by a reduction in gray matter volume, particularly in the angular gyrus region, and is hypothesized to precede AD. MCI patients may not meet the clinical criteria for AD. Positron emission tomography (PET) and magnetic resonance imaging (MRI) biomarkers can detect discriminating hypometabolism as well as other functional alterations linked to AD [14], such as vascular, lipid-metabolic, and immune-related abnormalities [15].

In medical imaging, computer-assisted diagnosis (CAD) is commonly used to diagnose a variety of disorders. Deep learning (DL) is a sort of CAD in which features are extracted automatically without the intervention of a user. Throughout the training, it develops effective representations of the underlying data distribution using information gathered from the input data. CNNs are DL methods that are well-known for handling image-based problems.

In the literature, various methods have been proposed for AD vs. normal control (NC) using discrete volume estimation CNNs deploying structural MRI scans, achieving normal weighted classification accuracies of 94.82 and 94.02% [16–18]; AD vs. NC and MCI-static vs. MCI-converter (MCI-S/MCI-C) classification tasks achieved balanced accuracies of 79.1 and 70.4% on these tasks, respectively, using autoencoder-based networks [19]; NC vs. very mild-demented vs. mild-demented vs. moderate-demented multiclass classification tasks achieved an accuracy of 95.23% using a deep CNN model [20]; a combination of random search, transfer learning and snapshot ensembles for AD vs. NC and MCI vs. AD tasks attained precisions of 99.05 and 98.71%, respectively [21–25].

Similarly, authors proposed a deep recurrent neural network for binary classification between AD and NC classes [26]; a deep CNN model comprising multiple layers, such as a pooling layer, feature mapping, and other layers for diverse classification tasks (binary and multiclass) using NC, AD, MCI-S, and MCI-C classes has been used to attain state-of-the-art performances on these tasks [27]; transfer learning based Inception-ResNet-V2 architecture for classification of AD [28]; as well as an ensemble of logistic regression and linear support vector machine for NC/MCI classification [29].

Correspondingly, authors provide a survey of latest techniques for AD diagnosis using deep learning methods [30]; MCI-S vs. MCI-C classification achieved an accuracy of 85.5% utilizing MRI and genotype data using the random forest algorithm and other methods [31]; AD vs. NC binary classification, via a mixture of discrete wavelet transform (DWT)-based features, color moments, and a feedforward artificial neural network (ANN) classifier achieved an accuracy of 97.14% [32]; 2D slice level CNN model for AD/NC and pMCI/sMCI classification tasks achieving an accuracy of 93% on AD/NC and 83% on pMCI/sMCI classification tasks [33]; an autoencoder-based architecture for classification

and prediction of AD using resting-state functional MRI data achieved an accuracy of approximately 95% using correlation coefficient data and approximately 90% using resting-state functional MRI data [34].

Batch normalization and dropout are widely adopted elements of modern deep CNN architectures. Dropout overcomes overfitting by dropping units and their connections randomly and stochastically, performing regularization in the process by providing an uncertainty estimate for a prediction that is dependent on its probabilities. Disharmony among batch normalization and dropout methods leads to deteriorating performances on discrimination tasks and is due to an internal covariance shift resulting in unstable numerical behaviors in inference due to their distinct test policies, causing an improper neural variance with a shift in information flow in inference, leading to erroneous predictions. Furthermore, batch normalization tries to reach a stable distribution of activation values throughout training by subtracting the batch mean and dividing by the batch standard deviation. To mitigate the variance shift phenomenon, the population variance needs to be recomputed to compensate for the variance shift caused by the inference mode of dropout [35–43].

In this study, the disharmony effects in dropout and batch normalization methods on the presentations of deep CNNs for the early detection of AD were examined. We used 3D scans from the MRI and PET neuroimaging modalities and sorted them into MCI, NC, and AD classes using 3D-CNN architectures. Four problems were considered: multiclass classifications of MCI, NC, and AD classes, as well as binary classifications of MCI and NC, MCI and AD, and NC and AD classes. For the AD vs. NC classification task, we used random zoomed-in/out data augmentation for the MRI modality, but no data augmentation for the PET modality (for the binary and multiclass classification tasks was deployed). We considered three distinct scenarios: (1) no dropout but batch normalization, (2) just a single layer of dropout before the softmax layer, and (3) a convolutional layer between the dropout and batch normalization layers.

The remainder of the paper is laid out below. Section 2 explains the datasets, whereas Section 3 describes the approach. Experiments and results are presented in Section 4, while Section 5 presents the discussion. Lastly, Section 6 concludes this work, giving pointers for further research in this field.

2. Datasets Description

Scans from the AD Neuroimaging Initiative (ADNI) database were considered in this study [44]. Tables 1 and 2 reflect the demographics of the participants considered in this study. For the experiments, we used entire brain scans in the 3D domain. We picked a volume size of $121 \times 145 \times 41$ for the T1-weighted MRI modality in the sagittal plane and a volume size of $79 \times 95 \times 69$ for the PET modality.

Table 1. The mean (min–max) from PET scans.

Demographics	NC	MCI	AD
Subjects	102	97	94
Age	76.01 (62.21–86.62)	74.54 (55.32–87.23)	75.82 (55.32–881)
Weight	75.72 (49–130.3)	77.13 (45.1–120.2)	74.12 (42.62–127.54)
Functional Activities Questionnaire Total Score	0.1863 (0–6)	3.163 (0–15)	13.672 (0–27)
Neuropsychiatric Inventory Questionnaire Total Score	0.4023 (0–5)	1.973 (0–17)	4.0741 (0–15)

Table 2. The mean (min–max) from MRI scans.

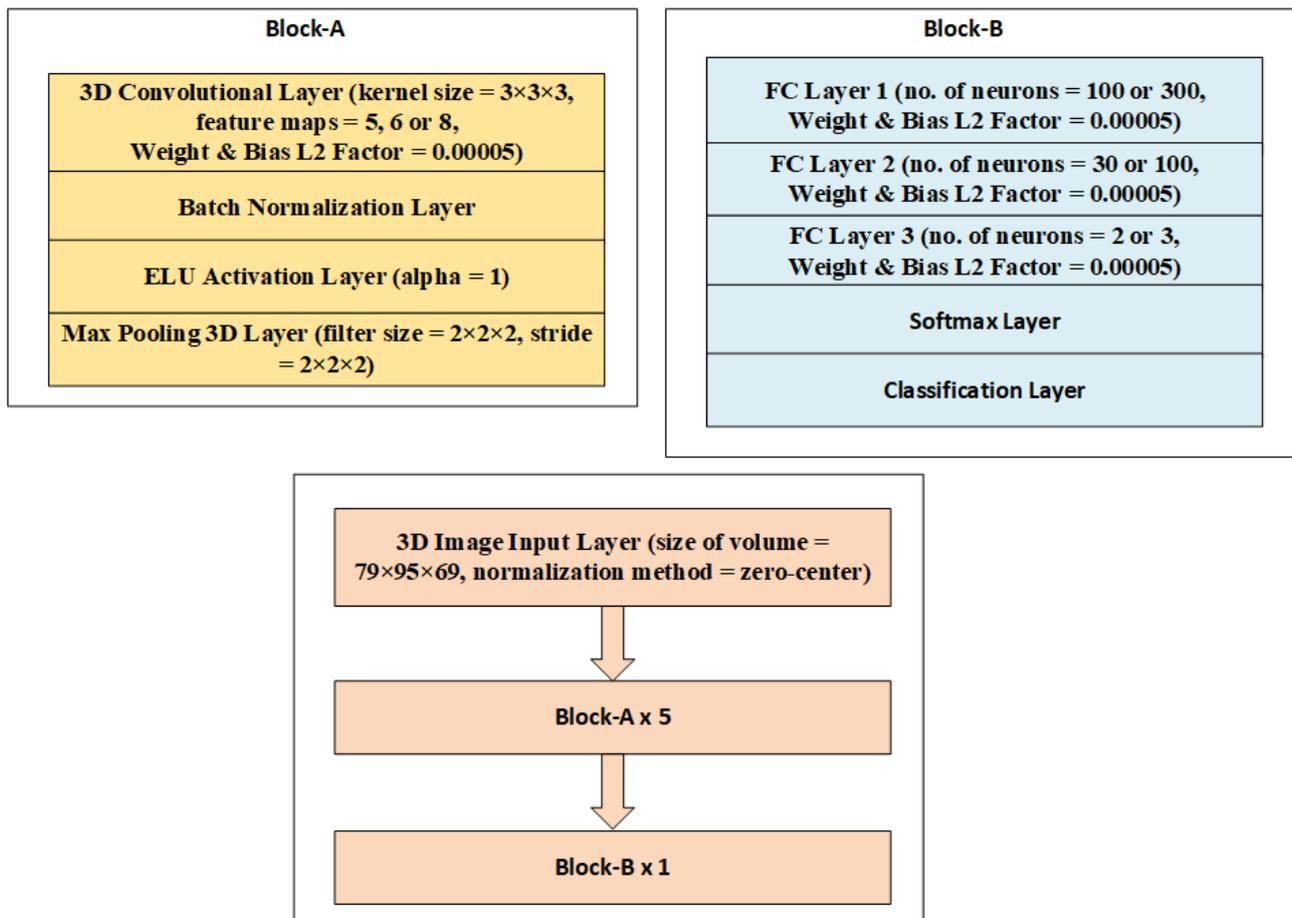| Demographics	NC	AD
Subjects Number	228	187
Age	75.97 (60.02–89.74)	75.4 (55.18–90.99)
Weight	75.91 (45.81–137.44)	72.03 (37.65–127.46)
Mini-Mental State Examination Score	29.11 (25–30)	23.26 (18–27)
Clinical Dementia Rating Global Score	0 (0–0)	0.75 (0.5–1)

3. Methodology

This research looks at four problems: a multiclass (three-class) classification problem between MCI, NC, and AD classes, and three binary classification problems that include MCI and NC, MCI and AD, and NC and AD classes. With the PET dataset, we looked at all four classification problems, whereas with the MRI dataset, we only looked at the AD vs. NC binary classification problem.

The designs for processing MRI and PET scans are shown in Figures 1 and 2 under scenario-1. After performing a zero-centered normalization procedure that subtracts the mean and divides it by the standard deviation, one input layer takes a volume size of $121 \times 145 \times 41$ or $79 \times 95 \times 69$, as illustrated in these diagrams. The input is subsequently sent via block-A, which is repeated five times. The block is composed of a 3D-convolutional layer for feature extraction, a batch normalization layer, and an exponential linear unit (ELU) non-linear activation layer, which works by gradually reducing the bias shift effect while reducing the cost to zero. Mathematically,

$$ELU = \begin{cases} z, & z > 0 \\ \alpha(e^z - 1), & z \le 0 \end{cases} \tag{1}$$
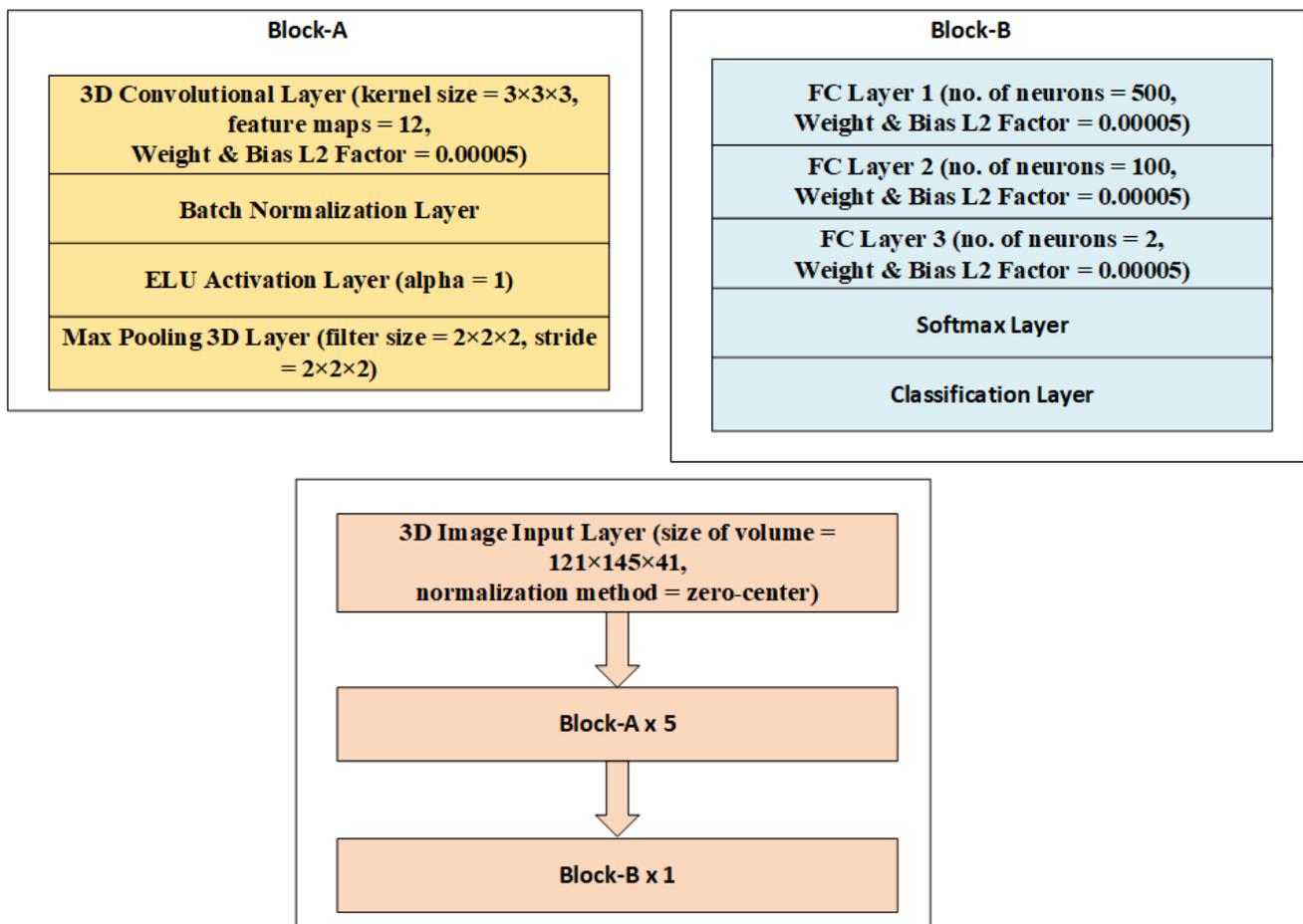
Figure 1. Architecture of 3D-CNN MRI scan processing for the classification of AD vs. NC binary tasks under scenario-1.

Block-A
3D Convolutional Layer (kernel size = 3×3×3, feature maps = 5, 6 or 8, Weight & Bias L2 Factor = 0.00005)
Batch Normalization Layer
ELU Activation Layer (alpha = 1)
Max Pooling 3D Layer (filter size = 2×2×2, stride = 2×2×2)

Block-B
FC Layer 1 (no. of neurons = 100 or 300, Weight & Bias L2 Factor = 0.00005)
FC Layer 2 (no. of neurons = 30 or 100, Weight & Bias L2 Factor = 0.00005)
FC Layer 3 (no. of neurons = 2 or 3, Weight & Bias L2 Factor = 0.00005)
Softmax Layer
Classification Layer

3D Image Input Layer (size of volume = 79×95×69, normalization method = zero-center)
Block-A x 5
Block-B x 1

Figure 2. Architecture of 3D-CNN PET scan processing for binary and multiclass classifications under scenario-1.

The max pooling 3D layer reduces the dimensionality of feature maps by using larger stride widths, making the entire process more computationally efficient. To reduce overfitting, the convolutional layers use small weights and bias L2 parameters. After repeating block-A five times, there is a single block named block-B that has three fully connected (FC), one softmax, and one classification layer. Each activation unit of one layer is connected to each activation unit of the next layer, which then compiles the data received by previous levels to produce the final output. In terms of temporal efficiency, these layers are only second to convolutional layers. Finally, a softmax layer operates by squashing the inputs in this range to normalize neural units between zero and one, allowing outputs to be interpreted as probabilities. Mathematically,

$$softmax\left(\vec{z}\right) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \tag{2}$$

The classification architectures in Figures 3 and 4 are nearly identical to those in Figures 1 and 2, except for a single dropout with a 0.1 probability right before the softmax layer in the entire design, which randomly removes neurons acting as the regularizer to minimize overfitting. Figures 5 and 6 depict structures that correspond to scenario-3. An input layer in these designs uses the zero-center normalization technique to center volume sizes of $121 \times 145 \times 41$ or $79 \times 95 \times 69$ toward the origin. Following this input, there is a single block labeled "block-A", four blocks labeled "block-B", and a single block labeled "block-C". These blocks are made up of convolutional, batch normalization, dropout layers, max-pooling layers, FC, ELU activation, and softmax layers, as shown in Figures 5 and 6.
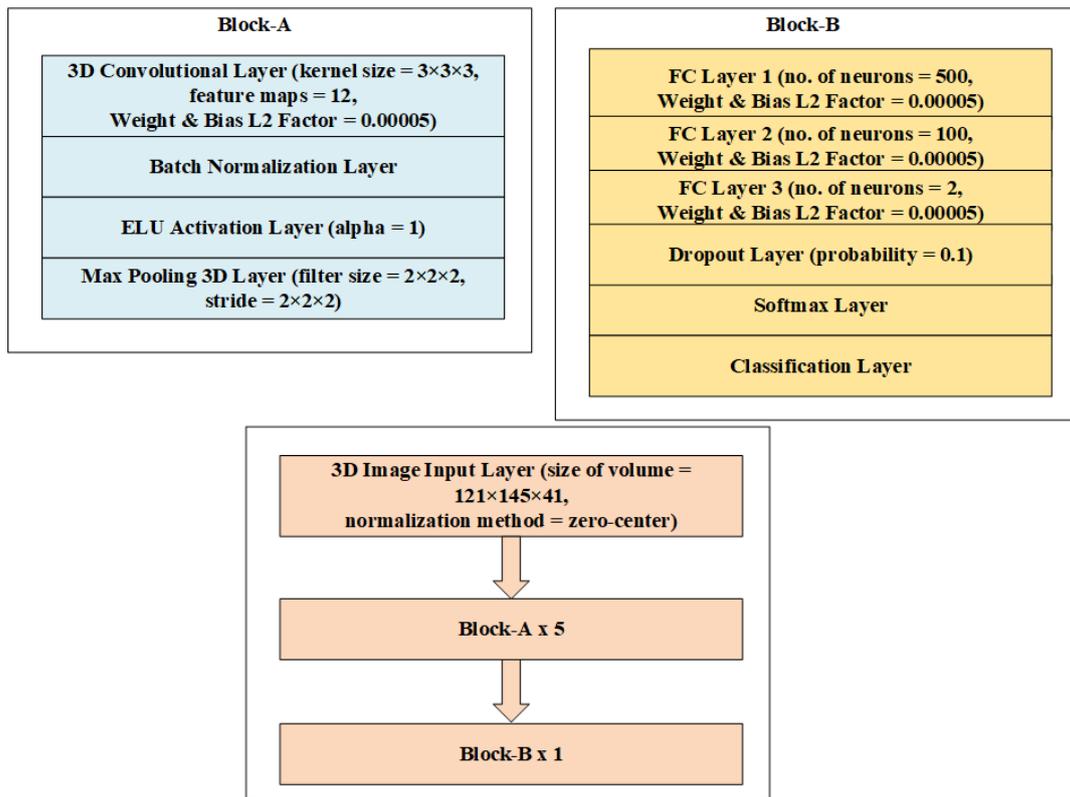
Block-A

3D Convolutional Layer (kernel size = 3×3×3, feature maps = 12, Weight & Bias L2 Factor = 0.00005)

Batch Normalization Layer

ELU Activation Layer (alpha = 1)

Max Pooling 3D Layer (filter size = 2×2×2, stride = 2×2×2)

Block-B

FC Layer 1 (no. of neurons = 500, Weight & Bias L2 Factor = 0.00005)

FC Layer 2 (no. of neurons = 100, Weight & Bias L2 Factor = 0.00005)

FC Layer 3 (no. of neurons = 2, Weight & Bias L2 Factor = 0.00005)

Dropout Layer (probability = 0.1)

Softmax Layer

Classification Layer

3D Image Input Layer (size of volume = 121×145×41, normalization method = zero-center)

Block-A x 5

Block-B x 1

Figure 3. Architecture of 3D-CNN MRI scan processing for AD vs. NC binary classification tasks under scenario-2.

Block-A

3D Convolutional Layer (kernel size = 3×3×3, feature maps = 5, 6 or 8, Weight & Bias L2 Factor = 0.00005)

Batch Normalization Layer

ELU Activation Layer (alpha = 1)

Max Pooling 3D Layer (filter size = 2×2×2, stride = 2×2×2)

Block-B

FC Layer 1 (no. of neurons = 100 or 300, Weight & Bias L2 Factor = 0.00005)

FC Layer 2 (no. of neurons = 30 or 100, Weight & Bias L2 Factor = 0.00005)

FC Layer 3 (no. of neurons = 2 or 3, Weight & Bias L2 Factor = 0.00005)

Dropout Layer (probability = 0.1)

Softmax Layer

Classification Layer

3D Image Input Layer (size of volume = 79×95×69, normalization method = zero-center)
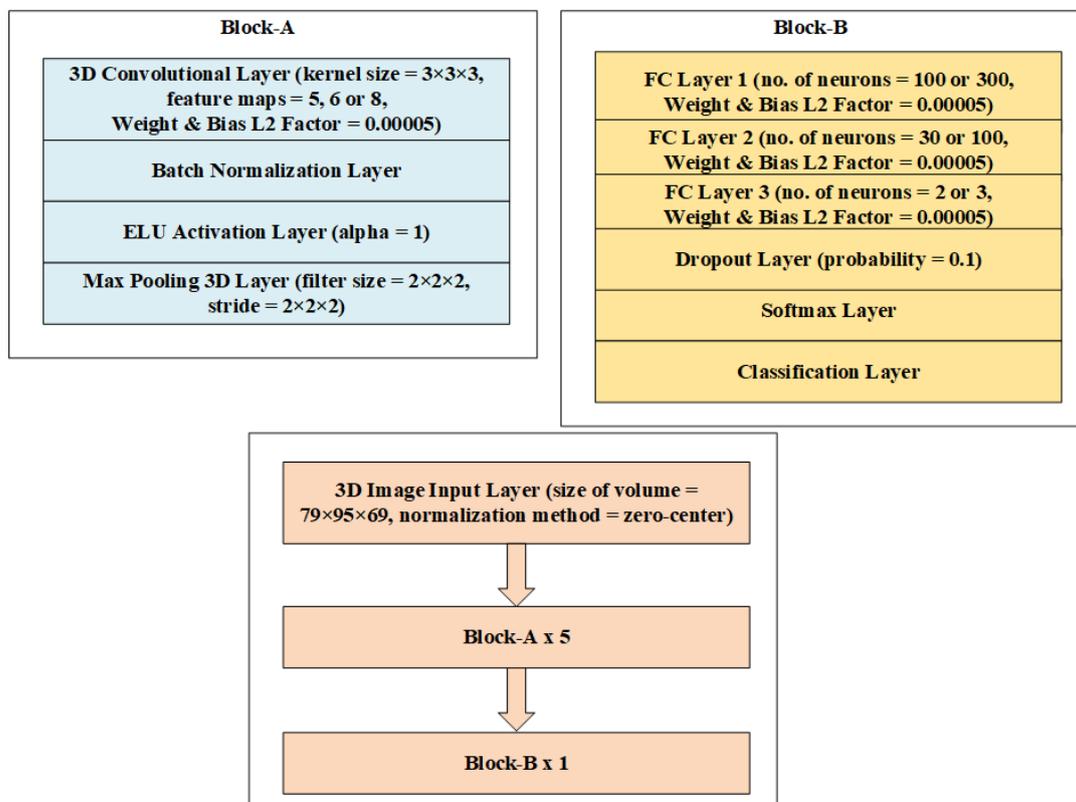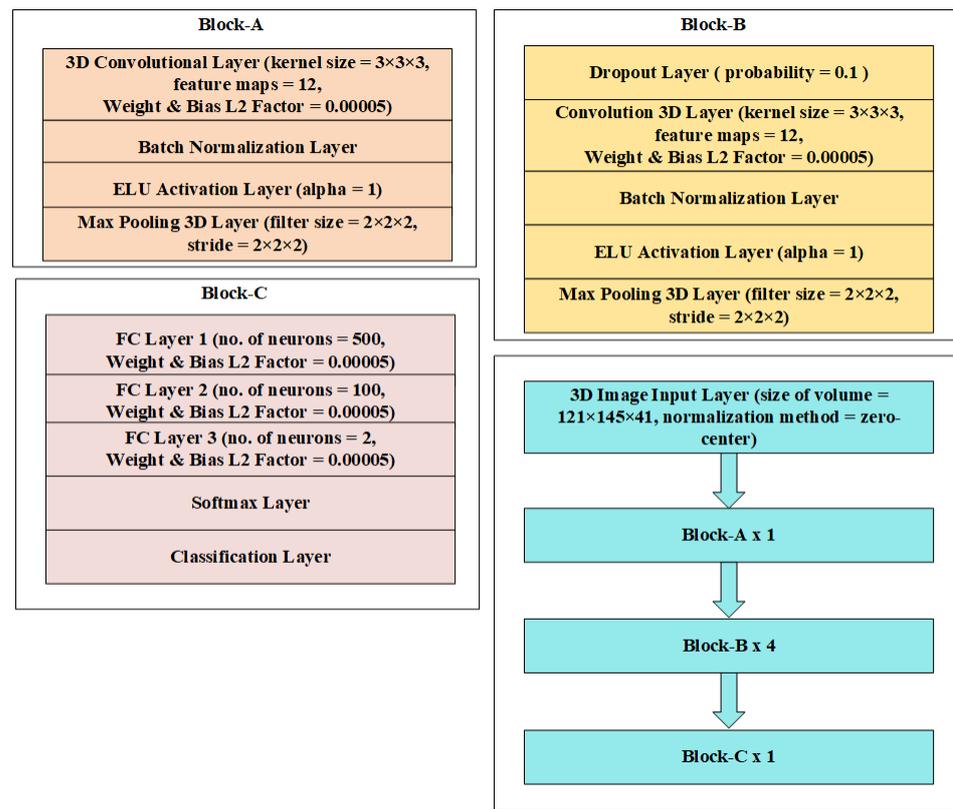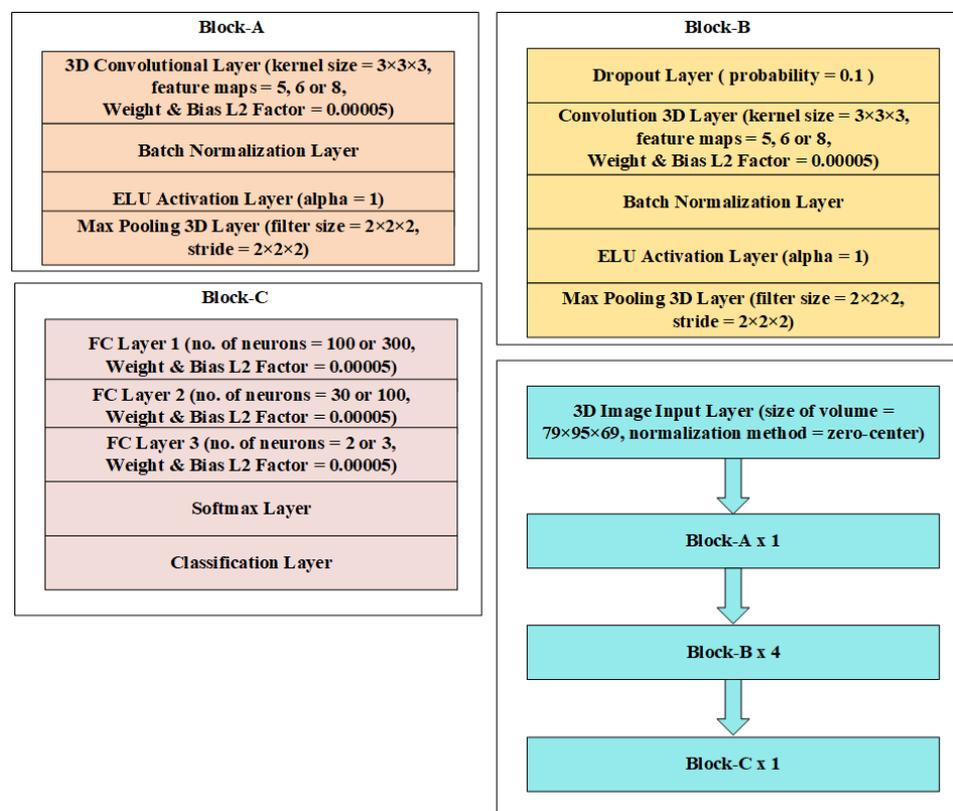
Block-A x 5

Block-B x 1

Figure 4. Architecture of 3D-CNN PET scan processing for binary and multiclass classification tasks under scenario-2.

Figure 5. Architecture of 3D-CNN MRI scan processing for the classification of AD vs. NC binary tasks under scenario-3.

Figure 6. Architecture of 3D-CNN PET scan processing for binary and multiclass classification tasks under scenario-3.

As given in Figure 6, for the AD vs. NC vs. MCI tasks, under scenario-3, the feature maps in the 3D convolutional layer in block-A and block-B are 8, and there are 300, 100, and 3 neurons in FC layer 1, FC layer 2, and FC layer 3, respectively. The feature mappings in the 3D convolutional layer in block-A and block-B for the NC vs. MCI and MCI vs. AD binary classification tasks are 6, followed by 100, 30, and 2 neurons in FC layers 1, 2, and 3, respectively.

4. Experiments and Results

In this research, a five-fold cross-validation (CV) method was applied in the experiments for hyperparameter selections. A separate test set was also generated, and it was never used during the training process for the performance assessment. In the test set, there were 12 NC class cases, 7 MCI class cases, and 4 AD instances from the PET data set, as well as 8 NC and 8 AD instances from the MRI data set. Since we utilized both balanced and unbalanced classes, we established a different set of performance criteria to investigate binary and multiclass classification tasks. For balanced multiclass and unbalanced binary classification tasks, confusion entropy (CEN), relative classifier information (RCI), Matthew's correlation coefficient (MCC), and geometric mean (GM) were used as performance measures. For balanced binary classification tasks, sensitivity (SEN), specificity (SPEC), F-measure, precision, and balanced accuracy were employed as performance metrics.

We did our best to avoid the data leakage problem mentioned by Junhao Wen et al. [45] by avoiding an incorrect data split, biased transfer learning, late split, and the lack of an autonomous test set. In addition, we followed the experimental design of Simeon Spasov et al. [46] and our previous work [47] for the experiments. Simeon Spasov et al. [46] reported only the validation results for the AD vs. NC classification tasks. We followed their procedure and that of [47] by including results from mutually validated and independent test sets by simply summing the results from these two sets, resulting in a total of six results for a task, and we used categorical cross-entropy as the loss function. For experimental purposes, regarding the independent test set, we considered models in the validation and training splits as training sets and samples in the independent test set as validation sets; we report the results on the last epoch. In fact, for all of the experiments, we report the results obtained in the last epoch. We ensured that our independent test set was never used for choosing the training hyperparameters since the validation set provides this functionality and, hence, was used for this purpose. In fact, the validation set provides dual functionality by providing the evaluation results as well as tuning the hyperparameters for the experiments. One reason for adopting this strategy is that we wanted to ensure that more samples were employed in the assessment of results and, secondly, we observed that the performance on the independent test sets was even better than the performance on validation splits.

For all classification problems employing the PET modality, there were a total of 72 subjects in the training split and 18 in the validation split for each class, such as AD, MCI, and NC. There were 44 NC class instances (as well as 35 or 36 AD class instances) in the validation split, and 176 NC class instances (and 144 or 143 AD class instances) in the training split, respectively, for trials employing the MRI modality. We used Adam as an optimizer and categorical cross-entropy as a loss function throughout the experiments.

RCI distinguishes among classes while taking class predictions into account. RCI is built using the predictability offered by the predicted classes and enables the comparison of classifiers in the same domain. CEN calculates the rate of misclassification between a single class and all other classes by using the off-diagonal components in a confusion matrix to utilize the class distribution information. It is highly dependent on the number of samples since a larger number of samples yields a better result for CEN. By integrating individual accuracy rates with total accuracy, IBA is intended to offer information about the dominating class. GM measures the amount to which a class recollection is aggregated by concentrating on the recall of a single class. Finally, MCC may be used to determine the Pearson coefficient of correlation between products and moments by correlating the difference between observed and actual readings.

The ratio of individuals who are expected to test positive for an illness compared to those who really have the ailment is known as sensitivity or recall. The ratio of individuals who are expected to test negative for an illness compared to those who do not really have the ailment is known as specificity. By balancing precision and recall, the F-measure gives information on a test's accuracy. The precision determines which forecasts in the positive class are correct. Finally, balanced accuracy is a desirable statistic for classification tasks in which sensitivity and specificity are equally important.

The results of the tests, as shown in Tables 3–6, will now be discussed. Table 3 shows that the structures used in scenarios-1 and 2 significantly reduce the disharmony between batch normalization and dropout methods for multiclass classification between AD, MCI, and NC classes using PET modalities, whereas scenario-3 significantly increases it. The best architecture, according to RCI and average CEN metrics, is 3D-CNN in scenario-1, while the worst performing architecture is 3D-CNN in scenario-3 using PET modality. Based on average IBA, GM, and MCC measures, the 3D-CNN trained under scenario-2 has the better performance, whereas the 3D-CNN trained under scenario-3 using the PET modality has the worst performance. Generally, the best-performing architecture is the 3D-CNN trained under scenario-2, whereas the worst-performing architecture is the 3D-CNN trained under scenario-3 using the PET modality. Because of the better performance of the 3D-CNN trained under scenario-2, which employs the PET modality, we can observe that dropout and batch normalization together may yield higher performance outcomes. It is also worth noting that the architecture trained under scenario-3 had the worst results, which could be explained by the disharmony in the dropout and batch normalization methods, which have lowered performances in this case due to the internal covariance shift phenomenon, which caused these methods to behave differently during the training and test phases of the neuronal units [43]. The results and rankings of the architectures trained under the three scenarios for the multiclass classification task are given in Figures 7 and 8.

Table 3. Multiclass classification among AD, MCI, and NC classes using PET modality.

Architecture Used	Metrics of Performance
3D-CNN applying PET data to study scenario-1	RCI = 0.2261, CEN = {'AD': 0.5054, 'MCI': 0.8572, 'NC': 0.4996}, Average CEN = 0.6207, IBA = {'AD': 0.4539, 'MCI': 0.1316, 'NC': 0.4969}, Average IBA = 0.3608, GM = {'AD': 0.7484, 'MCI': 0.4683, 'NC': 0.7344}, Average GM = 0.6503, MCC = {'AD': 0.5118, 'MCI': 0.019, 'NC': 0.4601}, Average MCC = 0.3303
3D-CNN applying PET data to study scenario-2	RCI = 0.1923, CEN = {'AD': 0.5448, 'MCI': 0.8405, 'NC': 0.5407}, Average CEN = 0.642, IBA = {'AD': 0.4674, 'MCI': 0.1516, 'NC': 0.4734}, Average IBA = 0.3641, GM = {'AD': 0.7454, 'MCI': 0.4982, 'NC': 0.7289}, Average GM = 0.6575, MCC = {'AD': 0.4953, 'MCI': 0.0721, 'NC': 0.4543}, Average MCC = 0.3405
3D-CNN applying PET data to study scenario-3	RCI = 0.1628, CEN = {'AD': 0.5692, 'MCI': 0.8398, 'NC': 0.5501}, Average CEN = 0.653, IBA = {'AD': 0.4407, 'MCI': 0.0411, 'NC': 0.5748}, Average IBA = 0.3522, GM = {'AD': 0.7424, 'MCI': 0.3597, 'NC': 0.6975}, Average GM = 0.5998, MCC = {'AD': 0.5026, 'MCI': −0.011, 'NC': 0.3881}, Average MCC = 0.2932

**Table 4.** Binary classifications among AD and MCI classes using PET modality.

Architecture Used	Metrics of Performance
3D-CNN applying PET data to study scenario-1	SEN = 0.6383, SPEC = 0.6907, F-measure = 0.6522, Precision = 0.6667, Balanced Accuracy = 0.6645
3D-CNN applying PET data to study scenario-2	SEN = 0.6702, SPEC = 0.7113, F-measure = 0.6811, Precision = 0.6923, Balanced Accuracy = 0.6908
3D-CNN applying PET data to study scenario-3	SEN = 0.7234, SPEC = 0.4845, F-measure = 0.6415, Precision = 0.5763, Balanced Accuracy = 0.6040

Table 5. Binary classifications among AD and NC classes using PET and MRI modalities.

Architecture	Performance Metrics
3D-CNN applying PET data to study scenario-1	SEN = 0.8511, SPEC = 0.8235, F-measure = 0.8333, Precision = 0.8163, Balanced Accuracy = 0.8373
3D-CNN applying PET data to study scenario-2	SEN = 0.8404, SPEC = 0.8824, F-measure = 0.8541, Precision = 0.8681, Balanced Accuracy = 0.8614
3D-CNN applying PET data to study scenario-3	SEN = 0.8617, SPEC = 0.6961, F-measure = 0.7864, Precision = 0.7232, Balanced Accuracy = 0.7789
3D-CNN applying MRI data to study scenario-1	RCI = 0.2194, CEN = {'AD': 0.7609, 'NC': 0.6752}, Average CEN = 0.71805, IBA = {'AD': 0.5468, 'NC': 0.6291}, Average IBA = 0.5879, GM = {'AD': 0.7668, 'NC': 0.7668}, Average GM = 0.7668, MCC = {'AD': 0.5367, 'NC': 0.5367}, Average MCC = 0.5367
3D-CNN applying MRI data to study scenario-2	RCI = 0.2517, CEN = {'AD': 0.7243, 'NC': 0.6525}, Average CEN = 0.6884, IBA = {'AD': 0.5979, 'NC': 0.6382}, Average IBA = 0.618, GM = {'AD': 0.7861, 'NC': 0.7861}, Average GM = 0.7861, MCC = {'AD': 0.5721, 'NC': 0.5721}, Average MCC = 0.5721
3D-CNN applying MRI data to study scenario-3	RCI = 0.238, CEN = {'AD': 0.7384, 'NC': 0.6643}, Average CEN = 0.7014, IBA = {'AD': 0.5825, 'NC': 0.6297}, Average IBA = 0.6061, GM = {'AD': 0.7785, 'NC': 0.7785}, Average GM = 0.7785, MCC = {'AD': 0.5573, 'NC': 0.5573}, Average MCC = 0.5573

Table 6. Binary classification tasks among NC and MCI classes using PET modality.

Architecture	Performance Metrics
3D-CNN applying PET data to study scenario-1	SEN = 0.5670, SPEC = 0.6471, F-measure = 0.5851, Precision = 0.6044, Balanced Accuracy = 0.6070
3D-CNN applying PET data to study scenario-2	SEN = 0.5567, SPEC = 0.6471, F-measure = 0.5775, Precision = 0.6000, Balanced Accuracy = 0.6019
3D-CNN applying PET data to study scenario-3	SEN = 0.2268, SPEC = 0.8137, F-measure = 0.3188, Precision = 0.5366, Balanced Accuracy = 0.5203
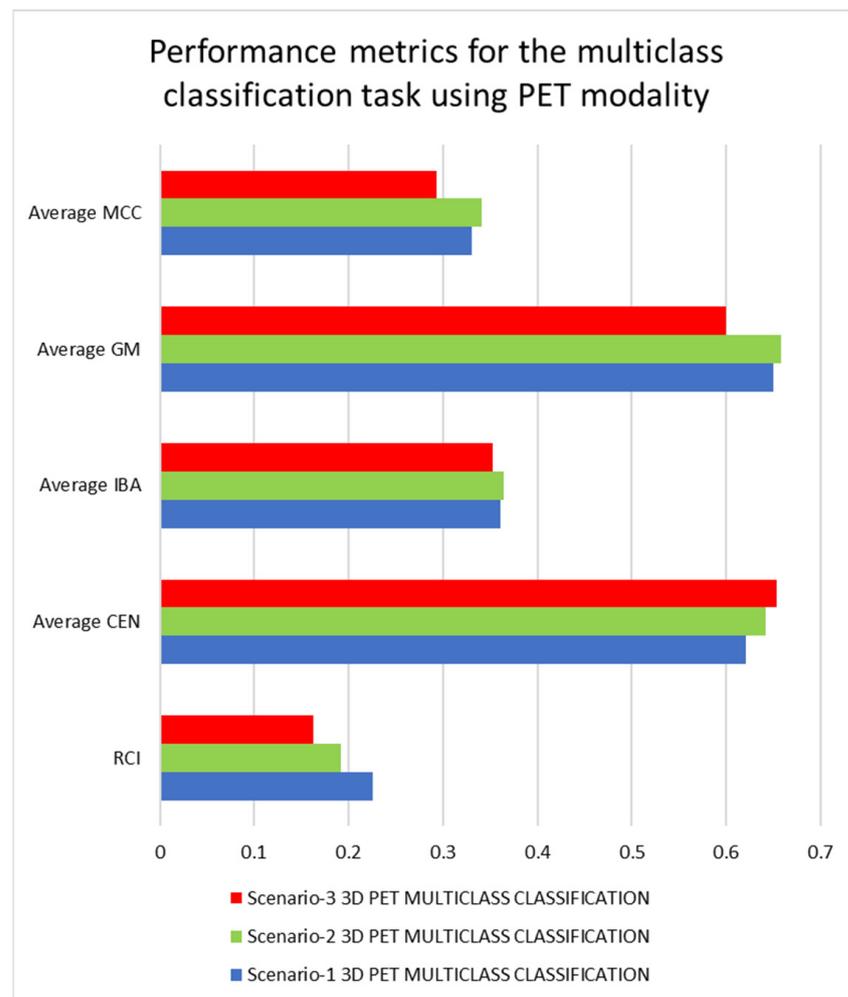
Figure 7. Performance metrics for the multiclass classification tasks under the three scenarios.
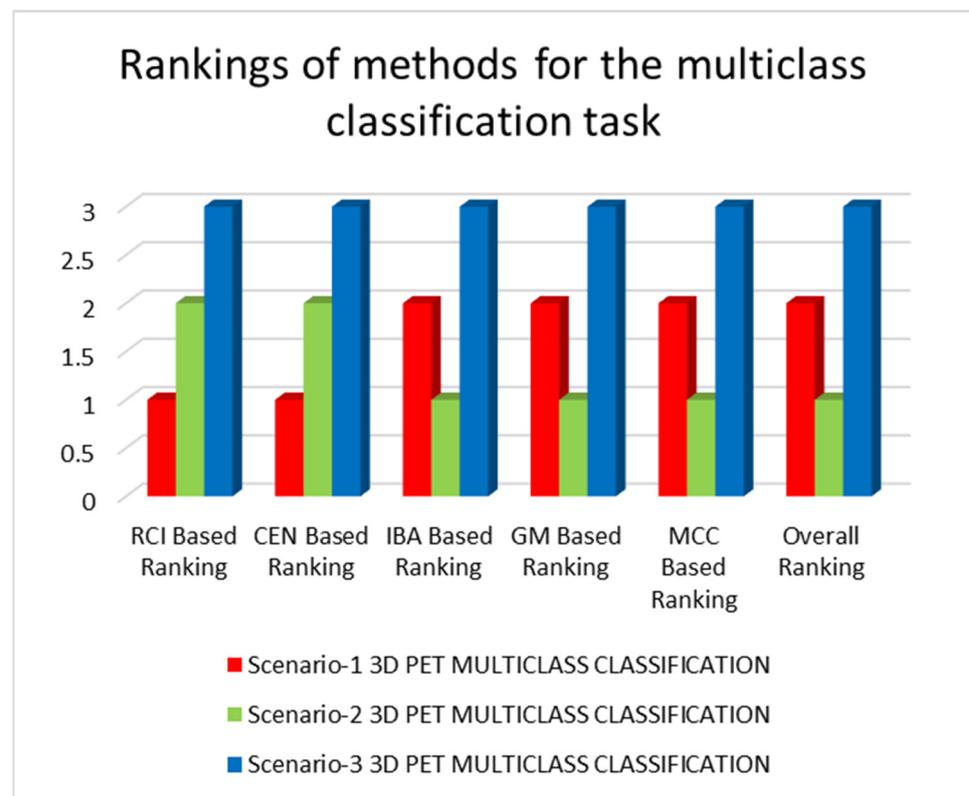
Figure 8. Rankings of methods for the multiclass classification tasks under the three scenarios.

Table 4 illustrates the binary classification outcomes between AD and MCI classes, applying the PET modality for the trained 3D-CNN architectures in each of the three scenarios. The 3D-CNN architecture trained under scenario-3 performed the best, whereas the 3D-CNN architecture trained under scenario-1 performed the poorest while considering only the SEN metric. Overall, the trained 3D-CNN architecture in scenario-2 yielded the best results, whereas the 3D-CNN architecture trained under scenario-3 yielded the worst results. These results illustrate the benefits of combining batch normalization and dropout approaches. A small value of dropout improves the performance; however, large amount of it reduces the performance. The findings and rankings of the architectures trained for the AD vs. MCI binary classification task in each of the three scenarios are shown in Figures 9 and 10.

The statistics for binary classification among AD and NC classes utilizing PET and MRI modalities are summarized in Table 5. Using PET modality, it can be observed that the best performing model is the 3D-CNN architecture trained under scenario-2, 3D-CNN architecture trained under scenario-1 performed the second best while the worst performing architecture is the one trained under scenario-3. Using MRI modality, it can be observed that the best performing model is the 3D-CNN architecture trained under scenario-2, 3D-CNN architecture trained under scenario-3 performed the second best while the worst performing architecture is the one trained under scenario-1. Here again, we can observe that a small value of dropout improves the performance. The findings and rankings of the architectures trained for the AD vs. NC binary classification task in each of the three scenarios are shown in Figures 11–14.
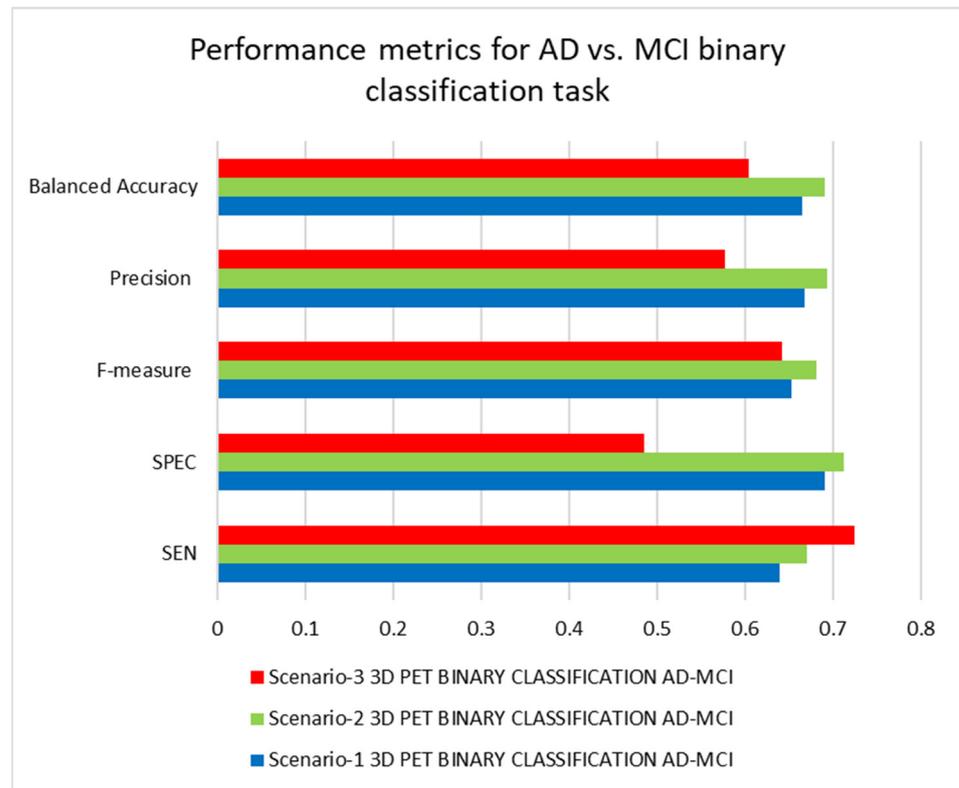
Figure 9. Performance metrics for AD vs. MCI binary classification tasks under the three scenarios.
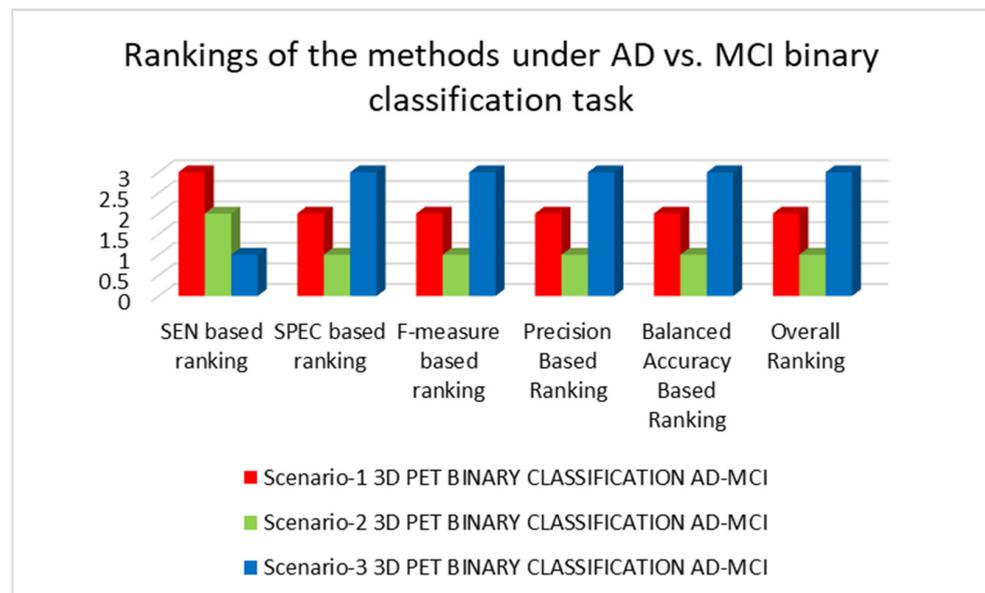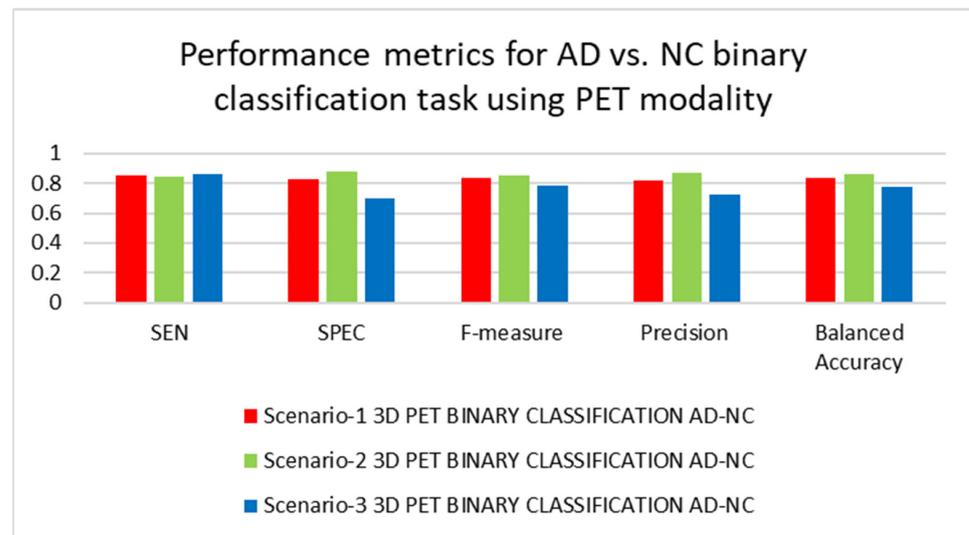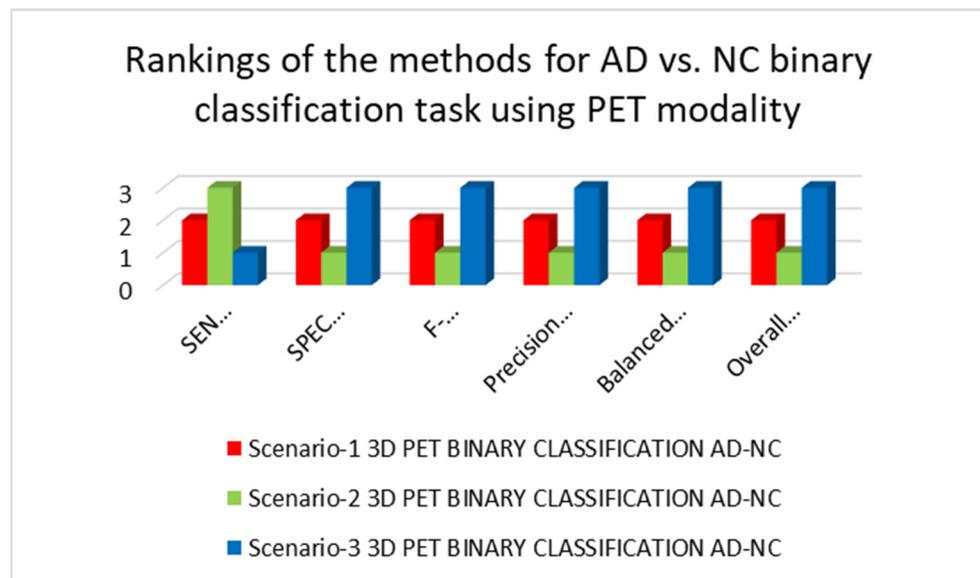

Figure 10. Rankings of the methods for AD vs. MCI binary classification tasks under the three scenarios.
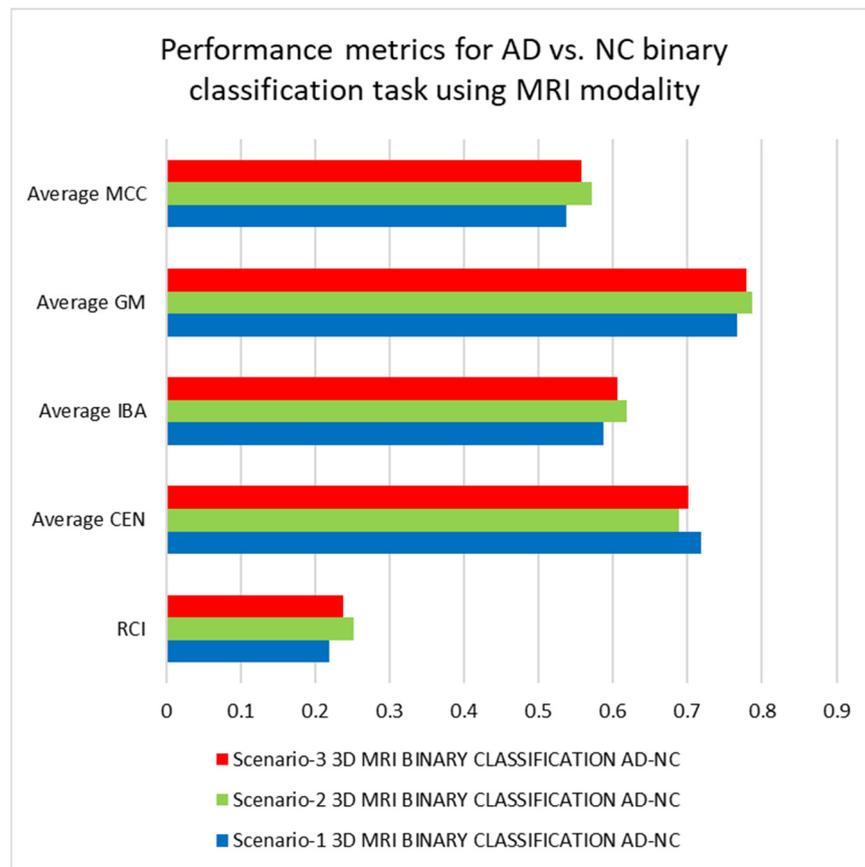
Figure 11. Performance metrics for AD vs. NC binary classification tasks using PET modality under the three scenarios.
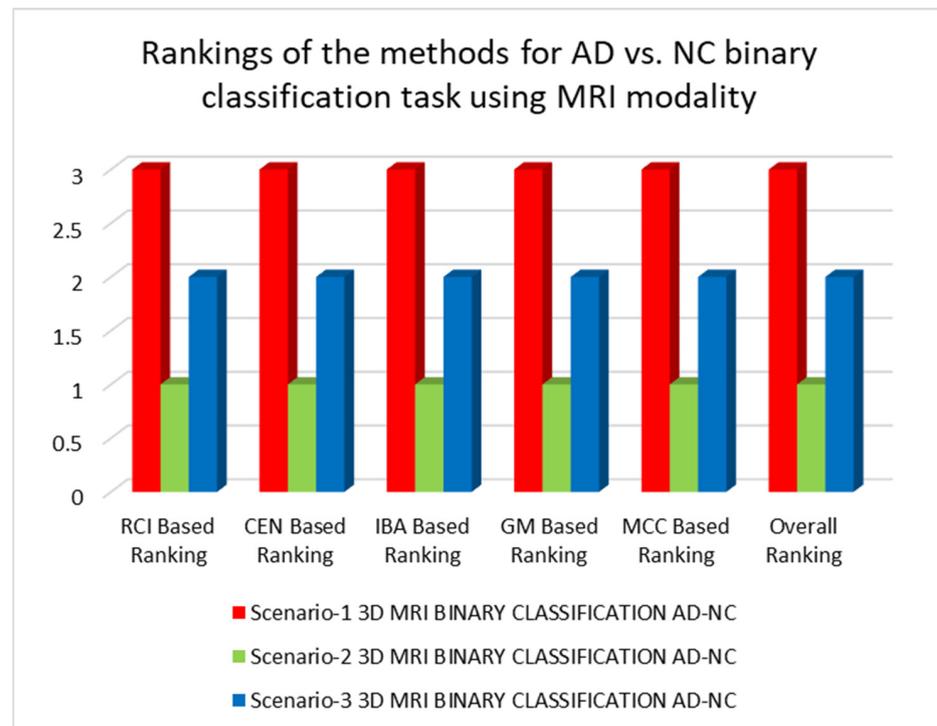

Figure 12. Rankings of the methods for AD vs. NC binary classification tasks using PET modality under the three scenarios.

Table 6 shows the data for binary classification among NC and MCI classes using PET modalities. According to SEN, F-measure, SPEC, and balanced accuracy performance metrics, the 3D-CNN model trained under scenario-1 performed the best, while the 3D-CNN model trained under scenario-3 performed the worst. Overall, the 3D-CNN architecture trained under scenario-1 is the best-performing architecture, whereas the 3D-CNN model trained under scenario-3 is the lowest-performing architecture. We notice an interesting trend here: the design with no dropout performed the best, the one with a small value of dropout performed the second best, and the one with a substantial amount of dropout performed the worst. These data show that utilizing dropout sparingly or not at all leads to better-performing designs than using it excessively. Figures 15 and 16 demonstrate the results and rankings of the architectures trained under the three scenarios for the NC vs. MCI binary classification tasks using PET modalities.
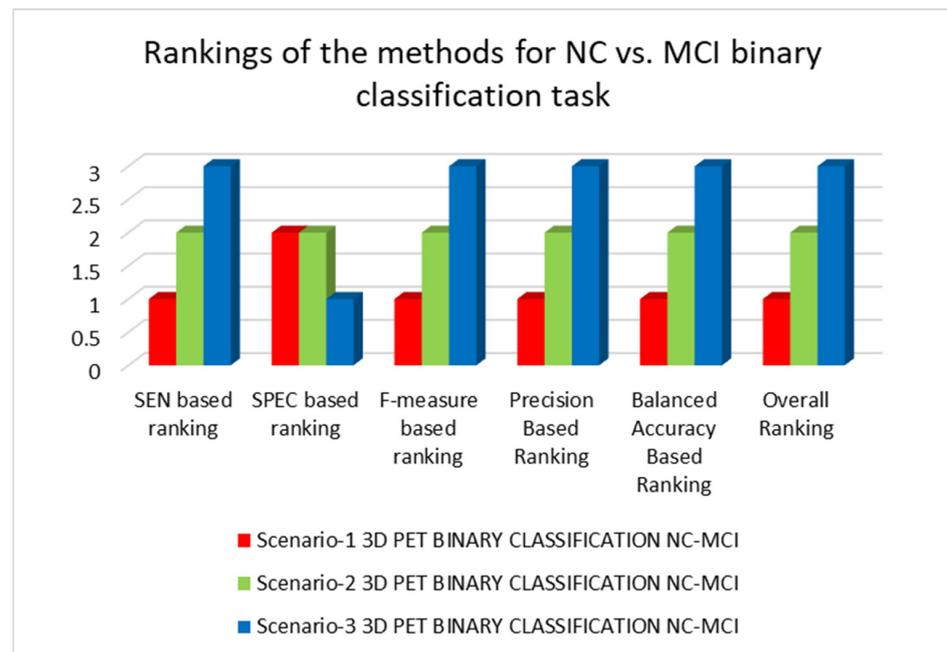
Figure 13. Performance metrics for AD vs. NC binary classification tasks using MRI modality under the three scenarios.

Figure 14. Rankings of the methods for AD vs. NC binary classification tasks using MRI modality under the three scenarios.
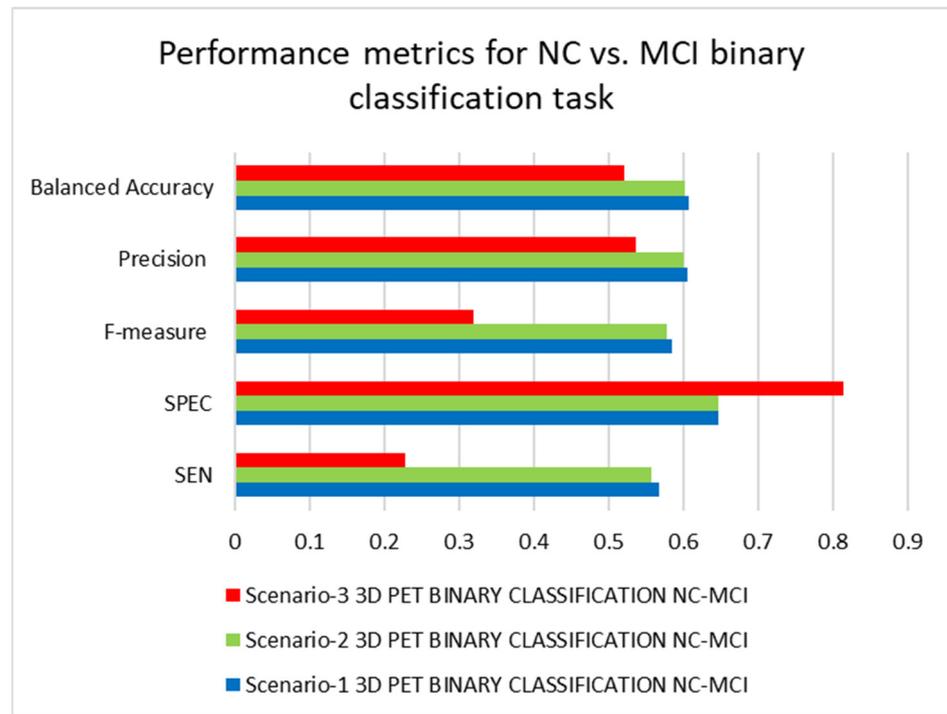
Figure 15. Performance metrics for NC vs. MCI binary classifications under the three scenarios.

Figure 16. Rankings of the methods for NC vs. MCI binary classifications under the three scenarios.

In general, we can observe from Tables 3–6 and Figures 7–16 that the architectures trained under scenario-2 fared the best, followed by the designs trained under scenario-1, while the architectures trained under scenario-3 performed the worst. We discovered that using small amounts of dropout in conjunction with batch normalization resulted in better-performing designs than using it excessively or not at all. In fact, because of the internal covariance shifts among batch normalization and dropout approaches, excessive usage of dropout resulted in bad-performing architectures. As we discovered in our work,

this shift in the variance of neuronal units over the training and testing phases leads to inferior performances across a variety of binary and multiclass classification tasks.

5. Discussion

In modern CNN designs, batch normalization and dropout are two of the most popular elements. They can create highly powerful discriminating or classification structures when they work together, which can help them obtain superior results across different tasks. In the later stages of AD, when a person's cognitive abilities begin to decline, whole brain-based approaches have significant discrimination capacities. We discovered that our 3D-CNN architecture performed worse on MCI vs. NC binary classification tasks than on AD vs. NC and MCI vs. AD binary classification tasks, which might be explained by the usage of whole brain slices. Approaches that use whole brain slices have the benefit of being able to catch changes in the brain that other methods, such as region-of-interest-based methods, may miss [48–54].

Despite using a more complex architecture for MCI vs. NC binary classification than for AD vs. NC binary classification, the comparative scenarios performed worse because whole brain slices failed to capture relevant discriminating cues for this task. Despite using the most sophisticated architecture, the performance on the multiclass classification task was the worst, which could be attributed to the fact that multiclass sorting is a more difficult problem in general than binary classification tasks. Because there are more observable changes inside the brain at this stage, the 3D-CNN design performed best on the AD vs. NC binary classification task. Another noteworthy aspect is that the 3D-CNN architecture trained under scenario-2 performed better than the other two scenarios, highlighting the importance of the requirement for a small value of dropout in the designs.

DL techniques [55,56] for the diagnosis of AD (utilizing multimodal neuroimaging data) are improving and appear to have promise. According to the findings of this study, these methods increased inter-subject variations and reduced intra-subject variations. The construction of individual-specific diagnostic models, as accomplished in this work, is a major focal point for research in AD diagnosis. We can enhance structural information preservation in the brain during processing. Furthermore, we can discover the sharing of common information across many aspects to aid prognosis and diagnosis in clinical applications.

In AD, neuropsychological testing can help measure present and prospective functional skills. Overall cognitive test performances and levels of functional impairment in AD are strongly correlated. Despite the usefulness of scores from neuropsychological batteries, a solitary summary of the test items may oversimplify complex characteristics. Another stumbling block to using DL is the requirement for a large-scale dataset. The benefit of DL algorithms is usually proportional to the sample size. The effectiveness of nonlinear models, such as CNNs, in employing neuropsychological characteristics to guess the practical outcome of AD might be explained by a number of factors. The findings in this paper strongly correlate with neuropsychological factors commonly employed in the assessment of AD, such as age. Despite using small-scale datasets, the impact of these factors on the performance of CNN architectures is profound, which highlights their importance. Furthermore, there is a need to add more factors to AD assessment to increase the chances of an accurate diagnosis in the early stages of AD.

Multimodal data, such as neuropsychological and clinical data, as well as other imaging modalities, may aid the designs in performing even better on classification tasks. Participants in this study may have performed poorly on several binary and multiclass classification tasks due to a lack of these data. Another limitation is the lack of generalization ability, as age is a factor in the order of AD, MCI, and NC subjects. Testing across many databases, such as OASIS, can help alleviate this problem. Because MCI is a phase in between NC and AD with a time period spanning from 0 to 36 months, longitudinal data might improve the classifiers' discrimination performances even further. Table 7 compares the projected approaches to other approaches available in the literature.

Table 7. Comparison with other approaches.

Authors	Data	Methods	Accuracy	Classification Task
Oh et al. [57]	MRI	Inception auto-encoder based CNN	84.51%	AD vs. NC Binary
Ekin et al. [58]	MRI	3D-CNN	73.4%	AD vs. NC Binary
Cosimo Ieracitano et al. [59]	MRI	Electroencephalographic signals	85.78%	AD vs. NC Binary
Rukesh Prajapati et al. [60]	MRI	DL model	85.19%	AD vs. NC Binary
Selene Tomassini et al. [61]	MRI	3D Convolutional Long Short-term Memory-based network	86%	AD vs. NC Binary
Rejusha T R et al. [62]	MRI	Deep convolutional GAN	83%	AD vs. NC Binary
Ekin et al. [63]	MRI	2D-CNN autoencoder	74.66%	AD vs. NC Binary
Ignacio Sarasua et al. [64]	Functional MRI	Template-based DL	77.3%	AD vs. NC Binary
Alex Fedorov et al. [65]	MRI	Multimodal	84.1%	AD vs. NC Binary
Proposed Model (Scenario-2)	**PET**	**3D-CNN Whole brain**	**86.22%**	**AD vs. NC Binary**
Karim A et al. [66]	MRI	2D CNNs hippocampal region	66.5%	AD vs. MCI Binary
Karim A et al. [67]	MRI	2D CNNs coronal, sagittal and axial projections	63.28%	AD vs. MCI Binary
Proposed Model (Scenario-2)	**PET**	**3D-CNN Whole brain**	**69.1%**	**AD vs. MCI Binary**
Tae-Eui K et al. [68]	Resting-state functional MRI	CNN framework	73.85%	NC vs. MCI Binary
Olfa B A et al. [69]	MRI	Circular Harmonic Functions	69.45%	NC vs. MCI Binary
Proposed Model (Scenario-1)	**PET**	**3D-CNN Whole brain**	**60.8%**	**NC vs. MCI Binary**
Bijen K et al. [70]	PET, MRI	DL employing 3D-CNN layers	50.21%	AD vs. NC vs. MCI Multiclass
Eva Y P et al. [71]	MRI	Deep CNN having 3 convolutional layers	55.27%	AD vs. NC vs. MCI Multiclass
Proposed Model (Scenario-2)	**PET**	**3D-CNN Whole brain**	**56.31%**	**AD vs. NC vs. MCI Multiclass**

6. Conclusions

In this paper, we used PET and MRI neuroimaging modalities to train multiple 3D-CNN architectures for binary and multiclass classifications of Alzheimer's disease to investigate the impact of disharmony between dropout and batch normalization techniques on the performance of these architectures. We investigated three different scenarios: (1) training without dropout but with batch normalization, (2) training with a single dropout layer immediately preceding the softmax layer, and (3) training with a single convolutional layer between the dropout and the batch normalization layer. The findings show that low or no dropout in the network leads to higher-performing designs, whereas excessive use of dropout degrades performance. It was discovered that the architecture trained under scenario-2 performed the best, while the architecture trained under scenario-3 performed

the worst. In the future, we plan to investigate this impact in the frequency domain using other data augmentation strategies. We also plan to integrate MRI and PET datasets to improve the performances in a variety of binary and multiclass classification tasks for the initial screening of Alzheimer's disease.

References

1. Dubois, B.; Hampel, H.; Feldman, H.H.; Scheltens, P.; Aisen, P.; Andrieu, S.; Bakardjian, H.; Benali, H.; Bertram, L.; Blennow, K.; et al. Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria. *Alzheimers Dement.* **2016**, *12*, 292–323. [CrossRef] [PubMed]

2. Cao, J.; Hou, J.; Ping, J.; Cai, D. Advances in developing novel therapeutic strategies for Alzheimer's disease. *Mol. Neurodegener.* **2018**, *13*, 64. [PubMed]

3. Slot, R.E.R.; Sikkes, S.A.M.; Berkhof, J.; Brodaty, H.; Buckley, R.; Cavedo, E.; Dardiotis, E.; Guillo-Benarous, F.; Hampel, H.; Kochan, N.A.; et al. Subjective cognitive decline and rates of incident Alzheimer's disease and non-Alzheimer's disease dementia. *Alzheimers Dement.* **2019**, *15*, 465–476. [CrossRef] [PubMed]

4. Tufail, A.B.; Ma, Y.K.; Kaabar, M.K.; Martínez, F.; Junejo, A.R.; Ullah, I.; Khan, R. Deep learning in cancer diagnosis and prognosis prediction: A minireview on challenges, recent trends, and future directions. *Comput. Math. Methods Med.* **2021**, *2021*, 9025470.

5. Khan, R.; Yang, Q.; Ullah, I.; Rehman, A.U.; Tufail, A.B.; Noor, A.; Cengiz, K. 3D convolutional neural networks based automatic modulation classification in the presence of channel noise. *IET Commun.* **2021**, *16*, 497–509. [CrossRef]

6. Tufail, A.B.; Ma, Y.K.; Zhang, Q.N.; Khan, A.; Zhao, L.; Yang, Q.; Ullah, I. 3D convolutional neural networks-based multiclass classification of Alzheimer's and Parkinson's diseases using PET and SPECT neuroimaging modalities. *Brain Inform.* **2021**, *8*, 1–9.

7. Ahmad, I.; Ullah, I.; Khan, W.U.; Rehman, A.U.; Adrees, M.S.; Saleem, M.Q.; Cheikhrouhou, O.; Hamam, H.; Shafiq, M. Efficient Algorithms for E-Healthcare to Solve Multiobject Fuse Detection Problem. *J. Healthc. Eng.* **2021**, *2021*, 9500304. [CrossRef]

8. De Ture, M.A.; Dickson, D.W. The neuropathological diagnosis of Alzheimer's disease. *Mol. Neurodegener.* **2019**, *14*, 32.

9. Mukherjee, S.; Mez, J.; Trittschuh, E.H.; Saykin, A.J.; Gibbons, L.E.; Fardo, D.W.; Wessels, M.; Bauman, J.; Moore, M.; Choi, S.-E.; et al. Genetic data and cognitively-defined late-onset Alzheimer's disease subgroups. *Mol. Psychiatry* **2019**, *25*, 2942–2951. [CrossRef]

10. Edwards, G.A., III; Gamez, N.; Escobedo, G., Jr.; Calderon, O.; Moreno-Gonzalez, I. Modifiable Risk Factors for Alzheimer's Disease. Front. *Aging Neurosci.* **2019**, *11*, 146.

11. Mendez, M.F. Early-onset Alzheimer Disease and Its Variants. *Continuum* **2019**, *25*, 34–51. [CrossRef] [PubMed]

12. Kim, M.; Snowden, S.; Suvitaival, T.; Ali, A.; Merkler, D.J.; Ahmad, T.; Westwood, S.; Baird, A.; Proitsi, P.; Nevado-Holgado, A.; et al. Primary fatty amides in plasma associated with brain amyloid burden, hippocampal volume, and memory in the European Medical Information Framework for Alzheimer's Disease biomarker discovery cohort. *Alzheimer's Dement.* **2019**, *15*, 817–827.

13. Sundas, A.; Badotra, S.; Bharany, S.; Almogren, A.; Tag-ElDin, E.M.; Rehman, A.U. HealthGuard: An Intelligent Healthcare System Security Framework Based on Machine Learning. *Sustainability* **2022**, *14*, 11934. [CrossRef]

14. Serrano-Pozo, A.; Growdon, J.H. Is Alzheimer's Disease Risk Modifiable? *J. Alzheimers Dis.* **2019**, *67*, 795–819. [CrossRef] [PubMed]

15. Rehman, A.U.; Naqvi, R.A.; Rehman, A.; Paul, A.; Sadiq, M.T.; Hussain, D. A Trustworthy SIoT Aware Mechanism as an Enabler for Citizen Services in Smart Cities. *Electronics* **2020**, *9*, 918. [CrossRef]

16. Tufail, A.B.; Anwar, N.; Othman, M.T.; Ullah, I.; Khan, R.A.; Ma, Y.K.; Adhikari, D.; Rehman, A.U.; Shafiq, M.; Hamam, H. Early-stage Alzheimer's disease categorization using PET neuroimaging modality and convolutional neural networks in the 2D and 3D domains. *Sensors* **2022**, *22*, 4609. [CrossRef]

17. Ahmad, S.; Ullah, T.; Ahmad, I.; Al-Sharabi, A.; Ullah, K.; Khan, R.A.; Rasheed, S.; Ullah, I.; Uddin, M.; Ali, M. A novel hybrid deep learning model for metastatic cancer detection. *Comput. Intell. Neurosci.* **2022**, *2022*, 8141530.

18. Basher, A.; Kim, B.C.; Lee, K.H.; Jung, H.Y. Volumetric Feature-Based Alzheimer's Disease Diagnosis From sMRI Data Using a Convolutional Neural Network and a Deep Neural Network. *IEEE Access* **2021**, *9*, 29870–29882. [CrossRef]

19. Martinez-Murcia, F.J.; Ortiz, A.; Gorriz, J.-M.; Ramirez, J.; Castillo-Barnes, D. Studying the Manifold Structure of Alzheimer's Disease: A Deep Learning Approach Using Convolutional Autoencoders. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 17–26. [CrossRef]

20. Murugan, S.; Venkatesan, C.; Sumithra, M.G.; Gao, X.-Z.; Elakkiya, B.; Akila, M.; Manoharan, S. DEMNET: A Deep Learning Model for Early Diagnosis of Alzheimer Diseases and Dementia from MR Images. *IEEE Access* **2021**, *9*, 90319–90329. [CrossRef]

21. Tanveer, M.; Rashid, A.H.; Ganaie, M.A.; Reza, M.; Razzak, I.; Hua, K.-L. Classification of Alzheimer's disease using ensemble of deep neural networks trained through transfer learning. *IEEE J. Biomed. Health Inform.* **2021**, *26*, 1453–1463. [CrossRef]

22. Tufail, A.B.; Ullah, I.; Khan, R.; Ali, L.; Yousaf, A.; Rehman, A.U.; Ma, Y.K. Recognition of Ziziphus lotus through Aerial Imaging and Deep Transfer Learning Approach. *Mob. Inf. Syst.* **2021**, *2021*, 4310321. [CrossRef]

23. Yousafzai, B.K.; Khan, S.A.; Rahman, T.; Khan, I.; Ullah, I.; Ur Rehman, A.; Cheikhrouhou, O. Student-performulator: Student academic performance using hybrid deep neural network. *Sustainability* **2021**, *13*, 9775.

24. Sadiq, M.T.; Yu, X.; Yuan, Z.; Zeming, F.; Rehman, A.U.; Ullah, I.; Li, G.; Xiao, G. Motor imagery EEG signals decoding by multivariate empirical wavelet transform-based framework for robust brain–computer interfaces. *IEEE Access* **2019**, *7*, 171431–171451.

25. Bin Tufail, A.; Ullah, I.; Khan, W.U.; Asif, M.; Ahmad, I.; Ma, Y.-K.; Khan, R.; Kalimullah; Ali, S. Diagnosis of Diabetic Retinopathy through Retinal Fundus Images and 3D Convolutional Neural Networks with Limited Number of Samples. *Wirel. Commun. Mob. Comput.* **2021**, *2021*, 6013448. [CrossRef]

26. Mahendran, N.; PM, D.R. A deep learning framework with an embedded based feature selection approach for the early detection of the Alzheimer's disease. *Comput. Biol. Med.* **2022**, *141*, 105056.

27. Han, R.; Chen, C.L.P.; Liu, Z. A Novel Convolutional Variation of Broad Learning System for Alzheimer's Disease Diagnosis by Using MRI Images. *IEEE Access* **2020**, *8*, 214646–214657. [CrossRef]

28. Lu, B.; Li, H.X.; Chang, Z.K.; Li, L.; Chen, N.X.; Zhu, Z.C.; Zhou, H.X.; Li, X.Y.; Wang, Y.W.; Cui, S.X.; et al. A practical Alzheimer's disease classifier via brain imaging-based deep learning on 85,721 samples. *J. Big Data* **2022**, *9*, 101. [CrossRef]

29. Syed, A.H.; Khan, T.; Hassan, A.; Alromema, N.A.; Binsawad, M.; Alsayed, A.O. An Ensemble-Learning Based Application to Predict the Earlier Stages of Alzheimer's Disease (AD). *IEEE Access* **2020**, *8*, 222126–222143. [CrossRef]

30. Saleem, T.J.; Zahra, S.R.; Wu, F.; Alwakeel, A.; Alwakeel, M.; Jeribi, F.; Hijji, M. Deep Learning-Based Diagnosis of Alzheimer's Disease. *J. Pers. Med.* **2022**, *12*, 815. [CrossRef]

31. Yuan, S.; Li, H.; Wu, J.; Sun, X. Classification of Mild Cognitive Impairment with Multimodal Data using both Labeled and Unlabeled Samples. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2021**, *18*, 2281–2290. [CrossRef] [PubMed]

32. Assam, M.; Kanwal, H.; Farooq, U.; Shah, S.K.; Mehmood, A.; Choi, G.S. An Efficient Classification of MRI Brain Images. *IEEE Access* **2021**, *9*, 33313–33322. [CrossRef]

33. Zhang, F.; Pan, B.; Shao, P.; Liu, P.; Shen, S.; Yao, P.; Xu, R.X. A Single Model Deep Learning Approach for Alzheimer's Disease Diagnosis. *Neuroscience* **2022**, *491*, 200–214. [PubMed]

34. Guo, H.; Zhang, Y. Resting State fMRI and Improved Deep Learning Algorithm for Earlier Detection of Alzheimer's Disease. *IEEE Access* **2020**, *8*, 115383–115392. [CrossRef]

35. Santurkar, S.; Tsipras, D.; Ilyas, A.; Madry, A. How Does Batch Normalization Help Optimization? *arXiv* **2019**, arXiv:1805.11604v5.

36. Bjorck, J.; Gomes, C.; Selman, B.; Weinberger, K.Q. Understanding Batch Normalization. *arXiv* **2018**, arXiv:1806.02375v4.

37. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv* **2015**, arXiv:1502.03167v3.

38. Daneshmand, H.; Joudaki, A.; Bach, F. Batch Normalization orthogonalizes representations in deep random networks. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 4896–4906.

39. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.

40. Labach, A.; Salehinejad, H.; Valaee, S. Survey of Dropout Methods for Deep Neural Networks. *arXiv* **2019**, arXiv:1904.13310v2.

41. Fan, X.; Zhang, S.; Tanwisuth, K.; Qian, X.; Zhou, M. Contextual dropout: An efficient sample dependent dropout module. *arXiv* **2021**, arXiv:2103.04181v1.

42. Xie, J.; Ma, Z.; Lei, J.; Zhang, G.; Xue, J.-H.; Tan, Z.-H.; Guo, J. Advanced Dropout: A Model-free Methodology for Bayesian Dropout Optimization. *arXiv* **2021**, arXiv:2010.05244v2.

43. Li, X.; Chen, S.; Hu, X.; Yang, J. Understanding the Disharmony between Dropout and Batch Normalization by Variance Shift. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 2677–2685. [CrossRef]

44. Weiner, M.W.; Aisen, P.S.; Jack, C.R., Jr.; Jagust, W.J.; Trojanowski, J.Q.; Shaw, L.; Saykin, A.J.; Morris, J.C.; Cairns, N.; Beckett, L.A.; et al. The Alzheimer's disease Neuroimaging Initiative: Progress report and future plans. *Alzheimers. Dement.* **2010**, *6*, 202–211. [CrossRef] [PubMed]

45. 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]

46. Spasov, S.; Passamonti, L.; Duggento, A.; Liò, P.; Toschi, N.; for the Alzheimer's Disease Neuroimaging Initiative. A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease. Neuroimage **2019**, *189*, 276–287. [CrossRef]

47. Tufail, A.B.; Ma, Y.-K.; Kaabar, M.K.A.; Rehman, A.U.; Khan, R.; Cheikhrouhou, O. Classification of Initial Stages of Alzheimer's Disease through Pet Neuroimaging Modality and Deep Learning: Quantifying the Impact of Image Filtering Approaches. *Mathematics* **2021**, *9*, 3101. [CrossRef]

48. 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]

49. Choi, J.Y.; Lee, B. Combining of Multiple Deep Networks via Ensemble Generalization Loss, Based on MRI Images, for Alzheimer's Disease Classification. *IEEE Signal Process. Lett.* **2020**, *27*, 206–210. [CrossRef]

50. Basheer, S.; Bhatia, S.; Sakri, S.B. Computational Modeling of Dementia Prediction Using Deep Neural Network: Analysis on OASIS Dataset. *IEEE Access* **2021**, *9*, 42449–42462. [CrossRef]

51. 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.* **2020**, *42*, 880–893. [CrossRef]

52. Er, F.; Goularas, D. Predicting the Prognosis of MCI Patients Using Longitudinal MRI Data. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2021**, *18*, 1164–1173. [PubMed]

53. Xia, Z.; Zhou, T.; Mamoon, S.; Lu, J. Recognition of Dementia Biomarkers with Deep Finer-DBN. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2021**, *29*, 1926–1935. [CrossRef] [PubMed]

54. Cuingnet, R.; Gerardin, E.; Tessieras, J.; Auzias, G.; Lehéricy, S.; Habert, M.-O.; Chupin, M.; Benali, H.; Colliot, O. 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. [PubMed]

55. Bin Tufail, A.; Ullah, K.; Khan, R.A.; Shakir, M.; Khan, M.A.; Ullah, I.; Ma, Y.-K.; Ali, S. On Improved 3D-CNN-Based Binary and Multiclass Classification of Alzheimer's Disease Using Neuroimaging Modalities and Data Augmentation Methods. *J. Healthc. Eng.* **2022**, *2022*, 1302170. [CrossRef]

56. Aderghal, K.; Afdel, K.; Benois-Pineau, J.; Catheline, G. Improving Alzheimer's stage categorization with Convolutional Neural Network using transfer learning and different magnetic resonance imaging modalities. *Heliyon* **2020**, *6*, e05652. [CrossRef]

57. Oh, K.; Chung, Y.C.; Kim, K.W.; Kim, W.S.; Oh, I.S. Author correction: Classification and visualization of Alzheimer's disease using volumetric convolutional neural network and transfer learning. *Sci. Rep.* **2020**, *10*, 5663.

58. Yagis, E.; Citi, L.; Diciotti, S.; Marzi, C.; Atnafu, S.W.; De Herrera, A.G.S. 3D Convolutional Neural Networks for Diagnosis of Alzheimer's Disease via Structural MRI. In Proceedings of the 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems, CBMS, Rochester, MN, USA, 28–30 July 2020.

59. Ieracitano, C.; Mammone, N.; Hussain, A.; Morabito, F.C. A Convolutional Neural Network based self-learning approach for classifying neurodegenerative states from EEG signals in dementia. In Proceedings of the 2020 International Joint Conference on Neural Networks, IJCNN, Glasgow, UK, 19–24 July 2020.

60. Prajapati, R.; Khatri, U.; Kwon, G.R. An Efficient Deep Neural Network Binary Classifier for Alzheimer's disease Classification. In Proceedings of the 2021 International Conference on Artificial Intelligence in Information and Communication, ICAIIC, Jeju Island, Korea, 20–23 April 2021.

61. Tomassini, S.; Falcionelli, N.; Sernani, P.; Müller, H.; Dragoni, A.F. An End-to-End 3D ConvLSTM-based Framework for Early Diagnosis of Alzheimer's Disease from Full-Resolution Whole-Brain sMRI Scans. In Proceedings of the 2021 IEEE 34th International Symposium on Computer-Based Medical Systems, CBMS, Aveiro, Portugal, 7–9 June 2021.

62. Rejusha, T.R.; Vipin Kumar, K.S. Artificial MRI Image Generation using Deep Convolutional GAN and its Comparison with other Augmentation Methods. In Proceedings of the 2021 International Conference on Communication, Control and Information Sciences, ICCISc, Idukki, India, 16–18 June 2021.

63. Yagis, E.; de Herrera, A.G.S.; Citi, L. Convolutional Autoencoder based Deep Learning Approach for Alzheimer's Disease Diagnosis using Brain MRI. In Proceedings of the 2021 IEEE 34th International Symposium on Computer-Based Medical Systems, CBMS, Aveiro, Portugal, 7–9 June 2021.

64. Sarasua, I.; Lee, J.; Wachinger, C. Geometric Deep Learning on Anatomical Meshes for the Prediction of Alzheimer's disease. In Proceedings of the 2021 IEEE 18th International Symposium on Biomedical Imaging, ISBI, Nice, France, 13–16 April 2021.

65. Kaur, S.; Sharma, S.; Rehman, A.U.; Eldin, E.T.; Ghamry, N.A.; Shafiq, M.; Bharany, S. Predicting Infection Positivity, Risk Estimation, and Disease Prognosis in Dengue Infected Patients by ML Expert System. *Sustainability* **2022**, *14*, 13490. [CrossRef]

66. Aderghal, K.; Boissenin, M.; Benois-Pineau, J.; Catheline, G.; Afdel, K. Classification of sMRI for AD Diagnosis with Convolutional Neuronal Networks: A Pilot 2-D+ε Study on ADNI. In Proceedings of the International Conference on Multimedia Modeling, Reykjavik, Iceland, 4–6 January 2017.

67. Aderghal, K.; Benois-Pineau, J.; Afdel, K. FuseMe: Classification of sMRI images by fusion of Deep CNNs in 2D+ε projections. In Proceedings of the International Workshop on Content-Based Multimedia Indexing, CBMI, Florence, Italy, 19–21 June 2017.

68.   Kam, T.E.; Zhang, H.; Jiao, Z.; Shen, D. Deep Learning of Static and Dynamic Brain Functional Networks for Early MCI Detection. *IEEE Trans. Med. Imaging* **2020**, *39*, 478–487. [CrossRef]

69.   Ben Ahmed, O.; Mizotin, M.; Benois-Pineau, J.; Allard, M.; Catheline, G.; Ben Amar, C. Alzheimer's disease diagnosis on structural MR images using circular harmonic functions descriptors on hippocampus and posterior cingulate cortex. *Comput. Med. Imaging Graph.* **2015**, *44*, 13–25.

70.   Khagi, B.; Kwon, G.-R. 3D CNN Design for the Classification of Alzheimer's Disease Using Brain MRI and PET. *IEEE Access* **2020**, *8*, 217830–217847. [CrossRef]

71.   Puspaningrum, E.Y.; Wahid, R.R.; Amaliyah, R.P. Alzheimer's Disease Stage Classification using Deep Convolutional Neural Networks on Oversampled Imbalance Data. In Proceedings of the 2020 6th Information Technology International Seminar, ITIS, Surabaya, Indonesia, 14–16 October 2020.