



# BG-3DM2F: Bidirectional gated 3D multi-scale feature fusion for Alzheimer's disease diagnosis

Ibtissam Bakkouri<sup>1</sup> · Karim Afdel<sup>1</sup> · Jenny Benois-Pineau<sup>2</sup> · Gwénaëlle Catheline<sup>3</sup>  
For the Alzheimer's Disease Neuroimaging Initiative

Received: 20 December 2020 / Revised: 1 March 2021 / Accepted: 14 January 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

## Abstract

A computer-aided diagnosis system is one of the crucial decision support tools under the medical imaging scope. It has recently emerged as a powerful way to diagnose Alzheimer's Disease (AD) from structural magnetic resonance imaging scans. However, due to the deficit of recognition memory in the Mild Cognitive Impairment (MCI) stage, semantic feature ambiguity, and high inter-class visual similarities problems, computer-aided diagnosis of AD remains challenging. To bridge these gaps, this paper proposed a hippocampus analysis method based on a novel 3D convolutional neural network fusion strategy, called Bidirectional Gated 3D Multi-scale Feature Fusion (BG-3DM2F). The suggested BG-3DM2F framework consists of two modules: 3D Multi-Scale Chained Network (3DMS-ChaineNet) and Bidirectional Gated Recurrent Fusion Unit (Bi-GRFU). The 3DMS-ChaineNet architecture is introduced to design the subtle features and capture the variations in hippocampal atrophy, while the Bi-GRFU scheme is investigated to store 3DMS-ChaineNet levels in the forward and backward fashion and retain them in the decision-making process. For validation, our solution is completely evaluated on the public Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. Practically, we conducted empirical evaluations to verify the effect of BG-3DM2F components. In comparison with the current state-of-the-art methods, the experiments show that our proposed approach provides efficient results, achieving the accuracies of 98.12%, 95.26%, and 96.97% for binary classification of Normal Control (NC) versus AD, AD versus MCI, and NC versus MCI, respectively. Therefore, we can conclude that our proposed BG-3DM2F system has the potential to dramatically improve the conventional classification methods for assisting clinical decision-making.

**Keywords** Alzheimer's disease · 3D convolutional neural network · 3D multi-scale feature fusion · Hippocampal volumes of interest · Bidirectional gated recurrent unit

---

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database ([adni.loni.usc.edu](http://adni.loni.usc.edu)). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: [http://adni.loni.usc.edu/wp-content/uploads/how\\_to\\_protect\\_to\\_protect\\_apply/ADNI\\_protect\\_Acknowledgement\\_protect\\_List.pdf](http://adni.loni.usc.edu/wp-content/uploads/how_to_protect_to_protect_apply/ADNI_protect_Acknowledgement_protect_List.pdf)

---

✉ Ibtissam Bakkouri  
[ibtissam.bakkouri@gmail.com](mailto:ibtissam.bakkouri@gmail.com)

Extended author information available on the last page of the article.

## 1 Introduction

Alzheimer's disease is recently ranked as the most common and progressive form of chronic neurodegenerative dementia that interferes with memory, thinking, behavior, and other mental abilities with the hidden onset and progressive development [10, 90]. It usually starts slowly and gradually worsens over time associated with an ongoing decline of brain functioning [17, 85, 86]. It has risen dramatically during the last few decades due to neuroinflammation, oxidative stress, brain glucose hypometabolism, axonal and synaptic failures, insulin resistance, family history and lifestyle factor which are involved in the pathogenesis and progression of AD [16, 94]. Currently, neuroimaging researchers found that the gray matter volume in the bilateral medial temporal lobe declined notably in the AD group compared with the normal control patients [38, 41]. Further, an in-depth analysis of AD biomarkers reveals that the hippocampus contributes more significantly, which plays a crucial role in mediating declarative memory [20, 21, 25, 75, 80]. Hippocampal volume reduction and atrophy are noticeable features for the detection of MCI and AD [88]. Accordingly, some AD research studies suggested that the deterioration of the hippocampal structure, function, and biochemistry could predict the transformation of MCI to AD, which has high relevance in the diagnosis of AD [55, 108]. Thus, the volumetric hippocampal degeneration might be employed as a monitor for the progression of the disease [67]. Moreover, according to recent researches and studies, it has been reported that the Structural Magnetic Resonance Imaging (sMRI) modality is a very powerful tool in biomedical visualization and clinical research which gives complementary details in images of Alzheimer's disease with high resolution of anatomical structures, simple operation, and low cost that are more informative and suitable for visual perception or computer analysis [70, 81]. A variety of studies have been conducted for diagnosing AD or evaluating its progression by analyzing the sMRI data [2, 8, 91]. For these reasons, the quantitative analysis of the hippocampal Volumes Of Interest (VOI) using sMRI scans is of great significance in clinical to better assimilate the inter-individual anatomical variability in the structural and functional brain changes [2, 8, 12].

With the fast advance of technology in clinical decision making, the Computer-Aided Diagnosis (CAD) system has received significant attention over the recent years owing to the ever-increasing number of patients diagnosed with different diseases, which provides the second opinion and great assistance to radiologists to improve their confidence in accepting the results [6]. It has become one of the major research subjects in medical imaging and diagnostic radiology [7]. It is noticeable that the performance of a CAD system crucially depends on the feature representation and class discrimination [1]. Recently, building a thoroughly effective feature representation map without losing significant details and useful information is considered one of the key aspects in inter-class separation [72–74], and it heavily brought significant improvement in the diagnosis of neurodegenerative diseases [5, 15] and several medical research fields [69]. These successes of CAD system are benefited from using feature fusion in Convolutional Neural Network (CNN) architectures which is becoming the state-of-the-art method for several computer vision problems including image classification [4], object localization [110], and image segmentation [101]. Feature fusion solution is able of generating discriminative features from multi-level CNN to form gradually abstract descriptors, replacing the traditional approach of a single CNN model for feature map representation. Unfortunately, when dealing with volumetric neuroimaging data, we notice that the computation required for Three-dimensional Convolutional Neural Network (3D-CNN) is very complex in the volumetric domain [82]. The sparsity in the 3D

volumetric data can still introduce many redundant weights in the network. This can absolutely cause overfitting issues and lead to a huge number of parameters, in the end, [87, 100, 109]. Then, how can we alleviate such weaknesses?

In this paper, we present an empirical investigation of this question. One of the contributions of our work is to explore the application of 3D multi-scale feature fusion to evaluate the hippocampal volume for AD recognition using the volumetric sMRI modality. In fact, the hippocampus in mild (MCI) and severe (AD) stages is characterized by complex morphological structures like shape with convex outside borders and undefined edges, margin irregularity, and texture heterogeneity [11]. In addition, due to the high similarity of visual characteristics of the hippocampal atrophy patterns in MCI and AD in some cases, it is difficult to find a method able to provide a satisfactory recognition performance and to face the challenges pointed out above. To enable the feasibility of training multi-level 3D-CNN with alleviating of the semantic feature ambiguity and high inter-class visual similarities problems, a bidirectional sequence model [63, 105] was presented to control information flow extracted from each level of 3D-CNN and enhance the memory ability without suffering from vanishing gradient issue. This hierarchical fusion structure tends to perform well in maximally exploiting the input data and improving classification accuracy. Despite the popularity of 3D-CNNs and sequence models, there has only been preliminary research on hierarchical multi-scale feature fusion networks for Alzheimer's disease analysis [8]. Thus, the use of multi-level 3D-CNN architectures combined with sequence memories to this issue for different computer vision techniques such as classification, detection, and segmentation has yet to be fully investigated. To explore gaps in the topics of feature fusion in 3D-CNN architectures and bidirectional sequence model, this paper makes the following major contributions for AD recognition:

1. It introduces a new feature fusion paradigm in 3D-CNN, called BG-3DM2F, for AD recognition which is considered as one of the most important tasks in the CAD system. The BG-3DM2F framework provides a complementary decision tool and powerful joint representation for hippocampal atrophy in an efficient way to classify AD against NC (AD vs. NC) and MCI (AD vs. MCI), and NC against MCI (NC vs. MCI) with perfectly handling the high computational cost, overfitting, vanishing gradient and high inter-class visual similarity issues.
2. It designs an efficient 3D Multi-Scale Chained Network (3DMS-ChaineNet) architecture based on merging of multiple three-dimensional feature scales in order to strongly capture both the intrinsic and extrinsic semantic information of hippocampal volume and, as a consequence, to formulate a robust CAD system for AD recognition.
3. It proposes a simple but effective Bidirectional Gated Recurrent Fusion Unit (Bi-GRFU) approach to control 3DMS-ChaineNet information flow in a parallel way and boost the memory ability of the model in a bidirectional manner. The proposed Bi-GRFU contributes to better 3D multi-scale feature fusion, and it provides an inherent superiority for the BG-3DM2F system to achieve better performance, especially for AD vs. MCI classification task.
4. It presents a series of comparative experiments and an empirical evaluation analysis in order to select the optimal network structure and verify the effectiveness of different components of our proposed BG-3DM2F framework.

The proposed framework has been validated on the ADNI dataset [46], which is designed for AD diagnosis and has been compared with the state-of-the-art hippocampal atrophy classification approaches. The performance evaluation analysis shows the potential clinical

value of the proposed BG-3DM2F framework. To the best of our knowledge, no prior work exists on hippocampal volume characterization based on enhanced 3D multi-scale feature fusion and bidirectional sequence model with the empirical experimental analysis that is investigated here.

The remainder of this paper is organized as follows: In Section 2, we describe the most important methods applied to recognize Alzheimer's disease. In Section 3, we detail the proposed approach. Then, we present the results of experiments realized on the ADNI dataset in Section 4 and we discuss in detail the results of our proposed method in Section 5. Finally, we conclude this paper in Section 6.

## 2 Related works

Over the last few years, 3D-CNN has been extensively studied in the field of computer vision with high model performance and several interesting findings [57, 104]. In the context of Alzheimer's disease research, 3D-CNN has begun to achieve excellent machine learning performance on pattern recognition and diagnosis tasks. Currently, it attracts significant interest from many researchers in the field, and many studies have been developed to perform AD analysis and help radiologists give an accurate diagnosis. Inspired by the residual learning model, Karasawa et al. [49] proposed a 3D Convolutional Residual Network with 39 layers (3D-ResNet-39) which automatically learns feature representation from MRI images for diagnosis of AD. This method was applied to whole-brain images and evaluated on the ADNI dataset, and the empirical analysis indicates that the proposed 3D-ResNet-39 reached promising results. Tang et al. [92] developed a unified architecture named 3D fine-tuning convolutional neural network (3D-FCNN). The algorithm used the whole 3D MRI blocks as input to train a 3D convolutional neural network and then finetuned model parameters on the validation data. In the end, it can discriminate between diseased (AD) and healthy (NC) brains straightly. The experiments on the ADNI dataset indicate that the proposed 3D-FCNN model is superior to conventional classifiers both in accuracy and robustness. A Voxel-based 3D Convolutional Neural Network (V3D-CNN) was investigated in [32] to account for all voxels in the brain and capture the subtle local brain details with better pronounced global specifics of MRI images. The experiments on the ADNI dataset displayed important performance enhancements compared to several baseline and previous studies. In [22], a Cascaded 3D Convolutional Neural Networks (C3D-CNNs) was proposed to learn the multi-level imaging features for classifications of AD vs. NC using Positron Emission Tomography (PET) brain images. Conducting the experimental results and comparison on the ADNI database demonstrate the performance improvement of the proposed method for AD diagnosis. Yan et al. [103] suggested a 3D Residual Neural Network with 152 layers (3D-ResNet-152) for AD classification. They applied Conditional Generative Adversarial Networks (CGAN) to the generation of PET images from corresponding MRI images. The authors found that their contribution (3D-ResNet-152+CGAN) improved the performance of the network on the ADNI dataset. Inspired by independent component analysis (ICA), Kam et al. [48] introduced a novel 3D-CNN framework that directly works on the Brain Functional Networks (BFNs) abstracted by ICA. The proposed BFN+3D-CNN framework can automatically and deeply learn complex, high-level, and hierarchical diagnostic features from resting-state functional Magnetic Resonance Imaging (rs-fMRI) images. The results on a public ADNI dataset showed the effectiveness of the proposed framework for the early mild cognitive impairment diagnosis. To inherit the main drawbacks from the

manual feature extraction procedure, Shmulev et al. [89] modeled a 3D Embedding Residual Neural Network (3DE-ResNet) for diagnosis of AD using brain MRI ADNI dataset. The use of learned representation from the deep embedding allowed to increase the quality of prediction and yield significant performance. A 3D Deeply Supervised Adaptable CNN (3D-DSA-CNN) was investigated by Hosseini et al. [39] for early diagnosis of AD. They found that the 3D-CNN with fine-tuning reports better classification performance and alleviates the effects of overfitting. The experiments on the ADNI MRI and CADDementia datasets without skull stripping preprocessing showed that the proposed 3D-DSA-CNN outperforms previously proposed approaches. In this context, a similar approach based on the Residual and Plain 3D-CNN (ResNet+P3D-CNN) architectures was adopted in [53] for AD diagnosis to tackle with the complicated multi-stage pipelines for the hand-crafted feature extraction and conventional classification algorithms such as support vector machines or logistic regression. The authors evaluated the performance of the proposed framework based on the whole brain images stemming from the ADNI project that provides a dataset of sMRI scans, and they showed that applying ResNet+P3D-CNN to the sMRI classification problem yielded results comparable to previously used approaches. Additionally, Wang et al. [98] exploited Ensemble of 3D Densely Connected CNN (E3DC-CNN) for diagnosis of mild cognitive impairment and AD. E3DC-CNN can extract the most prominent features of dementia disease and capture brain anatomical variations from whole MRI scans. The experiments reported that the dense connections heavily drew significant improvement in the performance of 3D networks on the ADNI dataset. Lately, Region of Interest (ROI)-based approaches are broadly used due to the spatial dimension reduction and ability of the whole brain representation. Cui and Liu [26] offered an ambitious suggestion to find an apparent AD recognition solution by proposing an approach based on Deep 3D-CNN (D3D-CNN). They argue that hippocampal atrophy is a validated neurodegeneration biomarker of AD. Therefore, the bilateral hippocampal volume was used to train D3D-CNN, allowing the extraction of hierarchical features for each hippocampus. The proposed method was evaluated on 3D T1-weighted sMRI images extracted from the ADNI dataset, demonstrating the promising classification performance in the accuracy of AD diagnosis. The same idea has been explored in [27], by introducing a robust hippocampus analysis technique. To enhance AD classification, Cui and Liu combined 3D Densely Connected Convolutional Networks (3D-DenseNet) with Traditional Shape Features (TSF) which can capture complementary information. This jointly learning method was trained using sMRI images from the ADNI dataset to evaluate its performance.

Quite recently, RNN methods were incorporated with 3D-CNN to fulfill better pattern recognition performance [77, 93, 96]. In the Alzheimer's disease recognition context, LSTM and GRU schemes have currently been addressed by only a few research works. In [8], a computer-aided diagnosis system for AD was proposed. It was generated using the combination between 3D Multi-scale Feature (3DMF) blocks and Gated Recurrent Fusion Unit (GRFU) using hippocampal volumes as input. This method was evaluated on the sMRI ADNI dataset and the experimental analysis indicates that the proposed 3DMF+GRFU outperforms the conventional CAD methods. In a similar study, Li and Liu [56] evaluated the performance of 3D DenseNets (3D-DenseNets) to gradually learn the intensity and shape features of internal and external hippocampal patches. Additionally, the Bidirectional Gate Recurrent Unit (Bi-GRU), which is an extension of traditional Bidirectional Recurrent Neural Networks (Bi-RNN) is also investigated to build a hybrid deep learning model for AD diagnosis by capturing the feature similarity between the left and right hippocampus. The proposed method was evaluated with the baseline MRI images extracted

from the ADNI dataset, achieving better performance than the conventional volume and shape analysis methods. Another study performed in the hippocampal area has been suggested by Huang et al. [42]. They introduced a new framework based on 3D VGG-Variant CNN (3DVGG-CNN) to combine complementary features from T1-weighted MRI and Fluorodeoxyglucose Positron-Emission Tomography (FDG-PET) modalities. The proposed method has the potential to capture the metabolic activity of the tissues, give high-resolution information for the brain structure, and precisely evaluate anatomical metrics from MRI and FDG-PET scans like thickness, volume, and shape. The classification of Alzheimer's disease was carried out on the ADNI dataset, providing satisfactory performance. To fence this section, two recent issues similar to [56] based on the bidirectional sequence model have been suggested for AD diagnosis in [34, 35]. The first study explored the effect of 3D-CNN combined with Bi-RNN to obtain the deep feature representation from MRI and PET images while the second approach is focused on the virtues of 3D-CNN and Fully Stacked Bidirectional Long Short-Term Memory (FSBi-LSTM) architectures to capture deep feature representation from both MRI and PET modalities. The authors estimate that using a bidirectional sequence model can provide bidirectional information that includes previous and future contextual features to improve the predictions of the current point. These methods, 3D-CNN+Bi-RNN and 3D-CNN+FSBi-LSTM, were validated on the ADNI dataset, demonstrating the promising classification performance.

As a matter of fact, 3D-CNN is a crucial factor to effectively explore the feature domain and enhance performance on a visual recognition task. Irrespective of existing research studies that were previously mentioned and indicated in Table 1, the investigation of 3D-CNN on Alzheimer's disease research field is still relatively scarce. In addition, despite the high achievement of 3D-CNN for AD diagnosis, it needs some improvement concerning the systematical analysis of the factors that affect the multiple feature fusion in 3D-CNN without

**Table 1** Overview of the related works focusing on 3D-CNN for Alzheimer's disease recognition

Literature	Year	Modality	Method
Cheng et al. [22]	2017	PET	C3D-CNNs
Korolev et al. [53]	2017	sMRI	ResNet+P3D-CNN
Karasawa et al. [49]	2018	MRI	3D-ResNet-39
Tang et al. [92]	2018	MRI	3D-FCNN
Esmailzadeh et al. [32]	2018	MRI	V3D-CNN
Yan et al. [103]	2018	PET	3D-ResNet-152+CGAN
Kam et al. [48]	2018	rs-fMRI	BFN+3D-CNN
Shmulev et al. [89]	2018	MRI	3DE-ResNet
Hosseini et al. [39]	2018	MRI	3DDSA-CNN
Cui and Liu [26]	2018	sMRI	D3D-CNN
Feng et al. [34]	2018	MRI+PET	3D-CNN+Bi-RNN
Wang et al. [98]	2019	MRI	E3DC-CNN
Li and Liu [56]	2019	MRI	3D-DenseNets+Bi-GRU
Huang et al. [42]	2019	MRI+FDG-PET	3DVGG-CNN
Feng et al. [35]	2019	MRI+PET	3D-CNN+FSBi-LSTM
Bakkouri et al. [8]	2019	sMRI	3DMF+GRFU
Cui and Liu [27]	2019	sMRI	3D-DenseNet+TSF

**Table 2** Summary of related works classified by cerebral region

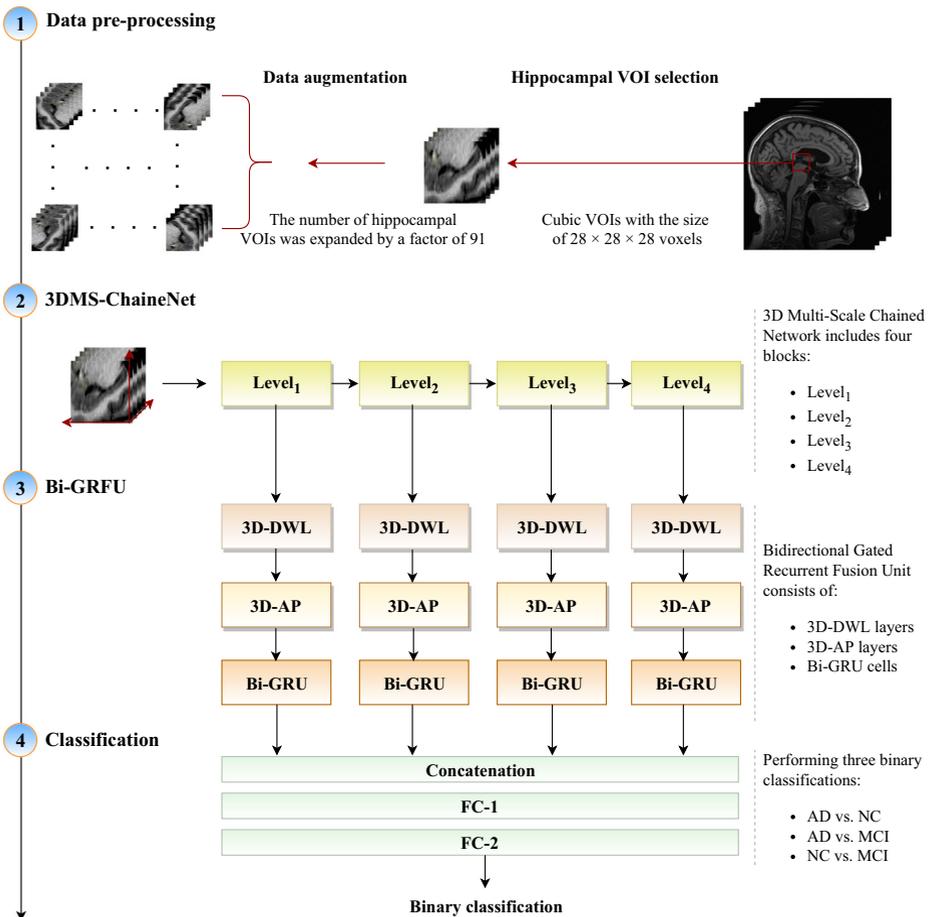
Cerebral region	Literature
	Karasawa et al. [49]
	Tang et al. [92]
	Esmailzadeh et al. [32]
	Cheng et al. [22]
	Yan et al. [103]
Whole brain	Kam et al. [48]
	Shmulev et al. [89]
	Hosseini et al. [39]
	Korolev et al. [53]
	Wang et al. [98]
	Feng et al. [34]
	Feng et al. [35]
	Cui and Liu [26]
Li and Liu [56]	
Hippocampal area	Huang et al. [42]
	Bakkouri et al. [8]
	Cui and Liu [27]

precisely harming the classification accuracy. Moreover, as indicated in Table 2, the main 3D-CNN approaches were carried out on full brain volumes rather than volumes of interest. Although recent studies suggest that the deterioration of hippocampal structure, function, and biochemistry has high relevance in the diagnosis of AD, the most current investigations of the 3D-CNN for AD diagnosis were still performed on the whole brain volume instead of hippocampal VOI, which require high cost of computation, memory allocation, and effort. In view of the scarce existing studies on the AD diagnosis domain, a detailed plan to overcome the problem of high computational costs, semantic feature ambiguity, and high inter-class visual similarities between the classes, especially AD and MCI, will be described in a fair amount of detail.

### 3 Proposed BG-3DM2F framework

With the rapid progress of artificial intelligence in diagnosing diseases, a CAD system is being effectively used in AD recognition to emulate radiologist cognition in improving AD diagnosis with the fast interpretation and making of outcome prediction and therapeutic decisions. Until now, the existing CAD methods used to benefit a lot from a single 3D-CNN feature map representation to capture AD biomarkers which always fail due to semantic feature ambiguity and high-level similarities between classes, especially AD and MCI. Aside from the gaps previously mentioned, investigating only one 3D-CNN pipeline for AD recognition is not good at handling complicated volumetric data for performing multiple classification analyses. Accordingly, the development of a powerful CAD system for AD diagnosis has crucially relied on the development of an intelligent system based on the fusion of multiple 3D-CNN features. In this paper, we proposed a novel CAD system, called BG-3DM2F, based on multi-scale 3D-CNN architecture combined with Bi-GRU

cells to carry out three binary classification problems: AD vs. NC, AD vs. MCI, and NC vs. MCI using the volumetric sMRI modality. The proposed process was developed in several stages. First, to improve model building time, we took the hippocampal volumes as input with a smaller size rather than the large-scale whole brain. To boost the performance of our model and avoid overfitting, data augmentation techniques were applied to all selected VOI for expanding the size of our dataset. Then, we built 3DMS-ChaineNet blocks to extract the most significant features. The extracted features from the last layer of each 3DMS-ChaineNet block were fed to Bi-GRU cells and preserved by applying Bi-GRFU architecture to save multi-scale feature abstraction, avoid vanishing gradient problem and embed each 3DMS-ChaineNet level into a low dimensional semantic space with enhancing the ability to capture long-term dependencies in the sequence learning model. Finally, the



**Fig. 1** The overview of the proposed BG-3DM2F framework. The process starts from Top to bottom. The BG-3DM2F system was organized into different stages including data pre-processing, 3D Multi-Scale Chained Network (3DMS-ChaineNet), Bidirectional Gated Recurrent Fusion Unit (Bi-GRFU), and classification, providing a completely integrated workflow to potentiate the achievement of CAD for AD. VOI stands for the volume of interest for the hippocampus. 3D-DWL and 3D-AP represent 3D depth-wise layer and 3D average pooling, respectively. Bi-GRU is the bidirectional gated recurrent unit while FC-1 and FC-2 are the fully connected layers

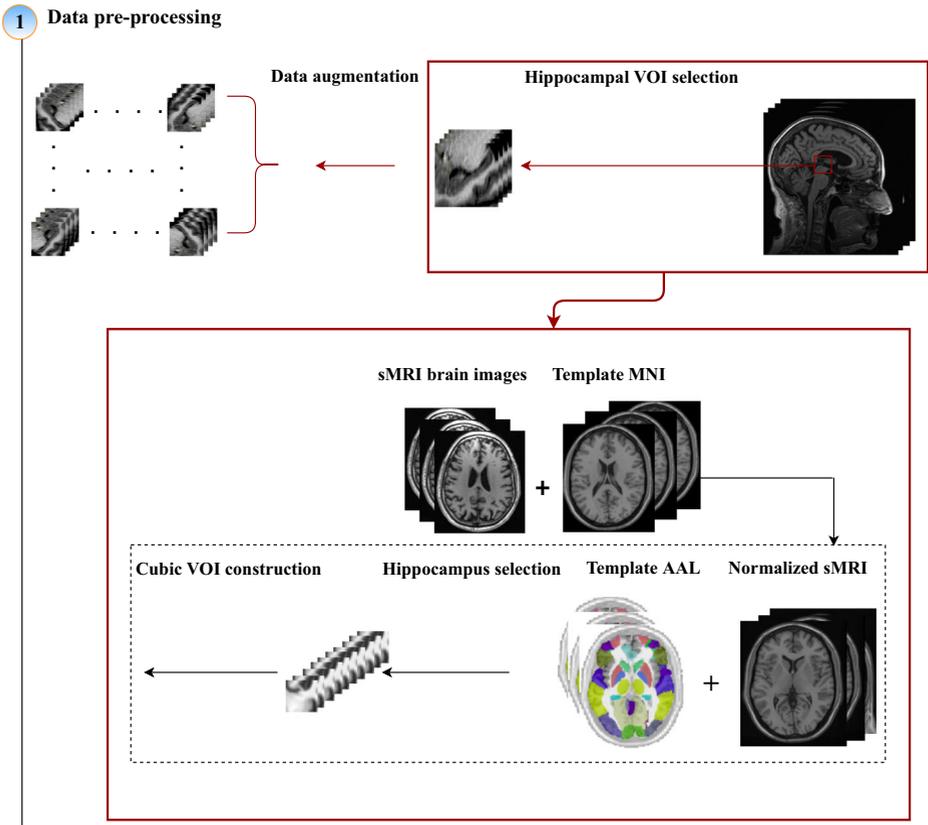
binary classification was performed using two fully connected layers (FC-1 and FC-2). An overview of the proposed approach is given in Fig. 1.

### 3.1 Hippocampal VOI extraction

Biological human brains differ from their size and shape [12]. Such morphological variation of brains leads to difficulties in neuroimaging studies in how to establish a spatial correspondence between 3D volumetric images for clinical applications like multimodal information fusion, comparison of between-subjects in a common coordinate space, stereotactic neurosurgery, and analysis of focal morphological differences to study human brain functions. The academic researchers are largely focusing on creating a reference brain template in a Cartesian coordinate system for a three-dimensional space as a common space for various cerebral subjects to align to [79]. In this study, to overcome the issue of brain shape and size variabilities from different subjects, a robust transformation of the original sMRI brain images to a common neuroanatomical space was involved [84]. It is done by spatial normalization through the affine registration procedure as mentioned in [8, 12–14] to the Montreal Neurological Institute (MNI-152) template [36], which is defined by the average of 152 normal MRI scans, without loss of cortical details and impacting on local image features. This spatial normalization process perfectly overcomes the inter-subject anatomical differences in shape, size, and relative orientation. Thus, images from the ADNI database were resized to  $[121 \times 145 \times 121]$  voxels with isotropic voxel sizes of  $1.5 \times 1.5 \times 1.5 \text{ mm}^3$ . Note that the voxels should be isotropic to allow the data to be rotated, re-sliced, and manipulated. Based on the Automated Anatomical Labelling (AAL) parcellation [83], we extracted the bilaterally symmetrical hippocampus from coronal and axial views and a single hippocampal volume from sagittal slices. To yield a hippocampal region representing the coronal and axial views for further analysis purposes, the average region of interest of the left and right hippocampus was calculated to get one region of interest per slice. In this way, the VOIs of dimension  $[28 \times 28 \times 121]$  were obtained. The spatial normalization and box extraction were fully automated. Since input data should not limit the use of 3D-CNN ideally, we constructed cubic hippocampal volumes representing various regions of interest of hippocampal margin and shape characteristics with the size of  $[28 \times 28 \times 28]$  voxels as indicated in [8] where the 28 slices were selected from the center of each 3D image volume. Finally, we obtained cubic VOI with a size of  $[28 \times 28 \times 28]$  to build 3D-CNN without requiring extremely long training time and high computational costs. An overview of the proposed hippocampal VOI extraction approach is given in Fig. 2.

### 3.2 Data augmentation

The main power of a 3D-CNN lies in its deep architecture, which allows for extracting a set of discriminating visual features at multiple levels of abstraction [82]. Since they have different kinds of volumetric layers, and each layer works differently than the other for the extraction of visual features, 3D-CNN needs to be equipped with a large amount of dataset to yield strong generalization ability [57, 104]. Unfortunately, as the number of the original 3D sMRI images is limited, 3D-CNNs are quite difficult to train due to the overfitting issue which considered as the main risk. To cope with this challenge, one solution is to create fake data and add them to the dataset by applying data augmentation techniques. This process artificially extends the size of samples by transforming them and generating extra images as well as balance their number in each class, intending to enhance the effectiveness of 3D-CNN. In this paper, we propose label-preserving transformations as a data-space solution



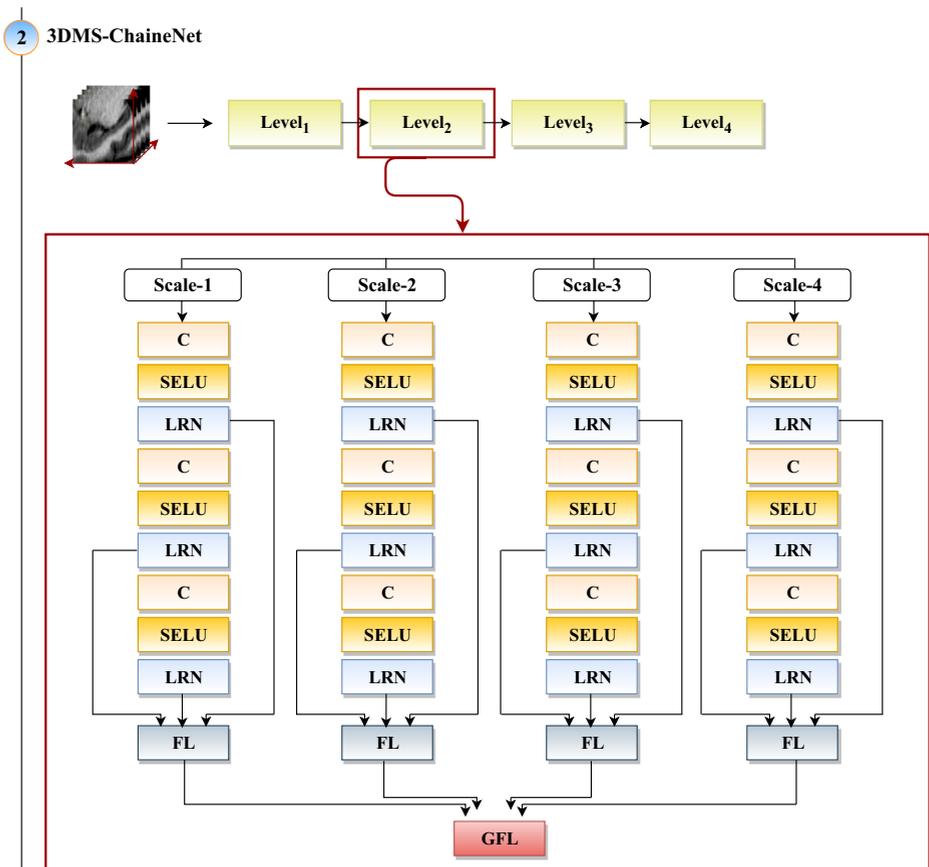
**Fig. 2** The overview of the hippocampal volume of interest selection workflow. The outline of the hippocampal VOI extraction process is composed of three main stages: spatial normalization of sMRI data to the Montreal Neurological Institute (MNI) space, bilateral hippocampal VOIs extraction using Automated Anatomical Labeling (AAL) method, and calculation of the average VOI of the two symmetric hippocampal volumes

to the problem of limited data which can lead to significant improvements in the 3D-CNN performance and make invariant predictions. The process of applying geometric transformation on each 3D image of a dataset is now a widely used approach and helps to augment the proportion of data for the training stability. As revealed in our previous works [6, 8], this strategy is based on rotating, translating, and reflecting the existing images [33]. Each extracted VOI was randomly rotated by 10 selected angles  $\theta$  in the interval  $0 \leq \theta \leq 360$ . The rotated VOIs were mirrored and translated by four values of horizontal and vertical pixel displacements. Resulting from this data augmentation process, the number of hippocampal VOIs was expanded by a factor of 91. This proposed approach enriches the available training data and consequently enhances the performance of the BG-3DM2F framework.

### 3.3 3D multi-scale chained network (3DMS-ChaineNet)

Recently, with the astonishing development of Two-Dimensional CNN (2D-CNN), state-of-the-art multi-scale 2D-CNN features-based pattern recognition frameworks have shown

remarkable strengths in automatically learning the specific features and substantially outperform traditional CNN and hand-crafted descriptors [43, 66, 107]. It is currently widely employed for image classification [18, 44], object localization [24], image super-resolution [30], and image segmentation [19]. Unfortunately, when dealing with 3D-CNN architectures, existing works do not pay attention to exploring the benefits of multi-scale 3D-CNN features for improving pattern recognition in volumetric data. Given the scarce existing studies on this domain, we investigated the benefit of the 3D multi-scale chained network for AD recognition. The proposed 3DMS-ChaineNet method was evaluated in four blocks denoted by  $Level_n$ , where  $n$  is the number of the 3D multi-scale chained blocks. As shown in Fig. 3, The constructed 3DMS-ChaineNet blocks, named  $Level_1$ ,  $Level_2$ ,  $Level_3$  and  $Level_4$ , are built in a similar structure. Each 3DMS-ChaineNet level has four parallel 3D-CNN scale paths, so that each of them consists of Convolutional (C) layers, Scaled Exponential Linear Unit (SELU), Local Response Normalization (LRN) layers, Fusion Layers (FL), and Global Fusion Layers (GFL).



**Fig. 3** The flowchart of the proposed 3DMS-ChaineNet method. The proposed 3DMS-ChaineNet method consists of four blocks, which are denoted by  $Level_1$ ,  $Level_2$ ,  $Level_3$  and  $Level_4$ , structured in similar way. Each level has four parallel 3D-CNN scales similar in architectures but different in convolutional kernel sizes and LRN, FL, and GFL parameters. The multiple kernel scales were performed to focus on low and high level localized features for coping with the complex morphological structures of hippocampal tissues

The 3D convolutional layer (C) is the core building component of a 3D-CNN. It consists of several kernels whose elements need to be trained for extremely reacting to a local area of the input. Specifically, the output of previous layers is convolved with multiple learned 3D weight matrices using the small receptive field. Then, the result is processed by an exponential linear operation to generate normalized output. To extract feature maps of  $C_i$ , the exponential linear function is applied to each element of  $C_{i-1}$ . The output of the 3D convolutional layer at position  $(x, y, z)$  is given as:

$$C_{ij}^{xyz} = SELU \left( b_{ij} + \sum_p \sum_{l=0}^{L_i-1} \sum_{m=0}^{M_i-1} \sum_{n=0}^{N_i-1} W_{ijp}^{lmn} \times C_{(i-1)p}^{(x+l)(y+m)(z+n)} \right) \tag{1}$$

Where  $i$  and  $j$  are two indexes of the layer and feature map, respectively.  $(L_i, M_i, N_i)$  is the 3D kernel size of  $i^{th}$  layer.  $W_{ijp}^{lmn}$  represents the weight of  $(l, m, n)^{th}$  position to the  $p^{th}$  feature map.  $b_{ij}$  denotes bias or offset of the  $j^{th}$  feature map at the  $i^{th}$  3D convolution layer, and SELU stands for scaled exponential linear unit activation function. It is firstly proposed by Klambauer et al. [51] in 2017 to regularize the high variance of the activations and avoid exploding and vanishing gradients. As shown below, SELU function gives an output  $a$  if  $a$  is positive and  $\alpha \times (exp(a) - 1)$  otherwise:

$$SELU(a) = \lambda \begin{cases} a & a \geq 0 \\ \alpha \times (exp(a) - 1) & \text{otherwise} \end{cases} \tag{2}$$

Where  $\lambda$  and  $\alpha$  are two fixed parameters. In this paper, SELU function was used with  $\lambda=1.0507$  and  $\alpha=1.6732$ .

Since the 3D-CNN model usually faces the challenge of unbounded activations, the local response normalization, LRN, is a better alternative to deal with this problem. LRN is firstly proposed by Krizhevsky et al. [54] in 2012 as a regularizer mechanism to diminish the test error of the AlexNet model. In this paper, the LRN was incorporated with the SELU function to improve the convergence speed and AD recognition performance by encouraging the competition for unrestrained activities among groups of neurons at the same location, but in different activation maps. Denoting by  $C^{xyz}$  the activity of a neuron computed by applying weight  $W$  at position  $(x, y, z)$  and the SELU exponential linearity, The LRN is expressed as follow:

$$LRN^{xyz} = C^{xyz} / \left( \lambda_1 + \lambda_2 \times \sum_{j=\max(0, i-\frac{z}{2})}^{\min(T-1, i+\frac{z}{2})} (C_j^{xyz})^2 \right)^{\lambda_3} \tag{3}$$

Where  $\lambda_1, \lambda_2$ , and  $\lambda_3$  are three fixed hyper-parameters.  $C^{xyz}$  represents the 3D convolutional layer output after applying the SELU function at position  $(x, y, z)$  in the feature map,  $T$  stands for a total number of weights, and  $z$  refers to the size of the normalization neighborhood.

Since the hippocampal VOIs are extremely complex and often show significant shape variations, especially in the MCI stage, a single 3D-CNN feature map is not greatly representative. Therefore, to generate latent representations for AD diagnosis, an innovative discriminant criterion using the 3D multi-scale feature fusion approach was investigated. To perform this task, each 3DMS-ChaineNet block was built by inserting the fusion layers, FLs, which enable the automatic fusion of the local response normalization layers. The FL is expressed as follow:

$$FL^{xyz} = \varphi_1 \times LRN_1^{xyz} \odot \varphi_2 \times LRN_2^{xyz} \odot \varphi_3 \times LRN_3^{xyz} \tag{4}$$

**Table 3** Structure parameters of 3DMS-ChaineNet block

Module	Layer	Kernel shape	Specifications
Scale-1	C	$[3 \times 3 \times 3] \times 128$	(P=1, S=1)
	LRN	-	$(\lambda_1 = 2, \lambda_2 = 10^{-4}, \lambda_3 = 0.7)$
	FL	-	$(\varphi_1 = 0.80, \varphi_2 = 0.75, \varphi_3 = 0.30)$
Scale-2	C	$[5 \times 5 \times 5] \times 128$	(P=2, S=1)
	LRN	-	$(\lambda_1 = 1, \lambda_2 = 10^{-5}, \lambda_3 = 0.5)$
	FL	-	$(\varphi_1 = 0.75, \varphi_2 = 0.50, \varphi_3 = 0.25)$
Scale-3	C	$[7 \times 7 \times 7] \times 128$	(P=3, S=1)
	LRN	-	$(\lambda_1 = 2, \lambda_2 = 10^{-4}, \lambda_3 = 0.75)$
	FL	-	$(\varphi_1 = 0.25, \varphi_2 = 0.20, \varphi_3 = 0.15)$
Scale-4	C	$[9 \times 9 \times 9] \times 128$	(P=4, S=1)
	LRN	-	$(\lambda_1 = 1, \lambda_2 = 10^{-4}, \lambda_3 = 0.5)$
	FL	-	$(\varphi_1 = 0.15, \varphi_2 = 0.25, \varphi_3 = 0.15)$
Global fusion	GFL	-	$(\eta_1 = 0.8, \eta_2 = 0.95, \eta_3 = 0.85, \eta_4 = 0.55)$

Where  $\odot$  is element-wise product,  $\varphi_1, \varphi_2$ , and  $\varphi_3$  are three weights. Each element of  $LRN_1, LRN_2$  and  $LRN_3$  is multiplied by  $\varphi_1, \varphi_2$  and  $\varphi_3$ , respectively.

For the (GFL), the last layer of each 3DMS-ChaineNet block can be represented as:

$$GFL^{xyz} = \eta_1 \times FL_1^{xyz} \odot \eta_2 \times FL_2^{xyz} \odot \eta_3 \times FL_3^{xyz} \odot \eta_4 \times FL_4^{xyz} \quad (5)$$

Where  $\eta_1, \eta_2, \eta_3$ , and  $\eta_4$  are four weights of  $FL_1, FL_2, FL_3$ , and  $FL_4$ , respectively.

In all 3DMS-ChaineNet blocks, we use a multi-scale feature approach as the base architecture. In each 3DMS-ChaineNet block, four scales of convolutional layers with size of  $[3 \times 3 \times 3], [5 \times 5 \times 5], [7 \times 7 \times 7]$  and  $[9 \times 9 \times 9]$ , SELU activation function, LRN, and element-wise multiplication layers (FL) are involved. Within each 3D-CNN stream, the three convolutional layers are completely connected by summing the outputs of preceding layers as the inputs of the Fusion Layer (FL). The four parallel 3D-CNN paths were connected via Global Fusion Layer (GFL) which took all the previous FL layers as the input. The arrangement and description of the 3DMS-ChaineNet block are shown in Fig. 3. The constructed 3DMS-ChaineNet levels were built in a similar structure and by the same parameters, while the four parallel 3D-CNN scales of each 3DMS-ChaineNet block were implemented in a similar structure and built by the same number of kernels (n) and stride (S), and they differ essentially from the kernel shape  $[width \times height \times volume] \times n$ , padding (P), LRN and FL parameters. The 3D multi-scale chained network blocks are randomly initialized using the Xavier weight initialization algorithm [62] and trained with the sMRI dataset. The Table 3 shows the parameters specification of 3DMS-ChaineNet block.

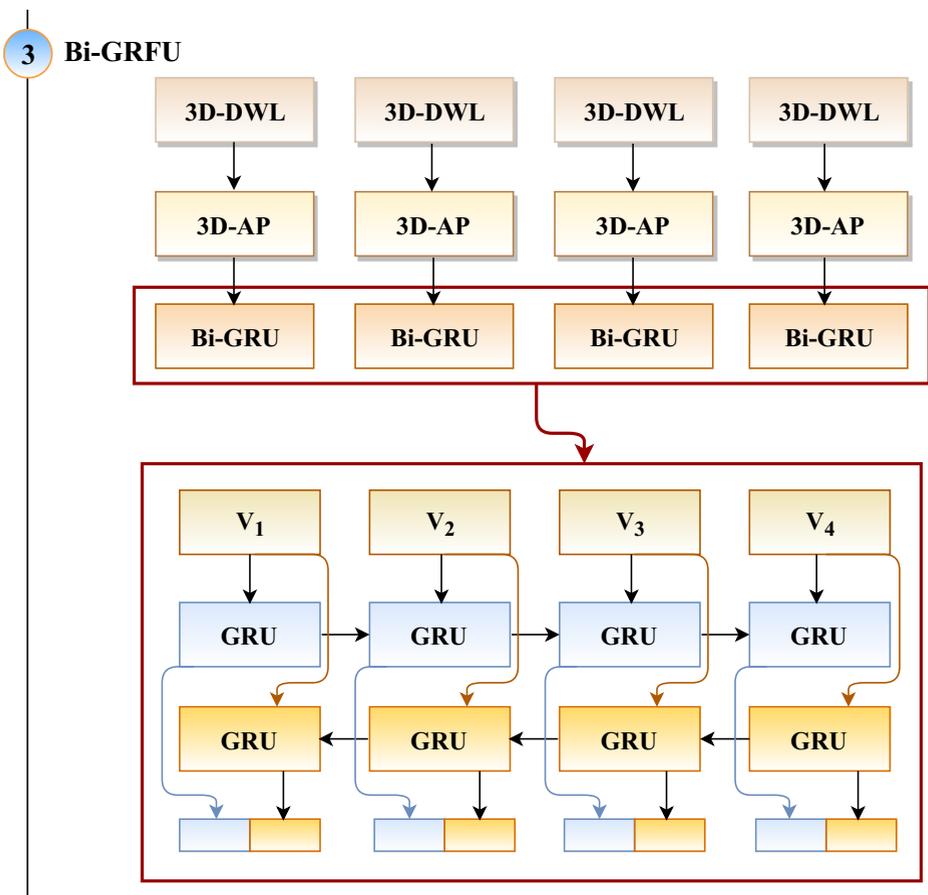
### 3.4 Bidirectional gated recurrent fusion unit (Bi-GRFU)

As introduced in the above sections, GRU is the most recently emerged type of RNN for sequential data analysis [64]. It is proved to be more powerful in the different fields of research applications and widely used as an alternative to the LSTM model [29]. To store hierarchical 3D multi-level features, we design a novel sequential model based on Bi-GRU [111] and marked as Bi-GRFU. Our proposed Bi-GRFU consists of 3D Depth-Wise Layer (3D-DWL), 3D Average Pooling (3D-AP), flatten layer, and Bi-GRU cells to embed each

3DMS-ChaineNet level into a low dimensional semantic space without losing relevant features of hippocampal VOI. An overview of the proposed Bi-GRFU approach is given in Fig. 4.

As the number of feature maps often increases with the depth of the network, the 3D Depth-Wise Layer (3D-DWL) [97], which is a linear projection layer, was used to reduce the number of parameters in 3D convolutions. Then, to decrease the dimension depth-wise, we performed one-to-one convolution with a kernel shape of  $[1 \times 1 \times 1] \times 32$ , which is regarded as a compression layer with a compression factor of 0.25.

3D Pooling module is a significant building block of 3D-CNN. The main idea regarding the application of the pooling component is to progressively down-sample an input representation with highlighting the more informative features, alleviating the drawback of 3D-CNN, and improving the structure information. The major property of the 3D pooling



**Fig. 4** The flowchart of the proposed Bi-GRFU method. This pipeline illustrates the low dimensional semantic space embedding and bidirectional sequence model phases. The low dimensional semantic space embedding was carried out using 3D average pooling and 3D depth-wise layers to cope with high computational costs, while Bi-GRU was proposed as a bidirectional sequence model to mitigate vanishing gradient problem and control each 3DMS-ChaineNet level stream. The vectors  $V_1$ ,  $V_2$ ,  $V_3$  and  $V_4$  represent flatten layers used to reshape 3D-AD matrices to have one dimensional maps

layer is to decrease the spatial resolution of the feature map without requiring a large number of resources and aggravating classification performance. Aside from the reduction of computational burdens, the pooling layer can maintain translation invariance. In this paper, we adopt a 3D Average Pooling (3D-AP) approach [40] which is a concept in 3D-CNN that calculates the average of the elements in each pooling region to obtain a condensed 3D feature map. Formally, the 3D Average Pooling (3D-AP) at pixel  $(x, y, z)$  with a filter size of  $L \times M \times N$  is expressed as:

$$3D - AP(x, y, z) = \frac{1}{L \times M \times N} \times \sum_{l=0}^{L-1} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I(x+l, y+m, z+n) \quad (6)$$

Where  $I$  is the input of pooling kernel window and  $I \in \mathbf{R}^{L \times M \times N}$ . In our case, we used 3D-AP with stride  $S = 2$  and kernel size  $2 \times 2 \times 2$ .

In order to deal with long-term dependencies in practice with fewer parameters and less computational time, GRU was investigated in this study. It can be expressed as the following four steps:

$$z_t = \delta(W_z \times h_{t-1} + W_z \times x_t + b_z) \quad (7)$$

$$r_t = \delta(W_r \times h_{t-1} + W_r \times x_t + b_r) \quad (8)$$

$$h'_t = \tanh(W_h \times (r_t \odot h_{t-1}) + W_h \times x_t + b_h) \quad (9)$$

$$\vec{h}_t = (1 - z_t) \odot h_{t-1} + z_t \odot h'_t \quad (10)$$

Where  $\odot$  is an element wise multiplication. The terms  $x_t, z_t, r_t, h'_t$  and  $\vec{h}_t$  are, respectively, the input sequence vector, the update gate, the reset gate, the new candidate state and the output state at current time  $t$ .  $\delta$  is the logistic sigmoid function and  $\tanh(\cdot)$  represents hyperbolic tangent function.  $W_z, W_r$  and  $W_h$  are the weighted matrices and  $b_z, b_r$  and  $b_h$  are the biases of GRU to be learned during training. The update gate  $z_t$  controls how much old information is kept and how much new information is added, whilst the reset gate  $r_t$  controls the size of contribution of past state to the candidate state. When  $r_t$  equals zero, it forgets the past state. Thus, the output representation of the forward GRU unit is denoted by  $\vec{h}_t$  and the output representation of the backward GRU unit is denoted by  $\overleftarrow{h}_t$ . The output representation of Bi-GRU is denoted by  $h_t = [\vec{h}_t : \overleftarrow{h}_t]$  where the operation  $(:)$  denotes the concatenation of the vectors. We apply the Bi-GRU which consists of a forward and backward GRU to provide complete previous and future contextual information for each point in the output layer, improve memory ability, and mitigate the vanishing gradients problem. After Bi-GRU processing, the forward and backward outputs are concatenated as the input of the first fully connected layer. The concatenated vector contains bidirectional

**Table 4** Structure parameters of the proposed BG-3DM2F architecture using hippocampal VOI as input

Module	Layer	Output shape
Input	Input	$[28 \times 28 \times 28] \times 1$
3DMS-ChaineNet	Level <sub><i>n</i></sub>	$[28 \times 28 \times 28] \times 128$
	3D-DWL	$[28 \times 28 \times 28] \times 32$
Bi-GRFU	3D-AP	$[14 \times 14 \times 14] \times 32$
	GRU	87808
	Concatenation	702464
Classification	FC-1	87808
	FC-2	2

information of the entire hippocampal volume which conveys more information to improve the representation. The structure parameters of the proposed BG-3DM2F framework are listed in Table 4.

## 4 Experiment analysis and evaluation

At this stage, the experimental results of the BG-3DM2F framework, which were pointed out in Section 3 to this paper, are shown and explained precisely and in exquisite detail. To perform three binary classifications, namely AD vs. NC, AD vs. MCI, and NC vs. MCI, we confirmed and validated the proposed BG-3DM2F system by iteratively training the network according to the number of 3DMS-ChaineNet blocks ( $Level_n$ ) and the Bi-RNN variants: Bi-LSTM and Bi-GRU. We demonstrate the robustness of the methodology on the ADNI dataset. To prove the efficiency of our proposed strategy, statistical measurements and quantitative evaluation are carried out. Our approach was compared with the most recent existing methods for AD recognition. All these methods were tested on the same dataset (ADNI) and compared using the same metrics (Accuracy, sensitivity, and specificity).

### 4.1 Data acquisition

The data used in this study were collected from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database [46] ([adni.loni.usc.edu](http://adni.loni.usc.edu)). The ADNI was launched in 2003 as a public-private partnership led by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies, and non-profit organizations. ADNI includes several biomarkers of Alzheimer's disease beyond neuropsychological tests, such as cerebrospinal fluid, structural and functional magnetic resonance imaging, Positron Emission Tomography, and other biological data. Ethical approval for the ADNI study was obtained by the ADNI investigators. For up-to-date information, see [www.adni-info.org](http://www.adni-info.org). In this paper, we used structural MRI data of patients with Alzheimer's disease (AD), normal controls (NC), and mild cognitive impairment (MCI) from ADNI-1 which is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. Particularly, the dataset was extracted from the ADNI-1 with 1.5T scanners that it contains 1075 T1-weighted sMRI images, comprising 188 AD, 401 MCI, and 229 NC images. The main demographic characteristics of 818 subjects are described in Table 5, including detailed information of NC, MCI, and AD classes such as the number of participants of each gender, mean age, and Mini-Mental State Examination (MMSE) score. As shown in Table 5, there is a limited number of brain scans of subjects that were annotated by specialized radiologists. To overcome this limitation, we applied data augmentation techniques to enlarge the dataset as described in Section 3.2. As indicated by Table 6, we obtained a

**Table 5** Demographic description of the ADNI-1 dataset group

Classes	Subjects	Age	Gender (M/F)	MMSE
AD	188	$75.36 \pm 7.5$	99/89	$23.28 \pm 2.0$
MCI	401	$74.85 \pm 7.4$	258/143	$27.01 \pm 1.8$
NC	229	$75.97 \pm 5.0$	119/110	$29.00 \pm 1.0$

**Table 6** A detailed description of the representative hippocampal VOIs divided into training, validation, and testing subsets

	AD	MCI	NC	Total
Tain	10260	10260	10260	30780
Validation	3420	3420	3420	10260
Test	3420	3420	3420	10260
Total	17100	17100	17100	51300

total amount of 51300 VOIs. The typical data set proportions are 60% for training, 20% for validation and 20% for testing [71, 102].

## 4.2 Implementation details

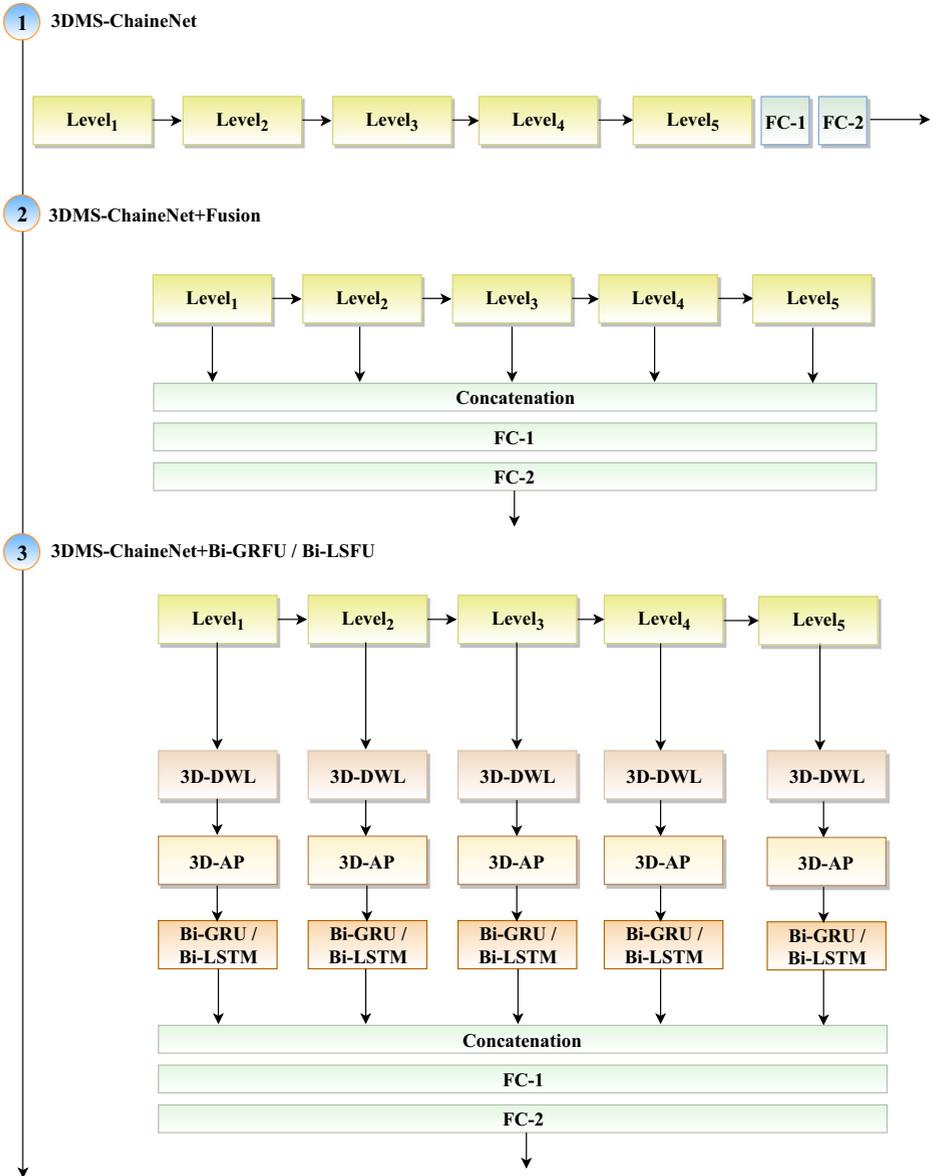
The hippocampal VOI selection was performed with the SPM12 program and MATLAB software with the MarsBaR toolbox [95]. The proposed method has been developed using the Keras library with python wrapper, and Compute Unified Device Architecture (CUDA) enabled a parallel computing platform to access the computational resources of the Graphics Processing Unit (GPU). The available hardware, used for training, is a PC with Intel Xeon E5-1603 v4 Quad-Core, 2.80 GHz processor base frequency, 10 MB cache, 16 GB RAM, and a single NVIDIA GeForce GTX 1080 Ti with 11 GB memory. The performance of a deep CNN model depends critically on its structure and the network configuration [61]. In this paper, there still remain a number of network hyper-parameters to be defined. Many of these hyper-parameters were chosen by experimenting until the model began to converge effectively. The proposed BG-3DM2F system was trained using Adaptive Moment Estimation (Adam) algorithm [50] with learning rate  $\alpha = 10^{-3}$ , first moment-decay  $\beta_1 = 0.9$  and second moment-decay  $\beta_2 = 0.999$  using Xavier algorithm for weight initialization and SELU activation function inserted after each convolutional layer, where the softmax activation function was combined with the cross-entropy for training the network. The optimization ran for 200 epochs with mini-batch size 32. The network was regularized using the dropout technique [9] with factor  $\rho = 0.25$  and L2 regularization [58] with  $\epsilon = 0.05$  as it penalized 5% of the current weight value to prevent the overfitting issues. As indicated in Table 7, to carry out the three binary classification problems (AD vs. NC, AD vs. MCI, and NC vs. MCI), the training took, on average, about 12 hours, 51 minutes, and 7.2 seconds. For testing, on average, it takes 4 hours, 33 minutes, and 54 seconds. The evaluation of the performance of our proposed system was carried out using two quantitative methods including graphical and statistical analysis as detailed in our previous works [6, 7]. The first method for evaluating our proposed system is the Receiver Operating Characteristic (ROC) curve, while the second consists of using six assessment metrics: accuracy (Acc), sensitivity (Sen), specificity (Spe), precision (Pre), F1-Score (FSc), and Area Under the ROC Curve (AUC).

**Table 7** The computational time in hours (h) required by the BG-3DM2F framework to recognize AD

Tasks	Training time (h)	Testing time (h)
AD vs. NC	12.794	4.562
AD vs. MCI	12.863	4.568
NC vs. MCI	12.899	4.566
Average	12.852	4.565

### 4.3 Ablation study

In this section, a series of comparative experiments are conducted to select the optimal network structure. As shown in Fig. 5, to verify the feasibility of the BG-3DM2F framework, the experiments are mainly divided into four parts, and each part was evaluated in



**Fig. 5** The overview of detailed ablation studies to investigate the efficiency gains obtained by the various components of the BG-3DM2F framework and better demonstrate the improvements obtained by 3DMS-ChaineNet, 3DMS-ChaineNet+Fusion, 3DMS-ChaineNet+Bi-LSFU, and 3DMS-ChaineNet+Bi-GRFU models at each level

**Table 8** Quantitative result of different 3DMS-ChaineNet levels with and without fusion and sequence models on the testing subset in terms of accuracy, sensitivity, specificity, precision, F1-Score, and AUC metrics

Methods	Level <sub>n</sub>	Tasks	Acc. (%)	Sen. (%)	Spe. (%)	Pre. (%)	FSc. (%)	AUC (%)
3DMS-ChaineNet	Level <sub>2</sub>	AD vs. NC	80.39	81.87	78.91	79.52	80.68	76.16
		AD vs. MCI	76.12	74.73	77.51	76.87	75.78	75.29
		NC vs. MCI	76.79	75.90	77.69	77.28	76.58	76.42
	Level <sub>3</sub>	AD vs. NC	80.90	<b>88.24</b>	73.56	76.95	82.21	76.59
		AD vs. MCI	75.81	76.37	75.26	75.53	75.95	76.14
		NC vs. MCI	76.37	74.47	<b>78.27</b>	<b>77.41</b>	75.91	75.38
	Level <sub>4</sub>	AD vs. NC	81.35	79.59	<b>83.12</b>	<b>82.50</b>	81.02	76.35
		AD vs. MCI	75.97	73.91	<b>78.04</b>	<b>77.09</b>	75.47	75.03
		NC vs. MCI	75.83	76.16	75.49	75.66	75.91	75.16
	Level <sub>5</sub>	AD vs. NC	<b>82.16</b>	83.21	81.11	81.50	<b>82.34</b>	<b>77.23</b>
		AD vs. MCI	<b>77.28</b>	<b>78.04</b>	76.52	76.87	<b>77.45</b>	<b>76.21</b>
		NC vs. MCI	<b>77.42</b>	<b>78.18</b>	76.66	77.01	<b>77.59</b>	<b>76.66</b>
Level <sub>2</sub>	AD vs. NC	83.05	83.09	83.01	83.02	83.06	78.43	
	AD vs. MCI	75.67	77.69	73.65	74.67	76.15	76.18	
	NC vs. MCI	80.90	73.74	<b>88.07</b>	<b>86.07</b>	79.43	77.00	
3DMS-ChaineNet + Fusion	Level <sub>3</sub>	AD vs. NC	84.25	82.77	85.73	85.29	84.01	81.56
		AD vs. MCI	80.90	81.92	79.88	80.28	81.09	79.71
		NC vs. MCI	83.66	85.11	82.22	82.72	83.90	80.39
Level <sub>4</sub>	AD vs. NC	86.38	83.45	<b>89.32</b>	<b>88.66</b>	85.97	83.75	
	AD vs. MCI	84.61	84.41	84.82	84.76	84.58	81.10	
	NC vs. MCI	85.77	<b>88.30</b>	83.24	84.05	86.12	82.42	
Level <sub>5</sub>	AD vs. NC	<b>86.88</b>	<b>85.61</b>	88.15	87.84	<b>86.71</b>	<b>83.91</b>	
	AD vs. MCI	<b>85.92</b>	<b>84.73</b>	<b>87.10</b>	<b>86.79</b>	<b>85.75</b>	<b>81.98</b>	
	NC vs. MCI	<b>86.03</b>	88.18	83.88	84.55	<b>86.33</b>	<b>82.59</b>	
Level <sub>2</sub>	AD vs. NC	92.83	94.09	91.57	91.78	92.92	90.66	
	AD vs. MCI	90.71	90.87	90.55	90.58	90.73	85.38	
	NC vs. MCI	91.65	90.11	93.18	92.97	91.52	89.41	
3DMS-ChaineNet + Bi-LSFU	Level <sub>3</sub>	AD vs. NC	94.13	93.97	94.29	94.27	94.12	93.40
		AD vs. MCI	92.32	92.33	92.30	92.31	92.32	91.27
		NC vs. MCI	93.15	93.85	92.45	92.56	93.20	92.13
Level <sub>4</sub>	AD vs. NC	<b>96.60</b>	<b>96.19</b>	<b>97.01</b>	<b>96.99</b>	<b>96.59</b>	<b>95.73</b>	
	AD vs. MCI	<b>93.69</b>	<b>94.21</b>	93.18	93.25	<b>93.73</b>	<b>93.19</b>	
	NC vs. MCI	<b>95.27</b>	94.03	<b>96.52</b>	<b>96.43</b>	<b>95.21</b>	<b>94.50</b>	
Level <sub>5</sub>	AD vs. NC	96.47	95.96	96.98	96.95	96.45	95.61	
	AD vs. MCI	92.55	91.25	<b>93.85</b>	<b>93.69</b>	92.46	92.53	

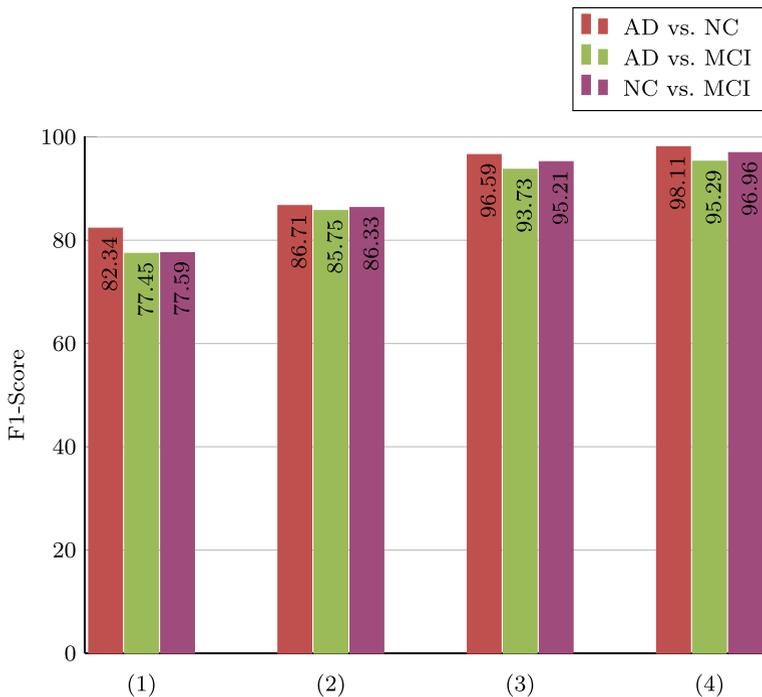
**Table 8** (continued)

Methods	Level <sub>n</sub>	Tasks	Acc. (%)	Sen. (%)	Spe. (%)	Pre. (%)	FSc. (%)	AUC (%)
	Level <sub>2</sub>	NC vs. MCI	94.80	<b>94.64</b>	94.97	94.95	94.80	93.74
		AD vs. NC	93.15	93.74	92.57	92.65	93.19	92.70
		AD vs. MCI	91.25	90.84	91.66	91.59	91.22	90.04
		NC vs. MCI	92.04	91.52	92.57	92.49	92.00	91.22
3DMS-ChaineNet +	Level <sub>3</sub>	AD vs. NC	94.75	94.50	95.00	94.97	94.73	94.65
		AD vs. MCI	92.99	93.36	92.63	92.68	93.02	93.27
Bi-GRFU	Level <sub>3</sub>	NC vs. MCI	93.94	93.83	94.06	94.05	93.94	93.85
		AD vs. NC	<b>98.12</b>	<b>97.51</b>	<b>98.74</b>	<b>98.72</b>	<b>98.11</b>	<b>96.58</b>
	Level <sub>4</sub>	AD vs. MCI	<b>95.26</b>	<b>95.99</b>	<b>94.53</b>	<b>94.61</b>	<b>95.29</b>	<b>94.86</b>
		NC vs. MCI	<b>96.97</b>	<b>96.72</b>	<b>97.22</b>	<b>97.20</b>	<b>96.96</b>	<b>95.32</b>
		AD vs. NC	97.11	96.40	97.83	97.80	97.09	96.43
	Level <sub>5</sub>	AD vs. MCI	93.68	95.00	92.36	92.56	93.76	93.18
		NC vs. MCI	95.83	96.14	95.52	95.55	95.84	94.50

The bold entries represent the best performances obtained for each ablation experiment

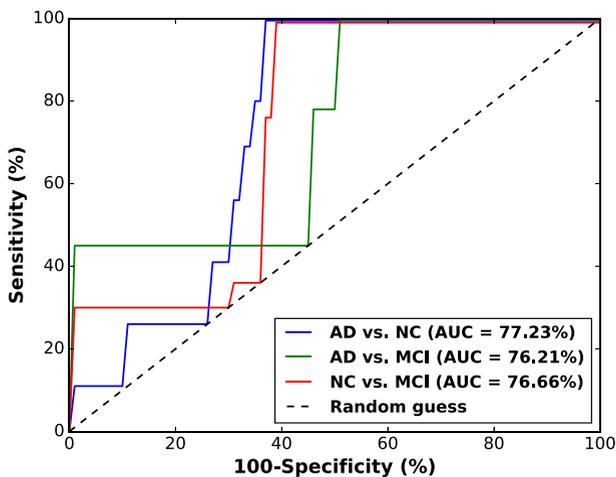
four iterations. The first part addresses the effect of the number of 3DMS-ChaineNet blocks without fusion on AD recognition performance. In the second part, we suggest connecting these 3DMS-ChaineNet blocks, at different levels, using an additional concatenation layer to combine the 3DMS-ChaineNet features of each block. In the third and fourth parts, the BG-3DM2F framework is iteratively trained using two categories of sequential architectures, namely Bidirectional Long Short-term Fusion Unit (Bi-LSFU) and Bi-GRFU by applying Bi-LSTM [35] and Bi-GRU [111], respectively, and the quality classification results are compared through testing. From Table 8, we can see that the different choices make different results. In order to figure out which component is working, we carried out four ablation experiments, termed as 3DMS-ChaineNet, 3DMS-ChaineNet+Fusion, 3DMS-ChaineNet+Bi-LSFU, and 3DMS-ChaineNet+Bi-GRFU, which were evaluated in four iterations denoted by Level<sub>n</sub>, where  $n \in \{2,3,4,5\}$ . Practically, we started the comparative experiments from the second level (Level<sub>2</sub>) and increased the number of levels up to 5 (Level<sub>5</sub>), because, from Level<sub>2</sub>, the convergence of the learning curves is clearly noticeable on the training stage and the plots of training and validation losses started to decrease to a point of stability which indicate that knowledge has been obtained. The highest accuracy, sensitivity, specificity, precision, F1-Score, and AUC for each ablation experiment are highlighted in bold.

From the results shown in Table 8, the 3DMS-ChaineNet at Level<sub>5</sub> gave better classification compared with 3DMS-ChaineNet at Level<sub>1</sub>, Level<sub>2</sub>, Level<sub>3</sub>, and Level<sub>4</sub> in terms of accuracy, F1-Score and AUC. Similar to 3DMS-ChaineNet, the results of 3DMS-ChaineNet+Fusion at Level<sub>5</sub> has also a better classification effect than other levels. Moreover, as the number of 3DMS-ChaineNet blocks increases, the accuracy, F1-Score, and AUC improve significantly and 3DMS-ChaineNet+Fusion in the fifth iteration (Level<sub>5</sub>) performs better than 3DMS-ChaineNet in all statistical metrics for the three tasks. When

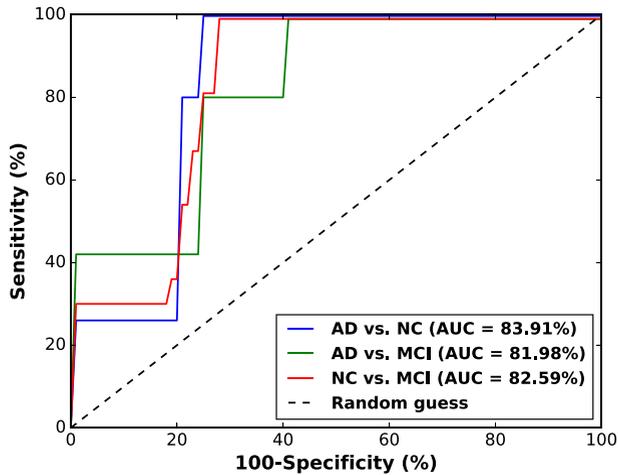


**Fig. 6** Bar graph illustrating the comparison of the performances of BG-3DM2F components in terms of F1-Score for AD vs. NC, AD vs. MCI, and NC vs. MCI classification. (1) is 3DMS-ChaineNet model at Level<sub>5</sub>. (2) represents 3DMS-ChaineNet+Fusion model at Level<sub>5</sub>. (3) stands for 3DMS-ChaineNet+Bi-LSFU model at Level<sub>4</sub>. (4) refers to the 3DMS-ChaineNet+Bi-GRFU model at Level<sub>4</sub>

compared to 3DMS-ChaineNet+Fusion at Level<sub>5</sub>, our algorithm 3DMS-ChaineNet+Bi-GRFU at Level<sub>4</sub> achieves an AUC gain of 12.67% for AD vs. NC, 12.88% for AD vs.



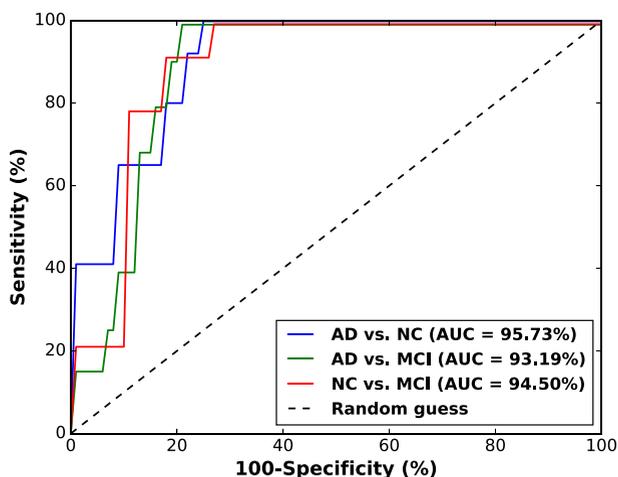
**Fig. 7** The ROC curves with the corresponding AUCs of 3DMS-ChaineNet model at Level<sub>5</sub> for the three binary classification tasks: AD vs. NC, AD vs. MCI, and NC vs. MCI



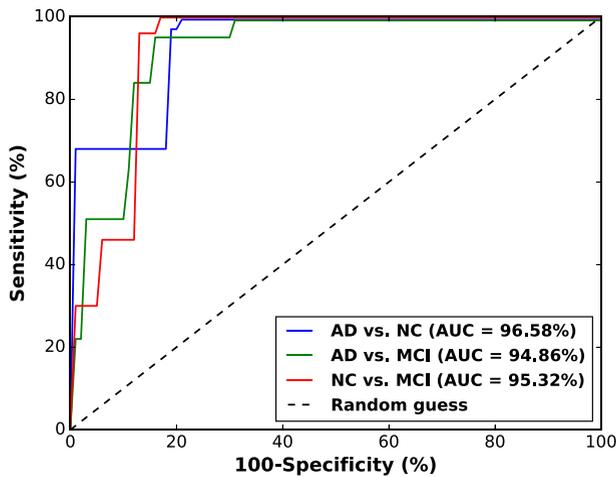
**Fig. 8** The ROC curves with the corresponding AUCs of 3DMS-ChaineNet+Fusion model at Level<sub>5</sub> for the three binary classification tasks: AD vs. NC, AD vs. MCI, and NC vs. MCI

MCI, and 12.73% for NC vs. MCI. 3DMS-ChaineNet+Bi-LSFU at Level<sub>4</sub> and Level<sub>5</sub> performs better than 3DMS-ChaineNet + Fusion but worse than 3DMS-ChaineNet+Bi-GRFU in the same iterations. Theoretically, the greater the number of convolutional layers, the higher the accuracy. Unfortunately, this is not the case in our model. From the empirical point of view, the Level<sub>5</sub> decreases the performance of both 3DMS-ChaineNet+Bi-LSFU and 3DMS-ChaineNet+Bi-GRFU models compared to Level<sub>4</sub>. The predictive models that yielded the highest peak performance in terms of F1-Score are shown in Fig. 6.

As illustrated in figures (Figs. 7, 8, 9 and 10), the ablation studies were evaluated using ROC curves which present the predictive models that achieved peak performance in terms of AUC. During the diagnosis procedure, the classification of AD against MCI is the



**Fig. 9** The ROC curves with the corresponding AUCs of 3DMS-ChaineNet+Bi-LSFU model at Level<sub>4</sub> for the three binary classification tasks: AD vs. NC, AD vs. MCI, and NC vs. MCI



**Fig. 10** The ROC curves with the corresponding AUCs of 3DMS-ChaineNet+Bi-GRFU model at Level<sub>4</sub> for the three binary classification tasks: AD vs. NC, AD vs. MCI, and NC vs. MCI

most challenging and very tedious task. From the results summarized in Table 8, the proposed BG-3DM2F system achieved the AUC of 94.86% for AD vs. MCI and 95.32% for NC vs. MCI. When AD was tested against NC, the results were notable with significant improvements in terms of AUC, yielding 96.58%.

#### 4.4 Comparison with the state-of-the-art

As shown in Table 9, we make a comparison between the proposed BG-3DM2F system and existing state-of-the-art methods to evaluate and confirm the quality of this study. Based on the summary and comparison of the AD research results discussed in Section 2, it is difficult to evaluate the BG-3DM2F system and ensure its quality that achieves high effectiveness in terms of AD diagnosis, because the dataset, the classification algorithm (binary or multi-class classification), and quantitative evaluation methods are not standardized. As a result, we should assure that for a dependable comparison and reliable clinical research, it would be necessary to compare the performance of the proposed method with those evaluated on the same dataset using the same evaluation metrics for validating the performance of these three binary classifications: AD vs. NC, AD vs. MCI, and NC vs. MCI. To perform this task, nine state-of-the-art methods for AD diagnosis are compared with our proposed algorithm. All these methods were carried out on the ADNI dataset using three evaluation metrics, namely accuracy, sensitivity, and specificity. The visual comparison is given in Table 9 demonstrates that our best performing network outperforms the Gauss-Laguerre (GL) circular harmonic functions combined with Bag-of-Visual-Words (BVW) and Late Fusion (LF) of classification results (GL-BVW-LF) on two biomarkers including the hippocampus and cerebrospinal fluid [12], GL circular harmonic functions combined with BVW, Principal Component Analysis (PCA) and Support Vector Machines (SVM) (GL-BVW-PCA-SVM) [14], Modified LeNet (MLeNet) [2], GL circular harmonic functions combined with BVW, Multiple Kernel Learning (MKL), and SVM (GL-BVW-MKL-SVM) [13], Cross-Modal Transfer Learning (CMTL) [3], Modified architectural pattern of DenseNet-121 (MDenseNet-121) [45], Multiple Kernel Boosting (MKB) [60], 3D Fine-Tuning CNN (3DFTCNN) [92], and

**Table 9** Comparison between the proposed method and other algorithms using the ADNI dataset

Methods	Tasks	Year	Acc.%	Sen.%	Spe.%
GL-BVW-LF [12]	AD vs. NC	2014	87.00	75.50	100
	AD vs. MCI		72.23	75.00	70.00
	NC vs. MCI		78.22	70.73	83.34
GL-BVW-PCA-SVM [14]	AD vs. NC	2015	83.70	78.80	85.70
	AD vs. MCI		76.50	78.90	52.80
	NC vs. MCI		66.70	62.00	68.30
MLeNet [2]	AD vs. NC	2016	82.80	79.60	85.90
	AD vs. MCI		62.50	60.00	64.00
	NC vs. MCI		66.00	73.70	58.70
GL-BVW-MKL-SVM [13]	AD vs. NC	2017	90.20	82.92	97.20
	AD vs. MCI		76.63	65.62	81.33
	NC vs. MCI		79.42	71.58	86.05
CMTL [3]	AD vs. NC	2018	92.50	94.70	90.40
	AD vs. MCI		85.00	93.70	79.10
	NC vs. MCI		80.00	92.80	73.00
MDenseNet-121 [45]	AD vs. NC	2018	94.97	94.33	95.89
	AD vs. MCI		91.98	90.47	95.38
	NC vs. MCI		74.70	70.96	78.20
MKB [60]	AD vs. NC	2018	94.65	95.03	91.76
	AD vs. MCI		88.63	91.55	86.25
	NC vs. MCI		84.79	88.91	80.34
3DFTCNN [92]	AD vs. NC	2018	96.81	96.30	97.20
	AD vs. MCI		88.43	88.00	88.90
	NC vs. MCI		92.62	92.20	93.10
3DMSF-GRU [8]	AD vs. NC	2019	97.03	96.28	97.77
	AD vs. MCI		93.42	95.20	91.63
	NC vs. MCI		95.90	97.04	94.76
BG-3DM2F (Ours)	AD vs. NC	2020	98.12	97.51	98.74
	AD vs. MCI		95.26	95.99	94.53
	NC vs. MCI		96.97	96.72	97.22

shallow 3D Multi-Scale Feature blocks combined with GRU cells (3DMSF-GRU) [8]. These state-of-the-art approaches can be categorized into four major following groups. The handcrafted methods [12–14, 60], shallow 2D-CNN approaches [2, 3], deep 2D-CNN [45] and 3D-CNN [8, 92]. More importantly, based on the shallow multi-scale 3D-CNN combined with unidirectional GRU as described in [8], by introducing the Bi-GRU cells and more fused 3D convolutional layers, we have an accuracy gain of 1.09%, 1.84% and 1.07% for AD vs. NC, AD vs. MCI and NC vs. MCI, respectively. Furthermore, when we compared the proposed method with those evaluated on both modalities; sMRI and Diffusion Tensor Imaging (DTI) [3, 13], we found that the proposed BG-3DM2F framework performs well in separating the classes, especially for the most challenging classification task (AD vs. MCI).

## 5 Discussion

In this paper, we proposed an optimal data representation framework for CAD system which is recently regarded as an indispensable system to support decision-making processes in the diagnosis of AD using volumetric neuroimaging data. Our work is motivated by the following observations. First, hippocampal atrophy is a key feature of neurodegeneration and it is widely considered as a structural magnetic resonance imaging biomarker for Alzheimer's disease diagnosis. Second, to perform visual recognition in the three-dimensional domain, 3D-CNN has achieved a significant breakthrough and become the mainstream approach in different challenging medical imaging tasks, such as image classification, object detection, semantic segmentation, and content-based image retrieval. In recent years, many medical imaging studies have applied 3D-CNNs and reported improved performance for a broad range of medical tasks. Third, currently, the bidirectional sequence model provides complementary advantages in modeling, including learning a continuous representation, encoding and managing sequential data, capturing contextual information, mitigating the vanishing gradient problem, and preserving the information state over long periods. Inspired by these observations, we find that the CAD systems can benefit from 3D-CNN structures and bidirectional sequence models to recognize AD. To this end, we proposed a novel approach to retain all 3D-CNN levels in the decision-making process based on a 3D multi-scale feature fusion and a bidirectional sequence model. Specifically, we used Bi-GRU to store hierarchical 3D multi-level features into a low dimensional semantic space. To evaluate our approach, we conduct experiments on the ADNI dataset. Among several evaluation procedures, the randomly partitioning of a dataset into 60% for training, 20% for validation, and 20% for testing are the most commonly used in the literature to evaluate the results of hippocampal atrophy classification, since it avoids overfitting when testing the generalization ability of the classification algorithm.

To better demonstrate the effectiveness of the proposed BG-3DM2F framework and the improvements obtained by different modules in its network, two main study axes are presented in this paper. In the first axis, we have analyzed the performance results with the change in the number of 3DMS-ChaineNet blocks, Bi-LSTM [35], and Bi-GRU [111] cells in our method by conducting empirical experiments on the ADNI dataset. The results of the evaluation related to four iterations were presented in Table 8. These findings deliver important intuitions regarding hippocampal volume features. The best performance was achieved at Level<sub>5</sub> when using five 3DMS-ChaineNet blocks with and without fusion. Given that deeper networks have a better ability for semantic abstraction than shallow ones, we supposed that integrating more 3DMS-ChaineNet blocks and Bi-LSTM or Bi-GRU cells into our proposed framework could yield outstanding performance. Unfortunately, from the perspective of modeling, both 3DMS-ChaineNet+Bi-LSFU and 3DMS-ChaineNet+Bi-GRFU models at Level<sub>5</sub> perform significantly worse compared to these models at Level<sub>4</sub> in terms of accuracy, F1-Score, and AUC metrics for all tasks. The second study axis is conducted to demonstrate the potentiality of the feature fusion, Bi-LSTM, and Bi-GRU. According to statistical outcomes shown in Table 8, we gain three significant reports. Firstly, it is noticeable that the single 3DMS-ChaineNet stream without fusion is not sufficient for accurate AD recognition. Secondly, these findings reaffirm the superiority of Bi-GRU which is more appropriate to preserve information than Bi-LSTM. Finally, the experiments on ADNI datasets reveal that integrating sequence models; Bi-LSTM and Bi-GRU cells into 3D-CNN can automatically focus on critical information, harvest discriminatory features, and strengthen the discriminative representation for AD diagnosis compared with the existing conventional 3D-CNN frameworks.

To confirm the quality of the BG-3DM2F algorithm, our results were compared with nine previous works. These methods can be broadly categorized into four classes: the handcrafted methods [12–14, 60], shallow 2D-CNN approaches [2, 3], deep 2D-CNN [45], and 3D-CNN [8, 92]. By comparing the results in Table 9, it is remarkable that our proposed system achieved promising results in terms of accuracy, sensitivity, and specificity. From the analysis of these results, it is evident that our proposed algorithm is more robust as compared to the three recent deep learning state-of-the-art works, namely modified DenseNet-121 [45], 3D fine-tuning convolutional neural network [92] and shallow multi-scale 3D-CNN combined with the GRU cells [8]. The optimization algorithm used in most of these previous works is Stochastic Gradient Descent (SGD), which is very slow, and causes convergence to the local minima rather than the global minima [52]. Although the Rectified Linear Unit (ReLU) is the most popular activation function and widely used in 3D-CNN, the use of ReLU can lead to the oscillation problem in optimization algorithms [106]. From the theoretical and practical points of view, the ReLU replaces the negative component by zero, occurring the bias shift of the outputs and the death of some neurons, which produces the impeding learning and vanishing gradient. In our study, we solved these issues by using Adam optimization algorithm [76] to train BG-3DM2F architecture, which combines the advantages of two recently popular methods; AdaGrad [31] and RMSProp [68]. Accordingly, Adam is a good alternative to SGD since it can provide improved robustness to the convergence to local minima and high memory consumption. To overcome the limitations of ReLU, SELU [51] is proposed as an activation function that can avoid the zero gradients in the negative part, allow BG-3DM2F to be trained faster, and provide easier sparse representations. In order to address the problem of overfitting, we used three techniques; inserting a dropout [9] layer after FC-1 layer with a probability of 0.25, using L2 regularization [58] with factor  $\epsilon=0.05$ , and applying data augmentation methods to expand the size of our dataset and increase the generalization ability of our model. Additionally, the hyper-parameters setting also plays a substantial role in the final performance of the systems. In this work, the hyper-parameters were experimentally chosen until the network began to converge effectively. From the practical point of views [23, 28, 47, 59, 64], the best way to alleviate the problem of semantic feature ambiguity, high inter-class visual similarities between the classes, and vanishing gradient issue is to integrate bidirectional sequence models into multi-scale feature fusion architectures to control information flow between layers in the network and improve the memory ability.

The computational time is one of the most essential factors of 3D-CNN. Consequently, we have always tried to insert after each 3DMS-ChaineNet block the three-dimensional depth-wise [97] and average down-sampling layers [78, 99]. The 3D depth-wise is a linear compression layer used to reduce the number of parameters in 3D convolutions while the 3D average pooling layer reduces the dimension of the features with completely keeping a balance between high-performance achievement and low computational times. Both layers maintain the local perception capacity of the model with lower computational cost [37, 65]. Moreover, the hippocampal volumes were taken as input with the smaller than the large-scale whole brain to improve model building time and mitigate the high computational complexity. By using the hippocampal VOI as input, the average computational time required to recognize AD is 12.852 hours and 4.565 hours for training and testing phases, respectively.

To the best of our knowledge, there have been very few instances in which 3D-CNN has been used for classification and prediction on volumetric neuroimaging data. The present work is unique in the field because we have, for the first time, used an improved multi-scale

feature fusion in 3D-CNN combined with bidirectional sequence models for the automated recognition system of AD. Based on the results, the summary of the proposed algorithm can be drawn as:

- This paper provides a series of comparative experiments to acquire the best performance of BG-3DM2F since the presence of 3DMS-ChaineNet and Bi-GRFU can be the major advantages.
- The proposed BG-3DM2F framework performs dramatically better compared to the traditional CAD systems for AD recognition, which is capable of outperforming the shallow 3D multi-scale feature blocks combined with GRU cells, as well as the well-known DenseNet-121 architecture, producing strong gains in terms of accuracy, sensitivity, and specificity.
- The proposed BG-3DM2F takes advantage of using the scaled exponential linear unit to resolve the bias shift, Adam optimization algorithm to avoid the convergence to local minima and high cost of memory accessibility, Bi-GRU to alleviate vanishing gradient issue, 3D multi-scale features combined with 3D fusion to mitigate feature ambiguity in complex visual discriminations, hippocampal volumes with the size of  $28 \times 28 \times 28$  voxels instead of large-scale whole brains and embedding of 3D space into a low dimensional semantic space to deal with long training time and high computational costs.
- The proposed method can be extended to integrate other neuroimaging biomarkers and formulate a more robust CAD system for AD recognition in its early stage.

Last but not least, this computational method based on 3D multi-scale feature fusion and Bi-GRU cells can perform better and more effectively in recognizing AD in volumetric sMRI modality. Besides, this method may fix and control the various gaps in the AD recognition area, starting from VOI selection, data augmentation, and ending with the complex hippocampal atrophy classification. our study is in progress. Despite these promising findings replace previous state-of-the-art methods as described in Section 4.4, there are still some limitations. For instance, the classification of AD against MCI is a tedious task and it needs some improvements to investigate the factors affecting its classification performance.

## 6 Conclusion and future work

In this paper, the BG-3DM2F framework for AD recognition in sMRI scans is proposed. It consists of two major sub-networks, namely 3DMS-ChaineNet and Bi-GRFU. The 3DMS-ChaineNet model accurately learns the most significant features, while the Bi-GRFU architecture controls 3DMS-ChaineNet information flow in a parallel way, captures contextual information, alleviates the vanishing gradient problem, and preserves the information state over long periods and in a bidirectional manner. To improve model building time, we took the hippocampal VOI as input with a smaller size than the large-scale whole brain, and each 3DMS-ChaineNet block was embedded into a low dimensional semantic space using the 3D depth-wise layer and 3D average pooling without losing valuable information. To demonstrate the effectiveness of the proposed BG-3DM2F framework, we conducted several experiments on the ADNI dataset. The experimental results show that the proposed BG-3DM2F generally outperforms other state-of-the-art approaches in AD recognition.

Despite the success achieved by this method for hippocampal atrophy classification, there are still some weaknesses that make it more difficult and challenging. Recognition

of mild cognitive impairment, which is considered as a transitional stage between normal control and Alzheimer's disease, is a difficult task because it is often confused with AD. In all likelihood, these confusions occur due to similar morphological features of hippocampal volumes in mild and severe stages. In future work, more efforts will be made to improve these results by exploring features from other data modalities, such as PET, DTI, and Functional Magnetic Resonance Imaging (fMRI), enhancing our BG-3DM2F with multi-resolution analysis, and investigating other cerebral regions that impacted by Alzheimer's degeneration.

**Acknowledgements** This work is supported in part by the PPR2-2015 project with the donation of the NVIDIA Geforce GTX 1080 Ti GPU used for this research under grant no 14UIZ2015, and in part by the Al Khawarizmi project financed by the Moroccan government through the CNRST funding program.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health ([www.fnih.org](http://www.fnih.org)). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

## References

1. Abdelaziz Ismael S, Mohammed A, Hefny H (2020) An enhanced deep learning approach for brain cancer MRI images classification using residual networks. *Artif Intell Med* 102:101779. <https://doi.org/10.1016/j.artmed.2019.101779>
2. Aderghal K, Boissenin M, Benois-Pineau J, Catheline G, Afdel K (2016) Classification of sMRI for AD Diagnosis with Convolutional Neuronal Networks: A Pilot 2-D+ $\epsilon$  Study on ADNI. *MultiMedia Modeling* 690–701. [https://doi.org/10.1007/978-3-319-51811-4\\_56](https://doi.org/10.1007/978-3-319-51811-4_56)
3. Aderghal K, Khvostikov A, Krylov A, Benois-Pineau J, Afdel K, Catheline G (2018) Classification of Alzheimer disease on imaging modalities with deep CNNs using cross-modal transfer learning. In: 2018 IEEE 31st international symposium on computer-based medical systems (CBMS). <https://doi.org/10.1109/cbms.2018.00067>
4. Amin S, Alsulaiman M, Muhammad G, Mekhtiche M, Shamim Hossain M (2019) Deep Learning for EEG motor imagery classification based on multi-layer CNNs feature fusion. *Futur Gener Comput Syst* 101:542–554. <https://doi.org/10.1016/j.future.2019.06.027>
5. Arifoglu D, Bouchachia A (2019) Detection of abnormal behaviour for dementia sufferers using Convolutional Neural Networks. *Artif Intell Med* 94:88–95. <https://doi.org/10.1016/j.artmed.2019.01.005>
6. Bakkouri I, Afdel K (2018) Multi-scale CNN based on region proposals for efficient breast abnormality recognition. *Multimed Tools Appl* 78(10):12939–12960. <https://doi.org/10.1007/s11042-018-6267-z>
7. Bakkouri I, Afdel K (2019) Computer-aided diagnosis (CAD) system based on multi-layer feature fusion network for skin lesion recognition in dermoscopy images. *Multimed Tools Appl* 79(29–30):20483–20518. <https://doi.org/10.1007/s11042-019-07988-1>
8. Bakkouri I, Afdel K, Benois-Pineau J, Catheline G (2019) Recognition of alzheimer's disease on sMRI based on 3D multi-scale CNN features and a gated recurrent fusion unit. In: 2019 international conference on content-based multimedia indexing (CBMI). <https://doi.org/10.1109/cbmi.2019.8877477>

9. Baldi P, Sadowski P (2014) The dropout learning algorithm. *Artif Intell* 210:78–122. <https://doi.org/10.1016/j.artint.2014.02.004>
10. Banning L, Ramakers I, Köhler S, Bron E, Verhey F, de Deyn P, Claassen J, Koek H, Middelkoop H, van der Flier W, van der Lugt A, Aalten P (2020) The association between biomarkers and neuropsychiatric symptoms across the alzheimer's disease spectrum. *Amer J Geriatric Psychiatry* 28:735–744. <https://doi.org/10.1016/j.jagp.2020.01.012>
11. Basaia S, Agosta F, Wagner L, Canu E, Magnani G, Santangelo R, Filippi M (2019) Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. *NeuroImage: Clinical* 21:101645. <https://doi.org/10.1016/j.nicl.2018.101645>
12. Ben Ahmed O, Benois-Pineau J, Allard M, Ben Amar C, Catheline G (2014) Classification of Alzheimer's disease subjects from MRI using hippocampal visual features. *Multimed Tools Appl* 74:1249–1266. <https://doi.org/10.1007/s11042-014-2123-y>
13. Ben Ahmed O, Benois-Pineau J, Allard M, Catheline G, Ben Amar C (2017) Recognition of Alzheimer's disease and Mild Cognitive Impairment with multimodal image-derived biomarkers and Multiple Kernel Learning. *Neurocomputing* 220:98–110. <https://doi.org/10.1016/j.neucom.2016.08.041>
14. Ben Ahmed O, Mizotin M, Benois-Pineau J, Allard M, Catheline G, Ben Amar C (2015) Alzheimer's disease diagnosis on structural MR images using circular harmonic functions descriptors on hippocampus and posterior cingulate cortex. *Comput Med Imaging Graph* 44:13–25. <https://doi.org/10.1016/j.compmedimag.2015.04.007>
15. Ben Miled Z, Haas K, Black C, Khandker R, Chandrasekaran V, Lipton R, Boustani M (2020) Predicting dementia with routine care EMR data. *Artif Intell Med* 102:101771. <https://doi.org/10.1016/j.artmed.2019.101771>
16. Budson A, Solomon P (2016) Evaluating the Patient with Memory Loss or Dementia. *Memory Loss Alzheimer's Disease, and Dementia*, 5–38. <https://doi.org/10.1016/b978-0-323-28661-9.00002-0>
17. Budson A, Solomon P (2016) Alzheimer's disease dementia and mild cognitive impairment due to alzheimer's disease. *Memory loss Alzheimer's Disease, and Dementia*, 47–69. <https://doi.org/10.1016/b978-0-323-28661-9.00004-4>
18. Bui T, Lee J, Shin J (2019) Incorporated region detection and classification using deep convolutional networks for bone age assessment. *Artif Intell Med* 97:1–8. <https://doi.org/10.1016/j.artmed.2019.04.005>
19. Cahall D, Rasool G, Bouaynaya N, Fathallah-Shaykh H (2019) Inception Modules Enhance Brain Tumor Segmentation. *Frontiers in Computational Neuroscience*. <https://doi.org/10.3389/fncom.2019.00044>
20. Castro A, Coria-Lucero C, Anzulovich A, Navigatore-Fonzo L, 2019 Effects of experimental intracerebral ventricular injection of amyloid beta peptide (1-42) aggregates on daily rhythms of A-degrading enzymes in the hippocampus: Relevance to Alzheimer's disease pathophysiology. *Pathophysiology*. <https://doi.org/10.1016/j.pathophys.2019.07.003>
21. Cattaud V, Bezzina C, Rey C, Lejards C, Dahan L, Verret L (2018) Early disruption of parvalbumin expression and perineuronal nets in the hippocampus of the Tg2576 mouse model of Alzheimer's disease can be rescued by enriched environment. *Neurobiol Aging* 72:147–158. <https://doi.org/10.1016/j.neurobiolaging.2018.08.024>
22. Cheng D, Liu M (2017) Classification of Alzheimer's Disease by Cascaded Convolutional Neural Networks Using PET Images. *Machine Learning in Medical Imaging*, 106–113. [https://doi.org/10.1007/978-3-319-67389-9\\_13](https://doi.org/10.1007/978-3-319-67389-9_13)
23. Cheng H, Xie Z, Wu L, Yu Z, Li R (2019) Data prediction model in wireless sensor networks based on bidirectional LSTM. *EURASIP Journal on Wireless Communications and Networking*. <https://doi.org/10.1186/s13638-019-1511-4>
24. Cogan T, Cogan M, Tamil L (2019) MAPGI: Accurate Identification of anatomical landmarks and diseased tissue in gastrointestinal tract using deep learning. *Comput Biol Med* 111:103351. <https://doi.org/10.1016/j.compbiomed.2019.103351>
25. Cruz-Alonso M, Fernandez B, Navarro A, Junceda S, Astudillo A, Pereiro R (2019) Laser ablation ICP-MS for simultaneous quantitative imaging of iron and ferroportin in hippocampus of human brain tissues with Alzheimer's disease. *Talanta* 197:413–421. <https://doi.org/10.1016/j.talanta.2019.01.056>
26. Cui R, Liu M (2018) Hippocampus analysis based on 3D CNN for Alzheimer's disease diagnosis. *Tenth International Conference on Digital Image Processing (ICDIP)*, 2018. <https://doi.org/10.1117/12.2503194>
27. Cui R, Liu M (2019) Hippocampus analysis by combination of 3-D DenseNet and shapes for Alzheimer's disease diagnosis. *IEEE J Biomed Health Inform* 23:2099–2107. <https://doi.org/10.1109/jbhi.2018.2882392>

28. Dai Y, Wang C, Dong J, Sun C (2019) Visual relationship detection based on bidirectional recurrent neural network. *Multimedia Tools and Applications*. <https://doi.org/10.1007/s11042-019-7732-z>
29. Dey R, Salemt F (2017) Gate-variants of Gated Recurrent Unit (GRU) neural networks. 2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS). <https://doi.org/10.1109/mwscas.2017.8053243>
30. Du X, Qu X, He Y, Guo D (2018) Single image Super-Resolution based on Multi-Scale competitive convolutional neural network. *Sensors* 18:789. <https://doi.org/10.3390/s18030789>
31. Duchi J, Hazan E, Singer Y (2011) Adaptive subgradient methods for online learning and stochastic optimization. *J Machine Learn Res* 12:2121–2159
32. Esmaeilzadeh S, Belivanis D, Pohl K, Adeli E (2018) End-To-End Alzheimer's Disease Diagnosis and Biomarker Identification. *Machine Learning in Medical Imaging*, 337–345. [https://doi.org/10.1007/978-3-030-00919-9\\_39](https://doi.org/10.1007/978-3-030-00919-9_39)
33. Fawzi A, Moosavi-Dezfooli S, Frossard P (2017) The robustness of deep networks: a geometrical perspective. *IEEE Signal Proc Mag* 34:50–62. <https://doi.org/10.1109/msp.2017.2740965>
34. Feng C, Elazab A, Yang P, Wang T, Lei B, Xiao X (2018) 3D convolutional neural network and stacked bidirectional recurrent neural network for Alzheimer's Disease diagnosis. *PRedictive Intelligence in Medicine*, 138–146. [https://doi.org/10.1007/978-3-030-00320-3\\_17](https://doi.org/10.1007/978-3-030-00320-3_17)
35. Feng C, Elazab A, Yang P, Wang T, Zhou F, Hu H, Xiao X, Lei B (2019) Deep learning framework for Alzheimer's disease diagnosis via 3d-CNN and FSBi-LSTM. *IEEE Access* 7:63605–63618. <https://doi.org/10.1109/access.2019.2913847>
36. Frisoni G, Testa C, Sabattoli F, Beltramello A, Soininen H, Laakso M (2005) Structural correlates of early and late onset Alzheimer's disease: voxel based morphometric study. *J Neurology Neurosurg Psychiatry* 76:112–114. <https://doi.org/10.1136/jnnp.2003.029876>
37. Gessert N, Beringhoff J, Otte C, Schlaefer A (2018) Force estimation from OCT volumes using 3D CNNs. *Int J CARS* 13:1073–1082. <https://doi.org/10.1007/s11548-018-1777-8>
38. Grajski K, Bressler S (2019) Differential medial temporal lobe and default-mode network functional connectivity and morphometric changes in Alzheimer's disease. *NeuroImage: Clinical* 23:101860. <https://doi.org/10.1016/j.nicl.2019.101860>
39. Hosseini Asl E, Ghazal M, Mahmoud A, Aslantas A, Shalaby A, Casanova M, Barnes G, Gimel'farb G, Keynton R, El Baz A (2018) Alzheimer's disease diagnostics by a 3D deeply supervised adaptable convolutional network. *Front Biosci* 23:584–596. <https://doi.org/10.2741/4606>
40. Hsiao T, Chang Y, Chou H, Chiu C (2019) Filter-based deep-compression with global average pooling for convolutional networks. *J Syst Archit* 95:9–18. <https://doi.org/10.1016/j.sysarc.2019.02.008>
41. Hu X, Teunissen C, Spottke A, Heneka M, Düzel E, Peters O, Li S, Priller J, Buerger K, Teipel S, Laske C, Verfaillie S, Barkhof F, Coll-Adrós N, Rami L, Molinuevo J, van der Flier W, Jessen F (2019) Smaller medial temporal lobe volumes in individuals with subjective cognitive decline and biomarker evidence of Alzheimer's disease—Data from three memory clinic studies. *Alzheimer's Dementia* 15:185–193. <https://doi.org/10.1016/j.jalz.2018.09.002>
42. Huang Y, Xu J, Zhou Y, Tong T, Zhuang X (2019) Diagnosis of Alzheimer's Disease via Multi-Modality 3D Convolutional Neural Network. *Frontiers in Neuroscience*. <https://doi.org/10.3389/fnins.2019.00509>
43. Huang S, Zhong Z, Jin L, Zhang S, Wang H (2018) Dropregion training of inception font network for high-performance Chinese font recognition. *Pattern Recogn* 77:395–411. <https://doi.org/10.1016/j.patcog.2017.10.018>
44. Islam M, Foysal F, Neehal N, Karim E, Hossain S (2018) Inceptb: A CNN based classification approach for recognizing traditional Bengali games. *Procedia Computer Science* 143:595–602. <https://doi.org/10.1016/j.procs.2018.10.436>
45. Islam J, Zhang Y (2018) Deep convolutional neural networks for automated diagnosis of Alzheimer's disease and mild cognitive impairment using 3D Brain MRI. *Brain Informatics*, 359–369. [https://doi.org/10.1007/978-3-030-05587-5\\_34](https://doi.org/10.1007/978-3-030-05587-5_34)
46. Jack C, Bernstein M, Fox N, Thompson P, Alexander G, Harvey D, Borowski B, Britson P, Whitwell LJ, Ward C, Dale A, Felmlee J, Gunter J, Hill D, Killiany R, Schuff N, Fox-Bosetti S, Lin C, Studholme C, DeCarli C, Krueger G, Ward H, Metzger G, Scott K, Mallozzi R, Blezek D, Levy J, Debbins J, Fleisher A, Albert M, Green R, Bartzokis G, Glover G, Mugler J, Weiner M (2008) The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *J Magn Reson Imaging* 27:685–691. <https://doi.org/10.1002/jmri.21049>
47. Jagannatha A, Yu H (2016) Bidirectional RNN for medical event detection in electronic health records. In: *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. <https://doi.org/10.18653/v1/n16-1056>

48. Kam T, Zhang H, Shen D (2018) A Novel Deep Learning Framework on Brain Functional Networks for Early MCI Diagnosis. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, 293–301. [https://doi.org/10.1007/978-3-030-00931-1\\_34](https://doi.org/10.1007/978-3-030-00931-1_34)
49. Karasawa H, Liu C, Ohwada H (2018) Deep 3D Convolutional Neural Network Architectures for Alzheimer's Disease Diagnosis. *Intelligent Information and Database Systems*, 287–296. [https://doi.org/10.1007/978-3-319-75417-8\\_27](https://doi.org/10.1007/978-3-319-75417-8_27)
50. Kingma D, Ba J (2015) Adam: A method for stochastic optimization. 2015 international conference on learning representations (ICLR). arXiv:1412.6980
51. Klambauer G, Unterthiner T, Mayr A, Hochreiter S (2017) Self-normalizing neural networks. In: 2017 Annual conference on neural information processing systems (NIPS). arXiv:1706.02515
52. Kleinberg R, Li Y, Yuan Y (2018) An Alternative View: When Does SGD Escape Local Minima? In: 2018 international conference on machine learning (ICML). arXiv:1802.06175
53. Korolev S, Safiullin A, Belyaev M, Dodonova Y (2017) Residual and plain convolutional neural networks for 3D brain MRI classification 2017. *IEEE 14th International Symposium on Biomedical Imaging (ISBI)*. <https://doi.org/10.1109/isbi.2017.7950647>
54. Krizhevsky A, Sutskever I, Hinton G (2017) Imagenet classification with deep convolutional neural networks. *Commun ACM* 60:84–90. <https://doi.org/10.1145/3065386>
55. Li H, Habes M, Wolk D, Fan Y (2019) A deep learning model for early prediction of Alzheimer's disease dementia based on hippocampal magnetic resonance imaging data. *Alzheimer's Dementia* 15:1059–1070. <https://doi.org/10.1016/j.jalz.2019.02.007>
56. Li F, Liu M (2019) A hybrid convolutional and recurrent neural network for hippocampus analysis in Alzheimer's disease. *J Neurosci Methods* 323:108–118. <https://doi.org/10.1016/j.jneumeth.2019.05.006>
57. Li J, Liu X, Zhang M, Wang D (2019) Spatio-temporal deformable 3D ConvNets with attention for action recognition. *Pattern Recogn* 98:107037. <https://doi.org/10.1016/j.patcog.2019.107037>
58. Li F, Zurada J, Liu Y, Wu W (2017) Input layer regularization of multilayer feedforward neural networks. *IEEE Access* 5:10979–10985. <https://doi.org/10.1109/access.2017.2713389>
59. Liu G, Guo J (2019) Bidirectional LSTM with attention mechanism and convolutional layer for text classification. *Neurocomputing* 337:325–338. <https://doi.org/10.1016/j.neucom.2019.01.078>
60. Liu J, Li M, Lan W, Wu F, Pan Y, Wang J (2018) Classification of alzheimer's disease using whole brain hierarchical network. *IEEE/ACM Trans Comput Biology Bioinform* 15:624–632. <https://doi.org/10.1109/tcbb.2016.2635144>
61. Liu W, Wang Z, Liu X, Zeng N, Liu Y, Alsaadi F (2017) A survey of deep neural network architectures and their applications. *Neurocomputing* 234:11–26. <https://doi.org/10.1016/j.neucom.2016.12.038>
62. Liu B, Wei Y, Zhang Y, Yang Q (2017) Deep neural networks for high dimension low sample size data. In: *Proceedings of the twenty-sixth international joint conference on artificial intelligence*. <https://doi.org/10.24963/ijcai.2017/318>
63. Liu F, Zheng J, Zheng L, Chen C (2019) Combining attention-based bidirectional gated recurrent neural network and two-dimensional convolutional neural network for document-level sentiment classification. *Neurocomputing* 371:39–50. <https://doi.org/10.1016/j.neucom.2019.09.012>
64. Lyu C, Chen B, Ren Y, Ji D (2017) Long short-term memory RNN for biomedical named entity recognition. *BMC Bioinformatics*. <https://doi.org/10.1186/s12859-017-1868-5>
65. Marteau P, Gibet S, Reverdy C (2016) Adaptive Down-Sampling and Dimension Reduction in Time Elastic Kernel Machines for Efficient Recognition of Isolated Gestures. *Advances in Knowledge Discovery and Management*, 39–59. [https://doi.org/10.1007/978-3-319-45763-5\\_3](https://doi.org/10.1007/978-3-319-45763-5_3)
66. McNeely-White D, Beveridge J, Draper B (2019) Inception and ResNet features are (almost) equivalent. *Cogn Syst Res* 59:312–318. <https://doi.org/10.1016/j.cogsys.2019.10.004>
67. Mufson E, Mahady L, Waters D, Counts S, Perez S, DeKosky S, Ginsberg S, Ikonomic M, Scheff S, Binder L (2015) Hippocampal plasticity during the progression of Alzheimer's disease. *Neuroscience* 309:51–67. <https://doi.org/10.1016/j.neuroscience.2015.03.006>
68. Mukkamala M, Hein M (2017) Variants of RMSProp and adagrad with logarithmic regret bounds. In: 2017 international conference on machine learning (ICML). arXiv:1706.05507
69. Nalepa J, Ribalta Lorenzo P, Marcinkiewicz M, Bobek-Billewicz B, Wawrzyniak P, Walczak M, Kawulok M, Dudzik W, Kotowski K, Burda I, Machura B, Mrukwa G, Ulrych P, Hayball M (2020) Fully-automated deep learning-powered system for DCE-MRI analysis of brain tumors. *Artif Intell Med* 102:101769. <https://doi.org/10.1016/j.artmed.2019.101769>
70. Nam D, Barrack R, Potter H (2014) What Are the Advantages and Disadvantages of Imaging Modalities to Diagnose Wear-related Corrosion Problems? *Clin Orthop Relat Res*® 472:3665–3673. <https://doi.org/10.1007/s11999-014-3579-9>

71. Nickerson P, Tighe P, Shickel B, Rashidi P (2016) Deep neural network architectures for forecasting analgesic response. In: 2016 38th Annual international conference of the IEEE engineering in medicine and biology society (EMBC). <https://doi.org/10.1109/embc.2016.7591352>
72. Ning X, Duan P, Li W, Zhang S (2020) Real-Time 3D Face alignment using an Encoder-Decoder network with an efficient deconvolution layer. *IEEE Signal Processing Letters* 27:1944–1948. <https://doi.org/10.1109/lsp.2020.3032277>
73. Ning X, Gong K, Li W, Zhang L, Bai X, Tian S (2020) Feature Refinement and Filter Network for Person Re-identification. *IEEE Transactions on Circuits and Systems for Video Technology*, 1–1. <https://doi.org/10.1109/tcsvt.2020.3043026>
74. Ning X, Li W, Tang B, He H (2018) BULDP: Biomimetic Uncorrelated locality discriminant projection for feature extraction in face recognition. *IEEE Trans Image Process* 27:2575–2586. <https://doi.org/10.1109/tip.2018.2806229>
75. Ofori E, DeKosky S, Febo M, Colon-Perez L, Chakrabarty P, Duara R, Adjouadi M, Golde T, Vaillancourt D (2019) Free-water imaging of the hippocampus is a sensitive marker of Alzheimer's disease. *NeuroImage: Clinical* 24:101985. <https://doi.org/10.1016/j.nicl.2019.101985>
76. Ohzeki M, Okada S, Terabe M, Taguchi S (2018) Optimization of neural networks via finite-value quantum fluctuations. *Scientific Reports*. <https://doi.org/10.1038/s41598-018-28212-4>
77. Ouyang X, Xu S, Zhang C, Zhou P, Yang Y, Liu G, Li X (2019) A 3d-CNN and LSTM based Multi-Task learning architecture for action recognition. *IEEE Access* 7:40757–40770. <https://doi.org/10.1109/access.2019.2906654>
78. Pagola M, Forcen J, Barrenechea E, Fernández J, Bustince H (2017) A Study on the Cardinality of Ordered Average Pooling in Visual Recognition. *Pattern Recognition and Image Analysis*, 437–444. [https://doi.org/10.1007/978-3-319-58838-4\\_48](https://doi.org/10.1007/978-3-319-58838-4_48)
79. Pirzada S, Uddin M, Figley T, Kornelsen J, Puig J, Marrie R, Mazerolle E, Fisk J, Helmick C, O'Grady C, Patel R, Figley C (2020) Spatial normalization of multiple sclerosis brain MRI data depends on analysis method and software package. *Magn Reson Imaging* 68:83–94. <https://doi.org/10.1016/j.mri.2020.01.016>
80. Pluta R, Bogucka-Kocka A, Ułamek-Kozioł M, Bogucki J, Januszewski S, Kocki J, Czuczwar S (2018) Ischemic tau protein gene induction as an additional key factor driving development of Alzheimer's phenotypic changes in CA1 area of hippocampus in an ischemic model of Alzheimer's disease. *Pharmacol Rep* 70:881–884. <https://doi.org/10.1016/j.pharep.2018.03.004>
81. Rabai F, Ramani R (2017) Magnetic resonance imaging: Anesthetic implications. *Essentials of Neuroanesthesia*, 519–532. <https://doi.org/10.1016/b978-0-12-805299-0.00031-2>
82. Rao C, Liu Y (2020) Three-dimensional convolutional neural network (3d-CNN) for heterogeneous material homogenization. *Comput Mater Sci* 184:109850. <https://doi.org/10.1016/j.commatsci.2020.109850>
83. Rolls E, Huang C, Lin C, Feng J, Joliot M (2020) Automated anatomical labelling atlas 3. *NeuroImage* 206:116189. <https://doi.org/10.1016/j.neuroimage.2019.116189>
84. Rolls E, Joliot M, Tzourio-Mazoyer N (2015) Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas. *NeuroImage* 122:1–5. <https://doi.org/10.1016/j.neuroimage.2015.07.075>
85. Scheltens P, Blennow K, Breteler M, de Strooper B, Frisoni G, Salloway S, Van der Flier W (2016) Alzheimer's disease. *Lancet* 388:505–517. [https://doi.org/10.1016/s0140-6736\(15\)01124-1](https://doi.org/10.1016/s0140-6736(15)01124-1)
86. Selkoe D (2015) Alzheimer disease. Rosenberg's Molecular and Genetic Basis of Neurological and Psychiatric Disease, 753–768. <https://doi.org/10.1016/b978-0-12-410529-4.00067-x>
87. Shen J, Wang D, Huang Y, Wen M, Zhang C (2019) Scale-out Acceleration for 3D CNN-based Lung Nodule Segmentation on a Multi-FPGA System. In: Proceedings of the 56th annual design automation conference 2019 on - DAC '19. <https://doi.org/10.1145/3316781.3317906>
88. Shinagawa S, Kobayashi N, Nagata T, Kusaka A, Yamada H, Kondo K, Nakayama K (2016) DNA Methylation in the NCAPH2/LMF2 promoter region is associated with hippocampal atrophy in Alzheimer's disease and amnesic mild cognitive impairment patients. *Neurosci Lett* 629:33–37. <https://doi.org/10.1016/j.neulet.2016.06.055>
89. Shmulev Y, Belyaev M (2018) Predicting Conversion of Mild Cognitive Impairments to Alzheimer's Disease and Exploring Impact of Neuroimaging. *Lecture Notes in Computer Science*, 83–91. [https://doi.org/10.1007/978-3-030-00689-1\\_9](https://doi.org/10.1007/978-3-030-00689-1_9)
90. Slot R, Sikkes S, Berkhof J, Brodaty H, Buckley R, Cavado E, Dardiotis E, Guillo-Benarous F, Hampel H, Kochan N, Lista S, Luck T, Maruff P, Molinuevo J, Kornhuber J, Reisberg B, Riedel-Heller S, Risacher S, Roehr S, Sachdev P, Scarmeas N, Scheltens P, Shulman M, Saykin A, Verfaillie S, Visser P, Vos S, Wagner M, Wolfsgruber S, Jessen F, van der Flier W (2018) Subjective cognitive decline and

- rates of incident Alzheimer's disease and non-Alzheimer's disease dementia. *Alzheimer's Dementia* 15:465–476. <https://doi.org/10.1016/j.jalz.2018.10.003>
91. Sørensen L, Igel C, Pai A, Balas I, Anker C, Lillholm M, Nielsen M (2017) Differential diagnosis of mild cognitive impairment and Alzheimer's disease using structural MRI cortical thickness, hippocampal shape, hippocampal texture, and volumetry. *NeuroImage: Clinical* 13:470–482. <https://doi.org/10.1016/j.nicl.2016.11.025>
  92. Tang H, Yao E, Tan G, Guo X (2018) A Fast and Accurate 3D Fine-Tuning Convolutional Neural Network for Alzheimer's Disease Diagnosis. *Communications in Computer and Information Science*, 115–126. [https://doi.org/10.1007/978-981-13-2122-1\\_9](https://doi.org/10.1007/978-981-13-2122-1_9)
  93. Thapar D, Jaswal G, Nigam A, Arora C (2019) Gait metric learning siamese network exploiting dual of spatio-temporal 3d-CNN intra and LSTM based inter gait-cycle-segment features. *Pattern Recogn Lett* 125:646–653. <https://doi.org/10.1016/j.patrec.2019.07.008>
  94. Tipps M, Buck K (2015) GIRK Channels: a potential link between learning and addiction. *Int Rev Neurobiol*, 239–277. <https://doi.org/10.1016/bs.irm.2015.05.012>
  95. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI Single-Subject brain. *NeuroImage* 15:273–289. <https://doi.org/10.1006/nimg.2001.0978>
  96. Umamakeswari A, Angelus J, Kannan M, Rashikha BS (2020) Action Recognition Using 3D CNN and LSTM for Video Analytics. *Intelligent Computing and Communication* 531–539. [https://doi.org/10.1007/978-981-15-1084-7\\_51](https://doi.org/10.1007/978-981-15-1084-7_51)
  97. Vorugunti C, Pulabaigari V, Gorthi R, Mukherjee P (2020) OSVFUsetnet: Online Signature Verification by feature fusion and depth-wise separable convolution based deep learning. *Neurocomputing* 409:157–172. <https://doi.org/10.1016/j.neucom.2020.05.072>
  98. Wang H, Shen Y, Wang S, Xiao T, Deng L, Wang X, Zhao X (2019) Ensemble of 3D densely connected convolutional network for diagnosis of mild cognitive impairment and Alzheimer's disease. *Neurocomputing* 333:145–156. <https://doi.org/10.1016/j.neucom.2018.12.018>
  99. Wijnands J, Thompson J, Nice K, Aschwanden G, Stevenson M (2019) Real-time monitoring of driver drowsiness on mobile platforms using 3D neural networks. *Neural Computing and Applications*. <https://doi.org/10.1007/s00521-019-04506-0>
  100. Xu X, Caulfield S, Amaro J, Falcao G, Moloney D (2019) 1.2 Watt Classification of 3D Voxel Based Point-clouds using a CNN on a Neural Compute Stick. *Neurocomputing*. <https://doi.org/10.1016/j.neucom.2018.10.114>
  101. Xu Q, Wang Z, Wang F, Gong Y (2019) Multi-feature fusion CNNs for Drosophila embryo of interest detection. *Physica A Stat Mech Appl* 531:121808. <https://doi.org/10.1016/j.physa.2019.121808>
  102. Xu J, Zhang Z, Friedman T, Liang Y, Broeck G (2018) A semantic loss function for deep learning with symbolic knowledge. In: *Proceedings of machine learning research (PMLR)*. arXiv:1711.11157
  103. Yan Y, Lee H, Somer E, Grau V (2018) Generation of Amyloid PET Images via Conditional Adversarial Training for Predicting Progression to Alzheimer's Disease. *PRedictive Intelligence in Medicine* 26–33. [https://doi.org/10.1007/978-3-030-00320-3\\_4](https://doi.org/10.1007/978-3-030-00320-3_4)
  104. Yang H, Yuan C, Li B, Du Y, Xing J, Hu W, Maybank S (2019) Asymmetric 3D Convolutional Neural Networks for action recognition. *Pattern Recogn* 85:1–12. <https://doi.org/10.1016/j.patcog.2018.07.028>
  105. Yildirim Ö (2018) A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. *Comput Biol Med* 96:189–202. <https://doi.org/10.1016/j.combiomed.2018.03.016>
  106. Zhang K, Guo L, Gao C (2018) Optimization method of residual networks of residual networks for image classification. In: *2018 IEEE international conference on big data and smart computing (BigComp)*. <https://doi.org/10.1109/bigcomp.2018.00054>
  107. Zhang Y, Muhammad K, Tang C (2018) Twelve-layer deep convolutional neural network with stochastic pooling for tea category classification on GPU platform. *Multimed Tools Appl* 77(17):22821–22839. <https://doi.org/10.1007/s11042-018-5765-3>
  108. Zhao K, Ding Y, Han Y, Fan Y, Alexander-Bloch A, Han T, Jin D, Liu B, Lu J, Song C, Wang P, Wang D, Wang Q, Xu K, Yang H, Yao H, Zheng Y, Yu C, Zhou B, Zhang X, Zhou Y, Jiang T, Zhang X, Liu Y (2020) Independent and reproducible hippocampal radiomic biomarkers for multi-site Alzheimer's disease: diagnosis, longitudinal progress and biological basis. *Sci Bull* 65:1103–1113. <https://doi.org/10.1016/j.scib.2020.04.003>
  109. Zhi S, Liu Y, Li X, Guo Y (2018) Toward real-time 3D object recognition: a lightweight volumetric CNN framework using multitask learning. *Comput Graphics* 71:199–207. <https://doi.org/10.1016/j.cag.2017.10.007>

110. Zhou T, Ruan S, Canu S (2019) A review: Deep learning for medical image segmentation using multi-modality fusion. *Array* 3–4:100004. <https://doi.org/10.1016/j.array.2019.100004>
111. Zhu Z, Dai W, Hu Y, Li J (2020) Speech emotion recognition model based on bi-GRU and Focal Loss. *Pattern Recogn Lett* 140:358–365. <https://doi.org/10.1016/j.patrec.2020.11.009>

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

## Affiliations

Ibtissam Bakkouri<sup>1</sup>  · Karim Afdel<sup>1</sup> · Jenny Benois-Pineau<sup>2</sup> · Gwénaëlle Catheline<sup>3</sup>  
For the Alzheimer's Disease Neuroimaging Initiative

<sup>1</sup> Laboratory of Computer Systems and Vision (LabSIV), Department of Computer Science, Faculty of Science, Ibn Zohr University, BP 8106, 80000 Agadir, Morocco

<sup>2</sup> Bordeaux Laboratory for Computer Science Research (LaBRI) UMR 5800, University of Bordeaux, CNRS, Bordeaux INP, F-33400, Talence, Cedex, France

<sup>3</sup> Aquitaine Institute for Cognitive and Integrative Neuroscience (INICIA) UMR 5287, University Victor Segalen Bordeaux 2, CNRS, F-33076, Bordeaux, Cedex, France