Neural Networks 126 (2020) 218–234

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Brain MRI analysis using a deep learning based evolutionary approach

Hossein Shahamat, Mohammad Saniee Abadeh *

Faculty of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran

ARTICLE INFO

Article history:
Received 16 April 2019
Received in revised form 29 December 2019
Accepted 16 March 2020
Available online 28 March 2020

Keywords:
3D-CNN
Genetic algorithm
Deep learning
Interpretable classifier
Brain MRI classification

ABSTRACT

Convolutional neural network (CNN) models have recently demonstrated impressive performance in medical image analysis. However, there is no clear understanding of why they perform so well, or what they have learned. In this paper, a three-dimensional convolutional neural network (3D-CNN) is employed to classify brain MRI scans into two predefined groups. In addition, a genetic algorithm based brain masking (GABM) method is proposed as a visualization technique that provides new insights into the function of the 3D-CNN. The proposed GABM method consists of two main steps. In the first step, a set of brain MRI scans is used to train the 3D-CNN. In the second step, a genetic algorithm (GA) is applied to discover knowledgeable brain regions in the MRI scans. The knowledgeable regions are those areas of the brain which the 3D-CNN has mostly used to extract important and discriminative features from them. For applying GA on the brain MRI scans, a new chromosome encoding approach is proposed. The proposed framework has been evaluated using ADNI (including 140 subjects for Alzheimer's disease classification) and ABIDE (including 1000 subjects for Autism classification) brain MRI datasets. Experimental results show a 5-fold classification accuracy of 0.85 for the ADNI dataset and 0.70 for the ABIDE dataset. The proposed GABM method has extracted 6 to 65 knowledgeable brain regions in ADNI dataset (and 15 to 75 knowledgeable brain regions in ABIDE dataset). These regions are interpreted as the segments of the brain which are mostly used by the 3D-CNN to extract features for brain disease classification. Experimental results show that besides the model interpretability, the proposed GABM method has increased final performance of the classification model in some cases with respect to model parameters.

1. Introduction

Deep learning models are composed of multiple processing layers to learn representations of the data with multiple levels of abstraction. These methods discover complicated structures in large data sets. Thus, they have dramatically improved the state-of-the-art in many fields of machine learning. Convolutional neural networks (CNNs) are a kind of deep learning methods, which have shown very good performance in processing images, video, speech and audio (LeCun, Bengio, & Hinton, 2015). CNNs are a type of representation-learning methods which can automatically identify the optimal representation from the row data without requiring prior feature selection (called end-to-end) (Vieira, Pinaya, & Mechelli, 2017). The end-to-end learning strategy makes CNN representations a black box and except for the final network output layer, it is difficult to understand the logic of the CNN predictions hidden inside the network. Deep learning algorithms, particularly CNNs, have rapidly became a methodology of choice for analyzing medical images (Litjens et al., 2017). Deep learning models can be used for medical image classification (Hosseini-Asl et al., 2018; Zhang et al., 2019), object detection (de Vos, Wolterink, de Jong, Viergever, & Isgum, 2016; Yang et al., 2015), segmentation (Çiçek, Abdulkadir, Lienkamp, Brox, & Ronneberger, 2016; Lu et al., 2019), registration (Cheng, Zhang, & Zheng, 2016; García et al., 2019), and other tasks (Anavi, Kogan, Gelbart, Geva, & Greenspan, 2016; Liu, Tizhoosh, & Kofman, 2016).

Neuroimaging technology has widely been used in the study of various brain diseases, such as autism spectrum disorder (ASD) (Ecker et al., 2010; Khosla, Jamison, Kuceyeski, & Sabuncu, 2018), Alzheimer (Liu, Li et al., 2018; Liu, Wang et al., 2017), and schizophrenia (Liu, Li et al., 2017; Liu, Wang, Zhang et al., 2017). In the recent years, there has been a growing trend in designing neuroimaging-based diagnostic tools to automatically classify patients from controls (Klöppel et al., 2012). In this regard, the machine learning algorithms have been successfully employed in the automated classification of magnetic resonance imaging (MRI) data. MRI is a powerful, widely used and non-invasive tool, which produces high quality 3D images of the brain structures (Kong et al., 2018). One of the challenges of applying deep learning techniques to the neuro imaging data is related to this new 3D data format (3D volumes). In comparison with the designed models for 2D data, these 3D models need a large amount of

* Corresponding author.
E-mail address: saniee@modares.ac.ir (M. Saniee Abadeh).

https://doi.org/10.1016/j.neunet.2020.03.017
0893-6080/© 2020 Elsevier Ltd. All rights reserved.

parameters. Prasoon et al. (2013) circumvented this problem by dividing the 3D volumes into 2D slices which are fed as different streams to a 2D network. Due to the 3D format of medical data, a full 3D CNN model (instead of 2D) can be used to classify patients vs. normal controls (NC) (Hosseini-Asl et al., 2018; Payan & Montana, 2015).

In addition to perform an accurate classification, interpreting and visualizing CNNs are important tasks to increase trust in automatic classification systems (Rieke, Eitel, Weygandt, Haynes, & Ritter, 2018). By interpreting and visualizing a 3D-CNN model which is trained using a number of MRI scans, the brain regions with sufficiently large discrimination power can be highlighted. These regions are those segments of the brain which are mostly used by the 3D-CNN to extract important features for classification task. In this paper, these regions are called knowledgeable brain regions. Knowledgeable brain regions are those areas of the brain which the 3D-CNN has mostly used to extract discriminative features from them. These regions can be interpreted as the most important brain regions in the classification task under the study. Knowledgeable brain regions could be interested for studying the diseases progress and monitoring the effect of treatments.

In this paper, a 3D-CNN model is designed for classification of brain MRI scans into two predefined groups (Patients vs. Normals). CNNs can identify the optimal representation from the row data but they are a type of black box methods. For interpreting the trained black box model, a novel genetic algorithm based brain masking (GABM) method is proposed. Besides, a new chromosome encoding technique is suggested for applying GA on the brain MRI scans. This proposed framework can discover the brain regions with sufficiently large discrimination power and report them as knowledgeable brain regions in the disease under study.

The rest of this paper is organized as follows. Section 2 reviews some state of the art methods. Section 3 explains our datasets and preprocessing step, a 3D-CNN model for MRI data classification, and the details of the proposed GABM method (chromosome encoding, fitness function, operators and parameters). Section 4 consists of the experimental results and discussion. Finally, Section 5 presents our conclusion.

2. Related works

Related works that are closely connected to this study are divided into three parts: Alzheimer classification, Autism classification, and CNN visualization. In this section some state of the art methods are reviewed for each part.

2.1. Alzheimer classification

The automatic classification of Alzheimer's disease (AD) using MRI data plays an important role in human health. Therefore, many researchers employed image classification methods to perform AD diagnosis (Beheshti, Demirel, & Initiative, 2016; Khedher et al., 2015; Long, Chen, Jiang, Zhang, & Initiative, 2017). In contrast to automatic feature extraction in deep learning approaches, the mentioned works need to extract features manually. On the other hand, many deep learning methods have been proposed to perform AD classification using MRI data (Li et al., 2015; Liu, Pan et al., 2018; Payan & Montana, 2015; Sarraf, DeSouza, Anderson, & Tofighi, 2017). Similar to present paper, some studies applied a 3D-CNN model on full brain structural MRI scans in order to AD classification. Korolev, Safiullin, Belyaev, and Dodonova (2017) applied a convolutional neural network for AD classification on a subset of 111 MRI scans from ADNI dataset (including 50 AD subjects and 61 NC subjects) and achieved an accuracy of 0.80. Rieke et al. (2018) obtained an accuracy of 0.77 in classification of 969 MRI scans (475 AD, 494 NC) from 344 subjects (193 AD, 151 NC). Yang, Rangarajan, and Ranka (2018) applied a 3D-CNN model on 103 brain MRI scans (47 AD and 56 NC) for AD classification. Using a 3D-ResNet model and a 5-fold cross validation strategy, they achieved an accuracy of 0.79.

2.2. Autism classification

Blumberg et al. (2013) reported 1 in 55 children aged 6–17 years are identified as patients with autism spectrum disorder (ASD). Thus, making an accurate diagnosis seems crucial for societies (Fein et al., 2013; Li, Karnath, & Xu, 2017). Many papers worked on ASD classification using brain MRI data and machine learning techniques (Bernas, Aldenkamp, & Zinger, 2018; Jung et al., 2017; Plitt, Barnes, & Martin, 2015; Tejwani, Liska, You, Reinen, & Das, 2017). They reported high classification accuracy, but used small datasets and it is hard to utilize these methods on other datasets. Dealing with large datasets, Sabuncu, Konukoglu, and Initiative (2015) worked on 935 MRI scans from the ABIDE dataset for ASD classification. They reported a classification accuracy of 0.60 under a 5-fold cross validation strategy. Monté-Rubio, Falcón, Pomarol-Clotet, and Ashburner (2018) worked on 1102 MRI scans from the ABIDE dataset. They explored several feature extraction methods and two types of classifiers for MRI classification. The maximum reported accuracy using a 5-fold cross validation mode is 0.62. Dvornek, Ventola, Pelphrey, and Duncan (2017) proposed a deep learning framework for ASD classification using functional MRI (f-MRI) data. They used entire ABIDE dataset and obtained a classification accuracy of 0.69. A methodology for incorporating phenotypic data with f-MRI data into a single deep learning framework has been proposed by Dvornek, Ventola, and Duncan (2018) which shown an accuracy of 0.70. Heinsfeld, Franco, Craddock, Buchweitz, and Meneguzzi (2018) applied a deep learning model to 964 f-MRI scans and achieved an accuracy of 0.70. Li, Parikh, and He (2018) applied deep learning on brain functional connectomes for ASD classification and achieved an accuracy of 0.70 by developing a deep transfer learning neural network framework.

2.3. CNN visualization

In the recent years, a growing number of researchers have realized that the CNNs model interpretability is an important issue and they have developed models with interpretable knowledge representation (Ventura, Masip, & Lapedriza, 2017; Zhang, Cao, Shi, Wu, & Zhu, 2017; Zhang & Zhu, 2018). One strategy for understanding and visualizing CNNs is to show the activations of the network during the forward pass. This is the most straightforward visualization technique. The second common approach is visualizing the network weights. Another technique is retrieving images that maximally activate a neuron. This technique needs taking a large dataset of images, feed them through the network and keep track of which images maximally activate some neurons. Then the images should be visualized to get an understanding of what the neuron is looking for in its receptive field (Girshick, Donahue, Darrell, & Malik, 2014). Another interesting method for understanding CNNs is occluding parts of the images. For investigating those parts of the image that a classification prediction is coming from, the probability of the class of interest should be plotted (as a function of the position of occluded part). This process is repeated over all regions of the image while looking at the probability of each class. Finally, each class probability is visualized as a 2D heat map (Zeiler & Fergus, 2014). This heat map shows important parts of the images according to the classification problem. In some classification domains, like hyperspectral image (HSI) classification (Wang, He,

& Li, 2018) and scene classification in VHR remote sensing images (Wang, Liu, Chanussot, & Li, 2018), the input data usually have a great number of spectral bands or a very high spatial resolution. Training a CNN and occluding input images or ignoring some spectral bands, and keep track of those parts of the images or spectral bands that maximally activate the neuron of the class of interest, can help to understand what the network is looking for. Focusing on these parts can help researchers to reduce negative effect of redundant areas and train models that are more efficient in terms of accuracy and computational complexity. For reducing the computational complexity, Wang, Liu, et al. (2018) applied an attention mask on intermediate data, after a set of convolution layers. This structure discarded non-critical information, improved the classification performance, and reduced the computational complexity at the same time. Visualizing a CNN model using image parts occluding approaches, can also be applied in 3D spaces. Huang et al. (2018) used subsets of video frames as 3D data for training a 3D-CNN model for dynamic scene classification. In this method, a given video is split into 16-frame long clips with a 15-frame overlap between two consecutive clips. Visualizing their network using a 3D image occluding approach can be useful to discover more interesting frames or more important areas of these frames. For interpreting and visualizing a 3D-CNN model trained by MRI data, Yang et al. (2018) proposed a segmentation based occlusion approach for sensitivity analysis of 3D-CNNs, which can identify the important brain regions involved in AD classification at different levels. A 3D-CNN model has been proposed to detect AD using brain structural MRI scans by Rieke et al. (2018). They introduced a brain area occlusion-based visualization method to highlight relevant brain areas in the input image. These papers explained the CNN decision for one specific sample at time. Furthermore, because of occluding one brain region at time, the correlations and interactions between brain regions may be ignored. The present paper purposes a new CNN visualization technique to discover important brain regions (knowledgeable brain regions) in a particular brain MRI classification task. This is done using a combination of an atlas based brain region occluding method and a genetic algorithm feature selection method (with a new chromosome encoding scheme). The proposed method explains the 3D-CNN decision for all training samples at time. Moreover, because of the nature of GA-based feature (region) selection methods the correlations between brain regions will be considered.

3. Material and methods

An overview of the proposed framework is summarized in Fig. 1. It consists of four major steps: (1) preprocessing, (2) classification, (3) genetic algorithm based brain masking (identification of knowledgeable brain regions), and (4) experimental results. The steps 1, 2, and 3 are described in the present section and step 4 is described in the next section.

3.1. Data and preprocessing

The Autism Brain Imaging Data Exchange I (ABIDE I) involved 17 international sites, sharing previously collected 1112 (MRI and fMRI) scans, including 539 individuals with ASD and 573 normal control subjects (ages 7–64 years, median 14.7 years across groups). Here, only the MRI scans collected from 1000 individuals (500 ASD and 500 NC) are used to evaluate our proposed method. For more details with the ABIDE I, please see http://fcon_1000.projects.nitrc.org/indi/abide/.

The Alzheimer's Disease Neuroimaging Initiative (ADNI) is a longitudinal multisite observational study of normal control, mild cognitive impairment (MCI), and Alzheimer's disease (AD) (Jack et al., 2008). ADNI researchers collect, validate and utilize data, including MRI and PET images, genetics, cognitive tests, CSF and blood biomarkers as predictors of the disease. In this paper, a set of 140 MRI scans (70 NC and 70 AD) has been downloaded from ADNI site (http://adni.loni.usc.edu/) and is used in our experiments.

All MRI scans are preprocessed using FSL software (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki). All scans are registered to MNI152_T1_2_mm standard space. After normalization, all MRI scans will have a size of 91 × 109 × 91 voxels. The scans are cropped to an 80×80×80 voxels sub-volume with the brain centered. This reduces the empty slices with non-valuable information from the margins.

3.2. Classification using 3D-CNN

In machine learning, a CNN is a class of deep learning generally for image analysis. The CNN models are based on the artificial neural networks. These models can automatically identify the optimal representation from the row data without requiring prior feature selection (Vieira et al., 2017). In this paper, a 3D-CNN model is designed and trained using a number of MRI scans. Then, the trained model is employed to predict the class label of a given MRI scan ("Patient" or "Normal"). The results show the efficiency of the proposed architecture on different brain MRI datasets. The process of building our 3D-CNN model involves these major steps: convolution, rectified linear unit (ReLU), pooling, dropout, flattening and fully connection (with sigmoid activation function). An overall view of the proposed 3D-CNN architecture for MRI classification is shown in Fig. 2.

3.3. Genetic algorithm based brain masking

In the first step of the proposed genetic algorithm based brain masking (GABM) method, 96 predefined brain regions are extracted from Harvard– Oxford cortical and subcortical structural atlas (see Section 3.3.1). These regions can be used as a mask to select the voxels inside them. Also, a mask can be generated using more than one brain region. Different subsets of the regions give different brain masks and have different effect on the system output. Analyzing the classifier output, by occluding some portions of the input scans using generated masks, reveals which regions of the brain are important for classification. As a result, the regions in the mask with best performance can be considered as knowledgeable brain regions. For Harvard–Oxford atlas with 96 predefined brain regions, there exist 2^{96} subsets of the regions. Thus, finding a subset of brain regions with sufficiently large discrimination power leads to a very large search space. Genetic algorithm (GA) is very effective in solving large scale problems and can be used to find an optimal (or near optimal) solution. GA starts with random population of trial solutions called individuals or chromosomes. In this paper, each individual would represent a brain mask which contains a subset of predefined brain regions. Normally, the quality of each individual is evaluated using a fitness function with respect to some measures of interest. Finally, GA finds the optimal solution through repetitive application of genetic operations on the chromosomes. More detailed descriptions of the proposed GA components are explained in next subsections. Fig. 3 demonstrates the overall diagram of the proposed GABM method.

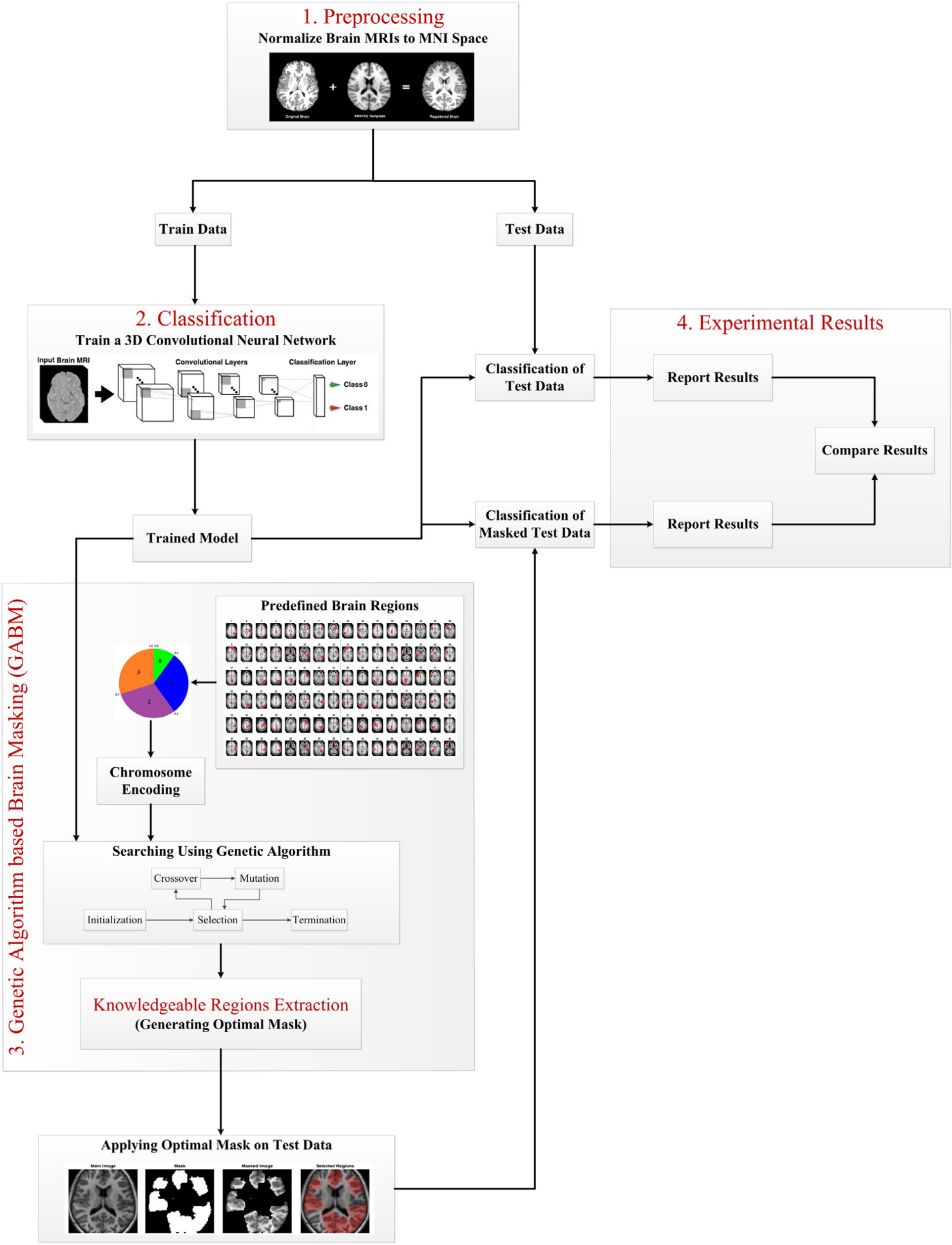


Fig. 1. Overview of the proposed framework.

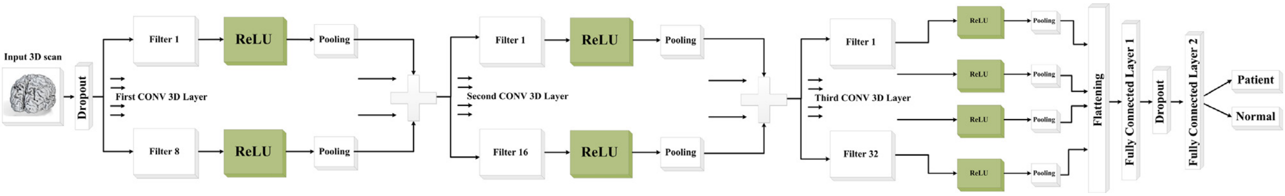


Fig. 2. The proposed 3D-CNN architecture for MRI classification.

3.3.1. *Brain regions extraction*

In the evaluation step of the proposed GABM method, we analyze the contribution of a subset of brain regions in the classification performance. In this regard, the MRI scans are parcellated into 96 predefined brain regions using Harvard– Oxford cortical and subcortical structural atlas. All predefined brain regions are extracted from the mentioned atlas (with threshold= 0) and they are stored in 3D matrices in the same size as the input scans. These 3D data or a subset of them can be used as brain masks to select the voxels inside them. The predefined brain regions are shown in Fig. 4. In this study, the knowledgeable brain regions are discovered using a GA based brain masking process. First, several brain masks are generated randomly and then will be evaluated via a fitness function. Then, the best mask will be optimized during GA generations. For applying a brain mask on the MRI scans, an element-wise multiplication operator is used.

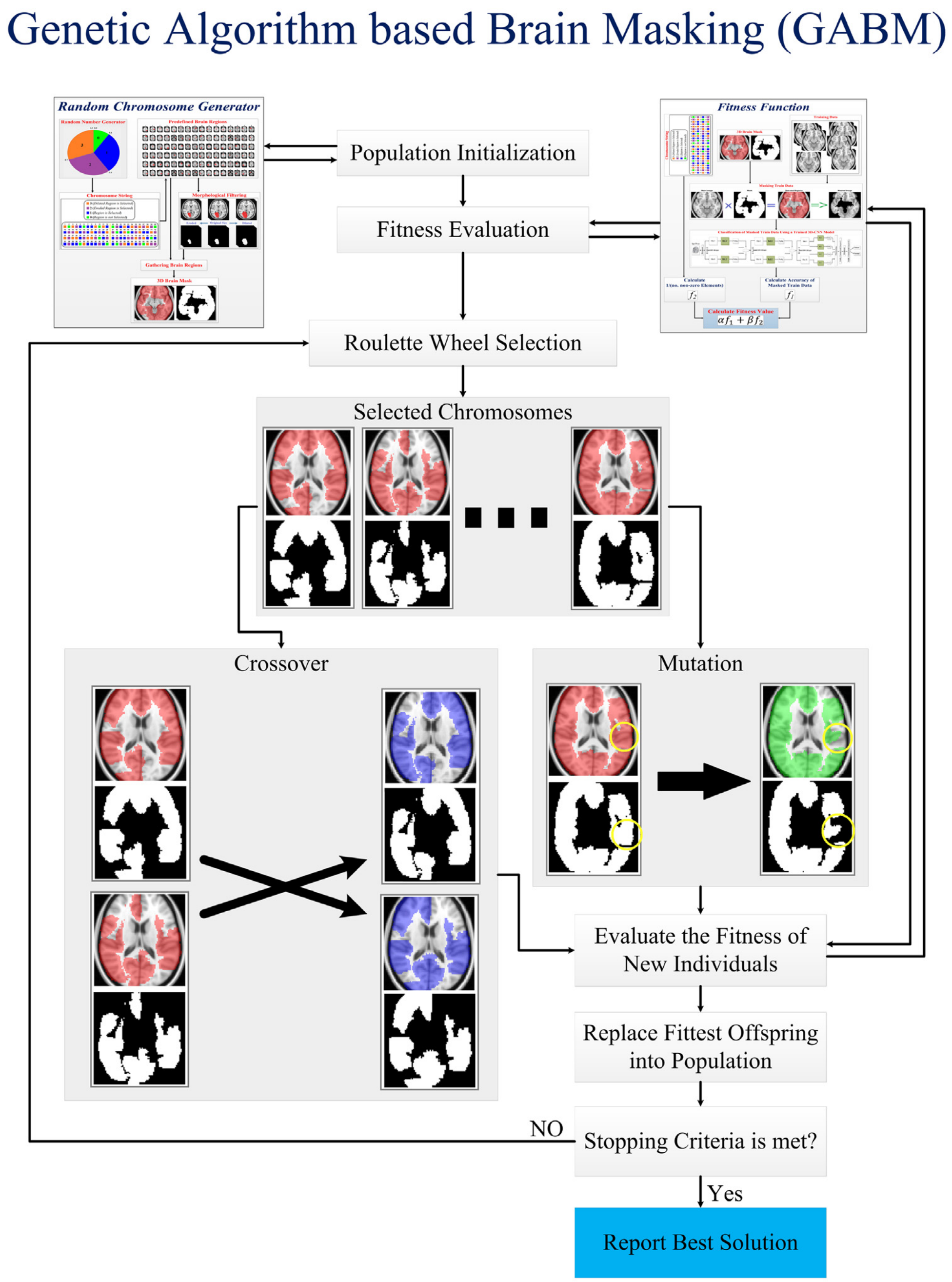

Genetic Algorithm based Brain Masking (GABM)
Random Chromosome Generator
Fitness Function
Population Initialization
Fitness Evaluation
Roulette Wheel Selection
Selected Chromosomes
Crossover
Mutation
Evaluate the Fitness of New Individuals
Replace Fittest Offspring into Population
NO
Stopping Criteria is met?
Yes
Report Best Solution

Fig. 3. The overall diagram of the proposed GABM method.

Table 1 shows the predefined brain regions' names and their corresponding IDs. After chromosome encoding, the ID field will be equal to the brain region location in the chromosome.

Table 1
Data structure including ID and Name of all predefined regions.

ID	Name
1	Left Angular Gyrus
2	Left Central Opercular Cortex
3	Left Cingulate Gyrus, anterior division
...	...
47	Left Temporal Occipital Fusiform Cortex
48	Left Temporal Pole
49	Right Angular Gyrus
50	Right Central Opercular Cortex
...	...
94	Right Temporal Fusiform Cortex, posterior division
95	Right Temporal Occipital Fusiform Cortex
96	Right Temporal Pole

3.3.2. *Chromosome encoding*

In this paper, a new chromosome encoding scheme is proposed to discover knowledgeable brain regions in a particular MRI classification problem. Each chromosome in the population represents a candidate solution or a brain mask. If $\boldsymbol{m}$ is the total number of the brain regions (here, $\boldsymbol{m}$ = 96), each chromosome is represented by a vector of dimension $\boldsymbol{m}$. Instead of using a binary representation, which is the simplest chromosome encoding scheme, we define a chromosome as a vector of integer values picked from 0, 1, 2, or 3. These values are known as genes. In this paper, when a gene value is "0" its corresponding brain region is not selected to generate the brain mask. If a gene value is "1"

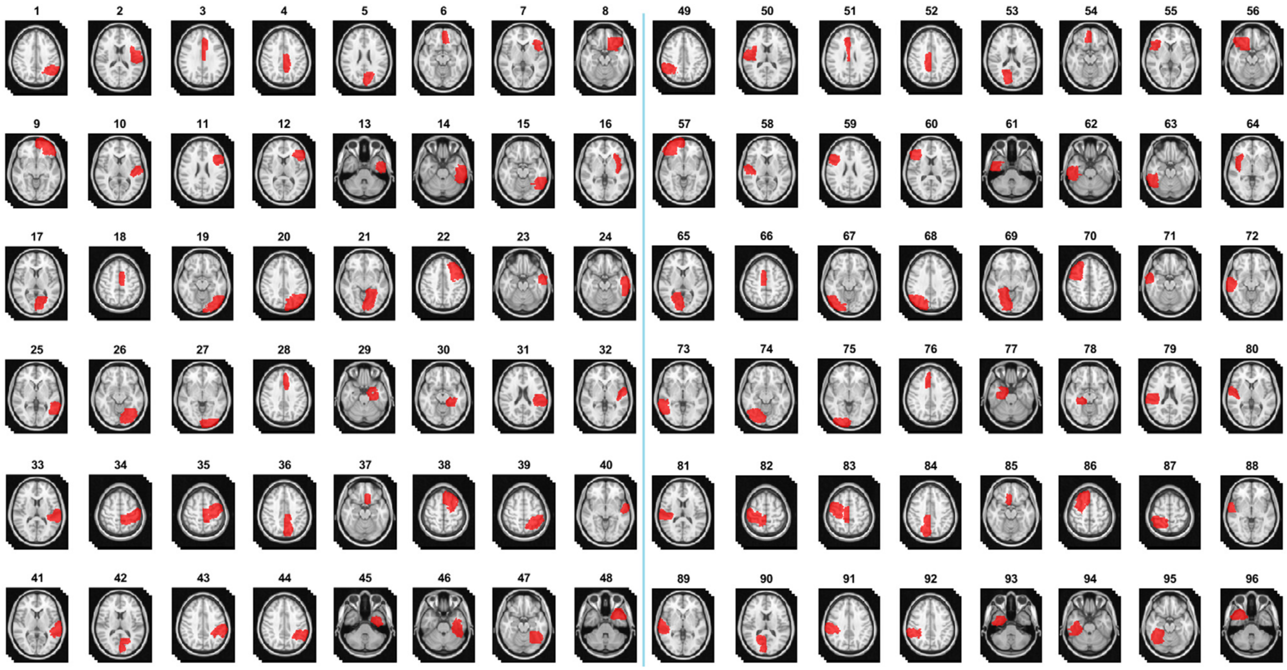


Fig. 4. All brain regions masks extracted from the Harvard–Oxford cortical and subcortical structural atlases.

its corresponding brain region is selected in its original shape. If a gene value is equal to “2” its corresponding brain region is shrunk by a morphological erosion operator. Erosion shrinks objects by etching away their boundaries (Gonzalez & Woods, 2002). Finally, if a gene’s value is equal to “3” its corresponding region is expanded using a morphological dilation operator. Fig. 5 shows the overall diagram of generating a random chromosome. Moreover, an example of applying the binary morphological filters on a particular brain region (Right Intracalcarine Cortex) is presented in Fig. 5. As can be seen the morphological erosion operator causes to exclude partial volume edges from the mask. Also, using the morphological dilation operator the mask will include outer area of the edges. These operators’ behaviors are very similar to changing threshold in the brain region extraction step. Using this new chromosome encoding framework, the boundaries of the predefined brain regions can be explored more accurately. This step can be interpreted as a local search step for the GA to find more optimal solutions.

3.3.3. Initial population

The first step of GA is generating an initial population randomly. In this study, each chromosome is created by randomly chosen values from set {0, 1, 2, 3}. In the population initialization step (also in the mutation step), the values are assigned to the genes using a predefined probability according to Eq. (1). For any particular gene G, a random real number R is generated in the range of (0, 1). Then, a value V is assigned to the gene G as follows:

$$V = \begin{cases} 0 & \mathit{if}\ 0 \le R < 0.1, \\ 1 & \mathit{if}\ 0.1 \le R < 0.4, \\ 2 & \mathit{if}\ 0.4 \le R < 0.7, \\ 3 & \mathit{if}\ 0.7 \le R < 1 \end{cases} \tag{1}$$

This probability confirms that the algorithm will select about 90% of the predefined brain regions for generation (and also mutation) of brain masks. Finally, it should be noted that the number of chromosomes in the initial population is an important issue for GA performance. A large population size leads to more genetic diversity but suffers from slower convergence. A very small population explores only a reduced part of the search space and it may converge to a local optima. This paper uses 200 individuals for initial population. Fig. 6 shows a randomly generated individual and its related brain mask.

3.3.4. Fitness function

In GA based approaches, fitness functions are used to evaluate the quality of the individuals. In this paper, an individual represents a brain mask and its quality is measured with respect to the accuracy of the model on the masked MRI scans and the inverse of the number of selected regions. Computing the fitness value of a chromosome contains two main steps. First, a 3D brain mask should be generated using the chromosome string. Next, the generated mask is applied on all input scans and the model accuracy will be calculated. Finally, the fitness value should be calculated using a weighted sum of the model accuracy and the inverse of the number of selected brain regions. By applying a mask on the brain MRI scans (in a voxel-wise multiplication mode), some regions will be suppressed and the other regions will be used for classification. As mentioned, a brain MRI scan is parcellated into 96 predefined regions. Thus, for a particular chromosome CH with 96 genes, the corresponding brain mask is generated as follows:

$$Msk = Zeros\_like(MRIscans) \tag{2}$$

where, we define an empty (initialized with zero) matrix MSK in the same size as the input MRI scans. MSK will be used to collect all selected brain regions into a single mask.

$$\mathit{if}\ CH_i > 0\ \mathit{then},\ \ Msk = Logical.OR\left(Msk, MF(BRM_i)\right). \qquad \mathit{for}\ i = 1..96 \tag{3}$$

where, MF is a function which applies a morphological filter on a 3D brain region mask based on its corresponding gene’s value (only for $CH_i > 1$). BRM is a set of 3D matrices containing predefined brain regions extracted from the Harvard– Oxford

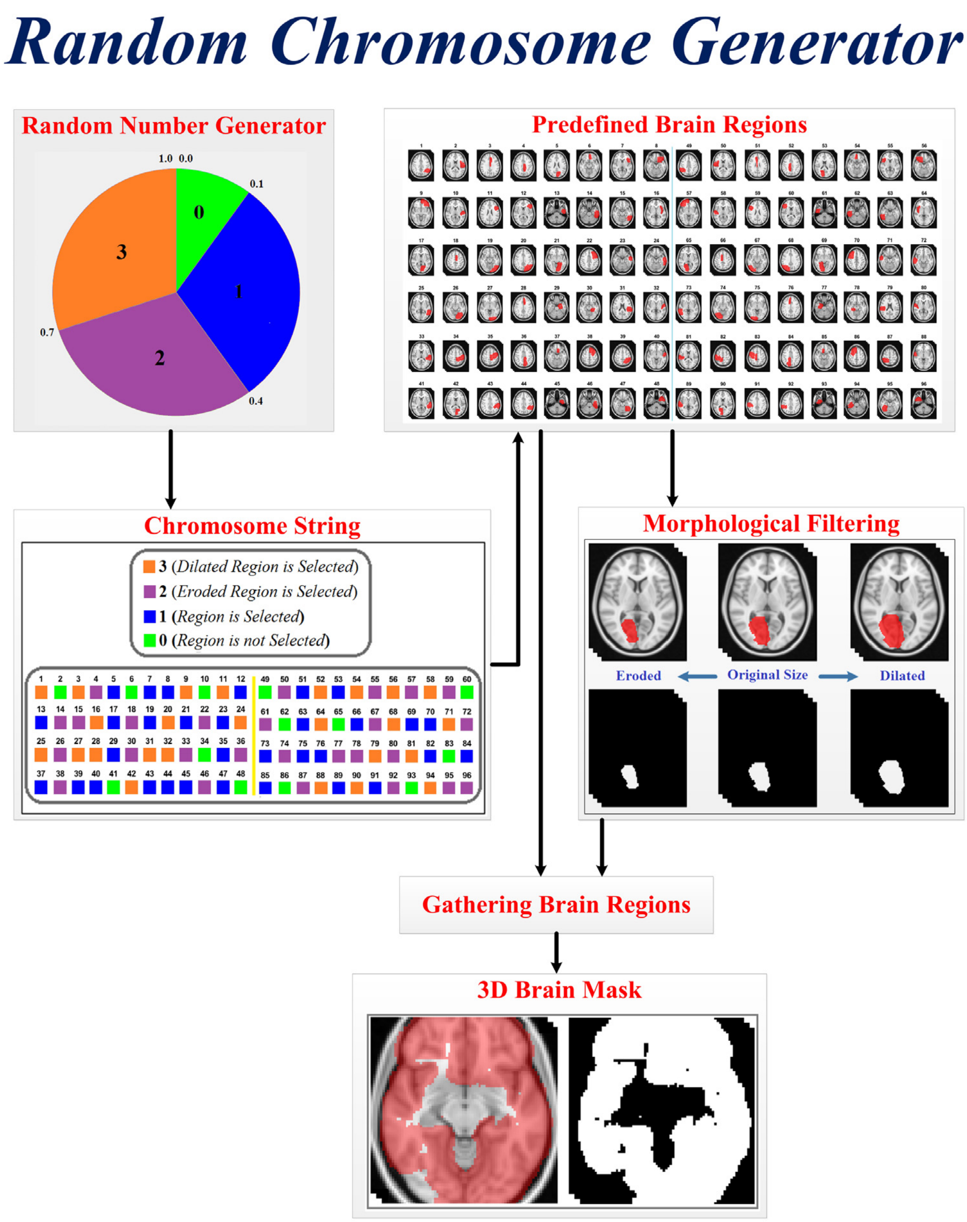


Fig. 5. The overall diagram of generation a random chromosome.

atlas. The *Logical-OR* function gives logical-or of two 3D matrices in a voxel-wise mode. This equation combines all selected brain regions masks to form a unified 3D brain mask. After converting a chromosome string to a brain mask, the classification accuracy of the trained 3D-CNN model should be calculated for the occluded versions of input scans. The classification accuracy is obtained according to Eq. (4):

$$f_1 = \frac{N_{correct}}{N_{total}} \tag{4}$$

where, f_1 denotes the classification accuracy, $N_{correct}$ is the number of correctly classified MRI scans and N_{total} is the total number of MRI scans. Additionally, to calculating the fitness value of a brain mask, we define a variable f_2 which denotes the inverse of the number of selected brain regions. Here, f_2 can be calculated as follows:

$$f_2 = \frac{1}{\sum_{i=1}^{96} CH_i > 0} \tag{5}$$

Finally, the fitness value for a particular solution s_i is calculated as follows:

$$Fitval(s_i) = \alpha f_1 + \beta f_2 \tag{6}$$

where, α and β are weighting parameters to balance between accuracy and the number of selected brain regions. Fig. 7 shows the overall diagram of the proposed fitness function. we consider $\alpha + \beta = 1$ and investigate different values for them to achieve different results. In this paper, GA is used to find an optimal brain mask (best chromosome) with highest classification accuracy and lowest number of selected regions. These regions can be interpreted as the most important brain regions in the classification problem under the study.

3.3.5. Operators and parameters of genetic algorithm

Selection method: In this paper, the roulette wheel method is used to select a number of individuals to be parents for later breeding. In the roulette wheel selection method, the probability

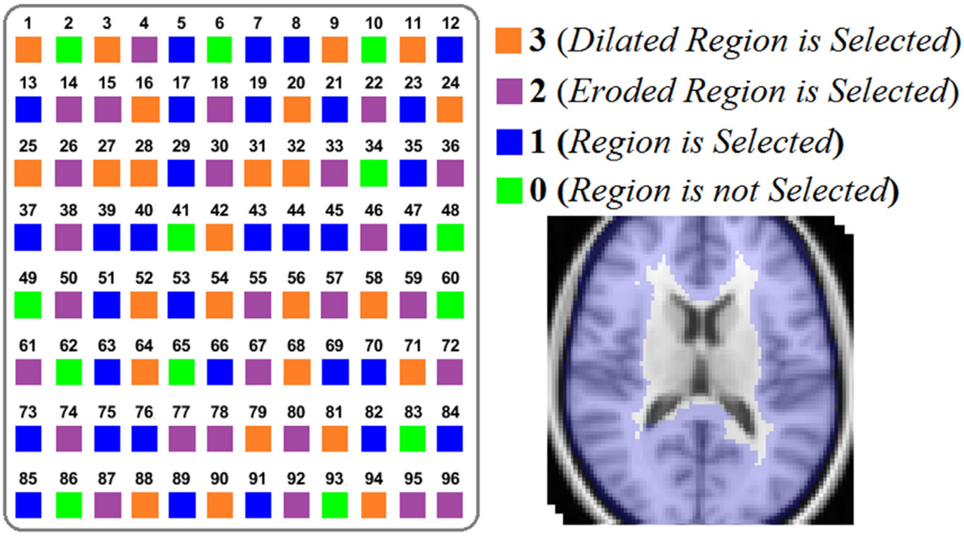


Fig. 6. A randomly generated chromosome and its corresponding brain masks.

Table 2
The parameters of the proposed 3D-CNN architecture.

Layer type	Kernel size	#Filters	Scan size
Input	––	––	80×80×80×1
Dropout	keep probability = 50%	––	80×80×80×1
Convolution	5×5×5	8	80×80×80×8
ReLU	––	––	80×80×80×8
Max Pooling	2×2×2	––	40×40×40×8
Convolution	3×3×3	16	40×40×40×16
ReLU	––	––	40×40×40×16
Max Pooling	2×2×2	––	20×20×20×16
Convolution	3×3×3	32	20×20×20×32
ReLU	––	––	20×20×20×32
Max Pooling	2×2×2	––	10×10×10×32
Flattening	––	––	1×32000
Fully connected	32000×1024	––	1×1024
Dropout	keep probability = 50%	––	1×1024
Fully connected	1024×2	––	1×2
Softmax Layer	––	––	1×2
Classification Layer	––	––	Patient vs. Normal

of selecting an individual s_i is given by:

$$P(s_i) = \frac{F(s_i)}{\sum_{j=1}^{n} F(s_j)} \tag{7}$$

where, $F(s)$ is the fitness value of the individual s, and n indicates the number of the population. The probability of selecting an individual is related to its own fitness and the fitness of the other competing individuals in the population (Tan, Fu, Zhang, & Bourgeois, 2008).

Crossover: A random single point crossover strategy is used to generate new solutions. This needs a crossover point ***i*** which is chosen randomly over the individuals' length. A new offspring will be created using first ***i*** genes of one parent and the remaining genes of the other parent.

Mutation: In the mutation step, a new value is assigned to a randomly chosen gene for all selected individuals. The new value is generated based on our predefined probability map.

GA parameters: Finally, the other GA parameters have been chosen as ***Population size***: 200, ***Number of generations***: 2000, ***Probability of crossover***: 0.4, and ***Probability of mutation***: 0.6.

4. Experimental results and discussion

In this paper, a 3D-CNN model has been designed and trained from the scratch. The input layer of this model has a size of 80×80×80 to accept preprocessed MRI scans. This layer is followed by a dropout layer with keep probability 50% to reduce over fitting. The first convolutional layer consists of 8 filters with size 5×5×5. After applying a ReLU activation function to the convolution's results, a max-pooling operator is used with window size 2×2×2. It reduces the input scan size to 40×40×40. The second convolutional layer has 16 filters with size 3×3×3. After applying Relu function and max-pooling operator, the data size is reduced to 20×20×20. The third convolutional layer has 32 filters with size 3×3×3. After applying Relu and pooling on the results, the data size is reduced to 10×10×10. Subsequently, two fully connected (FC) layers are used for data classification. The first FC layer has 32000 input and 1024 output neurons. This FC layer is followed by a dropout layer with keep probability 50%. The second FC layer has 1024 input and 2 output neurons (same as the number of classes). Finally, a softmax layer and a classification layer are used to provide labels for the input MRI scans. We trained this model with a cross-entropy loss function and the Adadelta optimizer (learning rate 0.05, decay rate 0.95, batch size 32 for ADNI and 64 for ABIDE, and 30000 iterations). All of the network parameters have been summarized in Table 2. The proposed 3D-CNN model has been evaluated using two different brain MRI datasets. In a 5-fold cross validation mode, the classification accuracy for the ADNI dataset was 0.85 and for the ABIDE dataset was 0.70. Fig. 8 shows the training accuracy during 3D-CNN optimization (average of all folds).

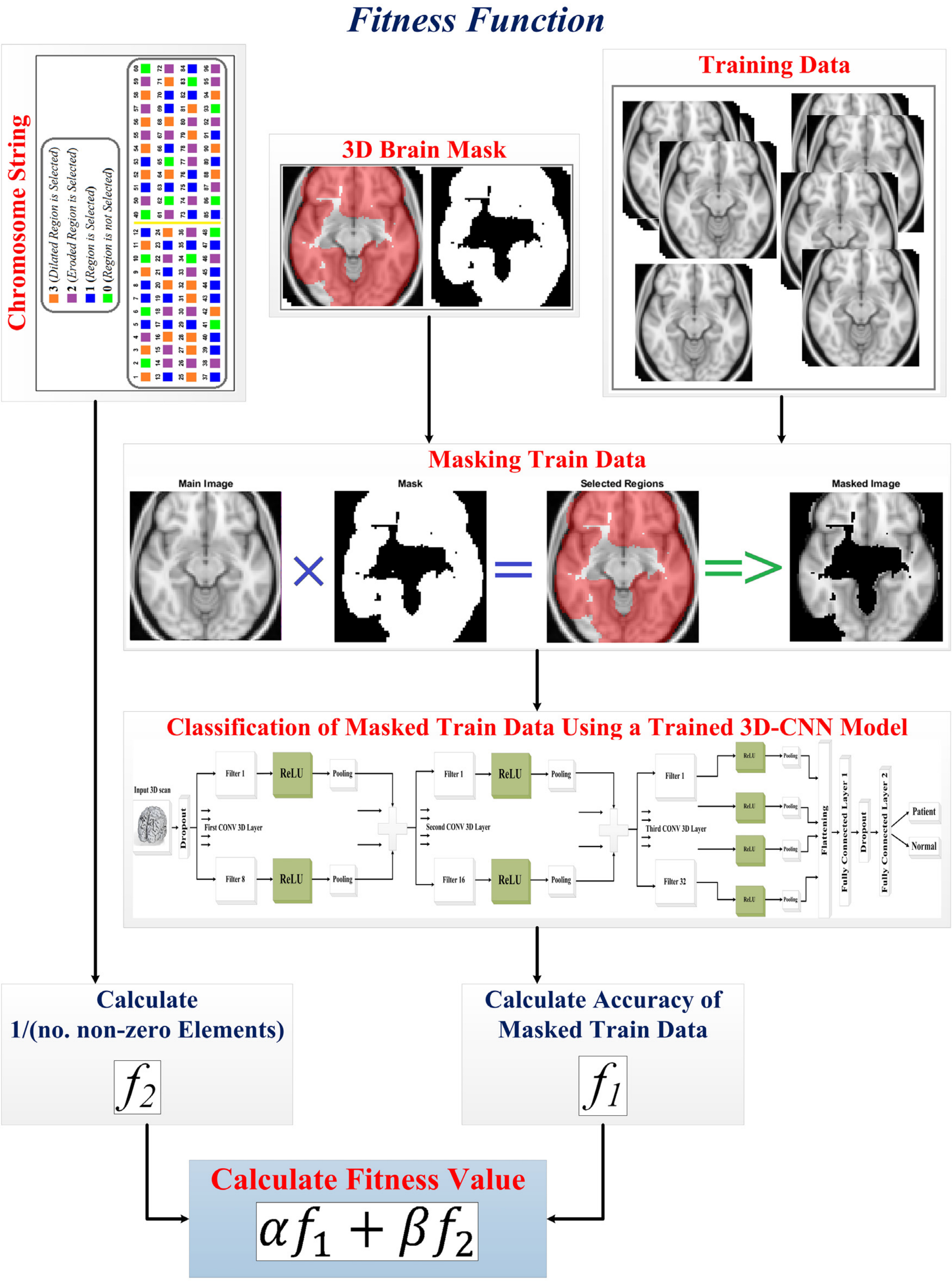


Fig. 7. The overall diagram of the proposed fitness function.

For identification of knowledgeable brain regions using the proposed GABM method, different values for α and β were investigated and different results have been achieved. The results of 4 different experiments are reported for both datasets. In the first experiment using the ADNI dataset, model parameters were selected as $\alpha = 0.99$ and $\beta = 0.01$. Any change in these parameters will affect the number of involved brain regions in the final mask. In all generations of the GABM method, best individual is applied on training data. Fig. 9 shows the classification accuracy (using best individuals) and the numbers of selected brain regions during 2000 GABM generations in some experiments (average over all folds). Fig. 9(b) shows the results of first experiment which has $\alpha = 0.99$ and $\beta = 0.01$. After GABM optimization, the classification accuracy using final brain mask (final best individual), was about 1.00 on the training data and 0.80 on the test data. These results have been achieved using 65 brain regions which can be reported as knowledgeable brain regions. This can be mentioned by removing 31 brain regions (about 1/3 of whole

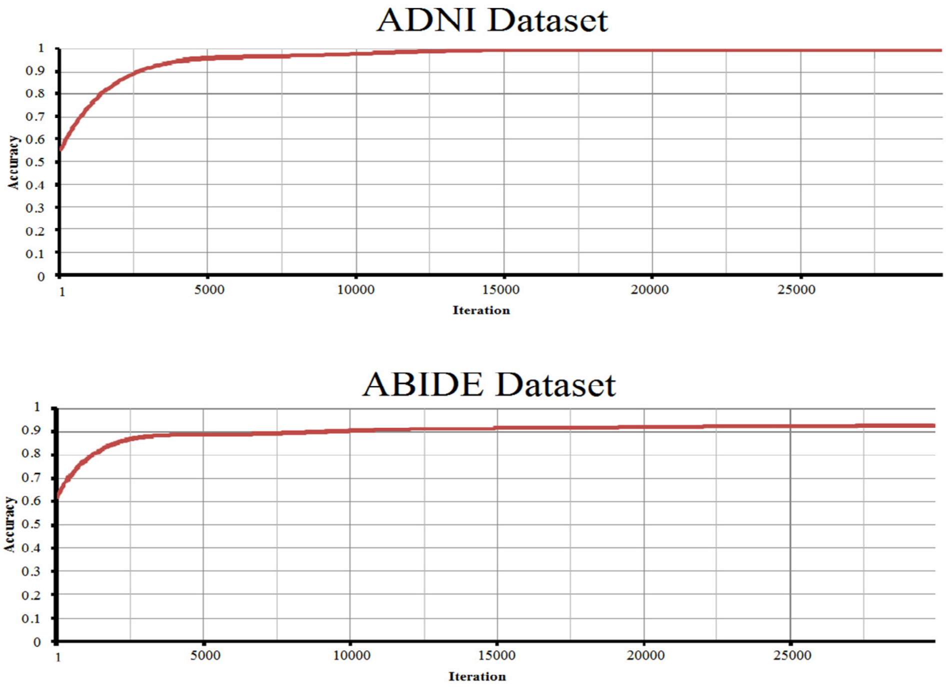


Fig. 8. Classification accuracy on training data during 3D-CNN training phase.

brain MRI scan) the accuracy on the training data still remains 1.00, but the accuracy on the test data was reduced to 0.80. As can be seen in Fig. 8(b), during GABM generations the classification accuracy is remains high, even when the number of selected brain regions is decreasing. These results prove that some brain regions do not have any effect on the classifier accuracy using the training data. Therefore, these brain regions can be ignored.

In the second experiment using ADNI dataset, the GABM parameters were selected as $\alpha = 0.03$ and $\beta = 0.97$. Fig. 9(c) shows the results during 2000 generations of the GABM method. After optimization, the classification accuracy for the final brain mask was 0.96 on the training data and 0.85 on the test data. These results achieved using about 41 brain regions which means by removing 55 brain regions (more than half of all brain regions) the accuracy on the training data was reduced to 0.96, but the accuracy on the test data still remained the same (0.85). This experiment proved by increasing the value of β, the number of selected brain regions in the final mask will be decreased.

In the third experiment using ADNI dataset, the GABM parameters were selected as $\alpha = 0.025$ and $\beta = 0.975$. After 2000 generations, the classification accuracy using final best solution was about 0.92 for the training data and 0.83 for the test data. These results have been achieved using 31 brain regions (removing 65 brain regions that is about 2/3 of the whole brain regions). Using the mentioned 31 brain regions, the accuracy on both training and test data sets are acceptable.

In the last experiment using ADNI dataset, the GABM parameters were selected as $\alpha = 0.02$ and $\beta = 0.98$. The obtained results during the GABM optimization process are shown in Fig. 9(d). After finding an optimal brain mask, the accuracy on the training data was about 0.75 and the accuracy on test data was about 0.63. These results achieved using only 6 brain regions. These regions can be interpreted as the most important brain regions in the classification task under the study.

Fig. 10 shows the brain mask obtained in the last experiment on ADNI dataset, its related chromosome, and the name of identified knowledgeable brain regions. The presented mask in Fig. 10 has been extracted using a majority voting on the best individuals of the folds. This mask contains 4 brain regions in original size and 1 dilated brain region. As a result, the proposed GABM method enables us to convert a 3D-CNN (a black box learning algorithm) to a tool for finding knowledgeable brain regions related to a particular brain MRI dataset.

Similarly, in the first experiment using the ABIDE dataset, the GABM parameters were selected as $\alpha = 0.99$ and $\beta = 0.01$. Using the final brain mask, the accuracy on the training data was 0.94 and the accuracy on the test data was 0.70. These results have been achieved using 75 brain regions.

In the second experiment using the ABIDE dataset, the GABM parameters were selected as $\alpha = 0.03$ and $\beta = 0.97$. In this case, the classification accuracy on the training data was reduced to 0.92, but the accuracy on the test data was increased to 0.73 (shows about 0.03 improvement). These results were achieved using 62 brain regions.

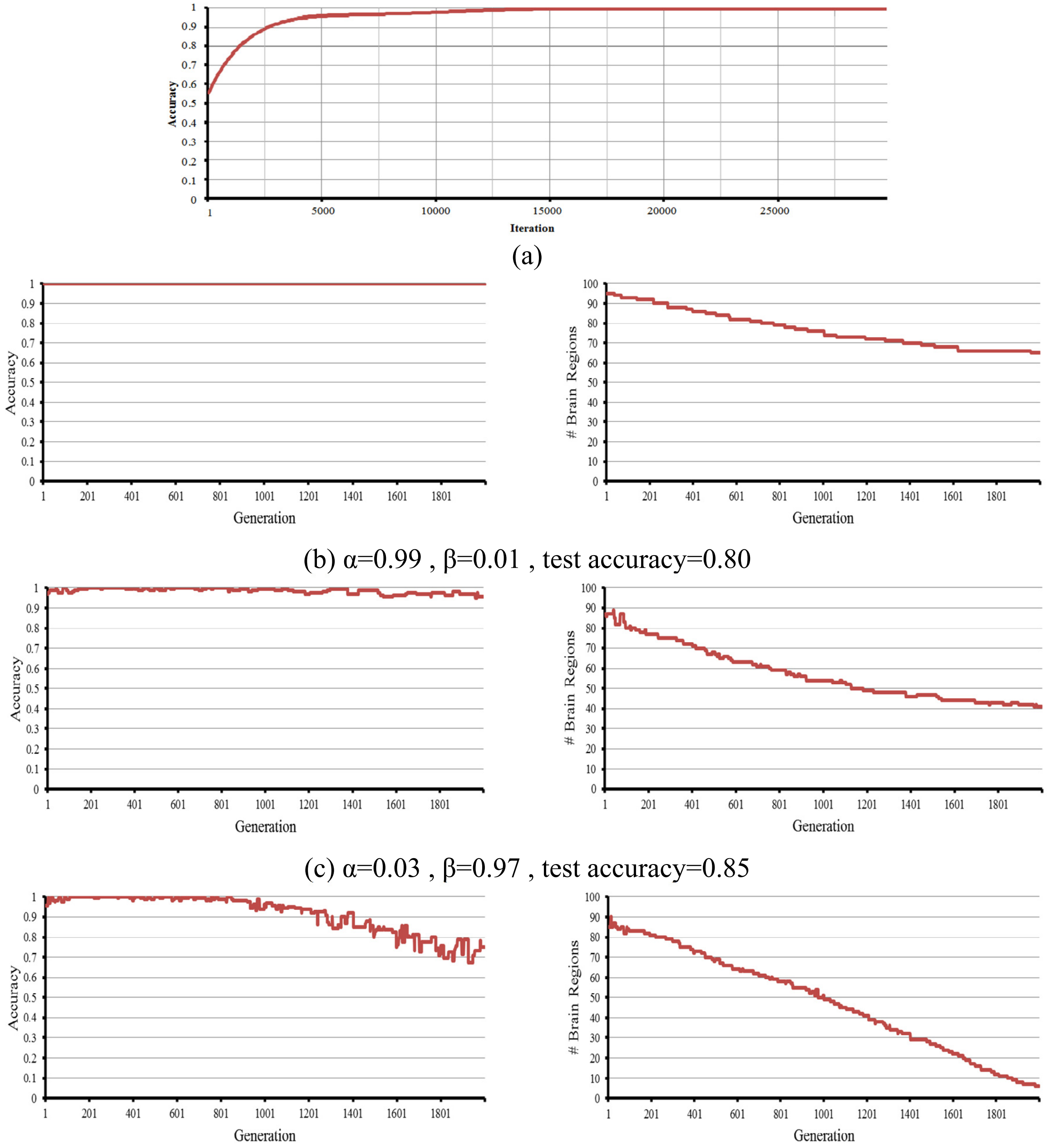
In the third experiment using the ABIDE dataset, the GABM parameters were selected as $\alpha = 0.025$ and $\beta = 0.975$. The classification accuracy using final brain masks was 0.89 for the training data and 0.67 for the test data. These results achieved using 53 brain regions.

In the last experiment using the ABIDE dataset, the GABM parameters were selected as $\alpha = 0.02$ and $\beta = 0.98$. Using final brain mask, the accuracy on the training data was about 0.76 and the accuracy on the test data was about 0.61. Fig. 11 shows obtained results in the last experiment on the ABIDE data. These results have been achieved using only 15 brain regions. These regions can be considered as the most important brain regions in ASD classification using a 3D-CNN model and the ABIDE dataset.

Fig. 12 shows final brain mask (obtained from all folds in the last experiment on the ABIDE dataset), its corresponding chromosome, and the regions' names. This mask contains 5 brain regions in original size, 3 dilated and 7 eroded regions. Here, the proposed GABM method used a 3D-CNN model to find knowledgeable brain regions for ASD classification using MRI data.

4.1. Discussion on classification accuracy

The results of all experiments are summarized in Table 3. When the 3D-CNN model without GABM was used for classification, all 96 predefined brain regions were involved for model



(a)

(b) α=0.99 , β=0.01 , test accuracy=0.80

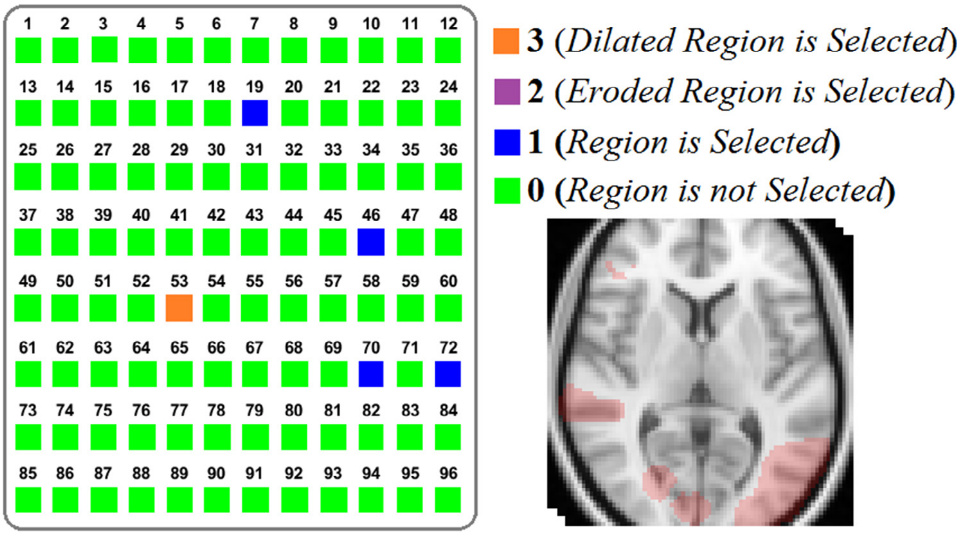
(c) α=0.03 , β=0.97 , test accuracy=0.85

(d) α=0.02 , β=0.98 , test accuracy=0.63

Fig. 9. Results on the ADNI dataset: (a) classification accuracy on training data during the 3D-CNN learning phase. In (b), (c), and (d) *left column* shows classification accuracy on the masked training data using best solution, and *right column* shows the number of involved brain regions during GABM optimization. (b) for $\alpha = 0.99$ and $\beta = 0.01$, (c) for $\alpha = 0.03$ and $\beta = 0.97$, and (d) for $\alpha = 0.02$ and $\beta = 0.98$.

training and testing. This method was evaluated using two brain MRI datasets. The obtained accuracy was acceptable (0.85 for ADNI and 0.70 for ABIDE) but we lost the ability of highlighting knowledgeable brain regions. Similarly, using the 3D-CNN + GABM method all brain regions were involved in model training, but only a number of them have been selected by the proposed method. The GABM method is similar to network pruning techniques. The idea is that among many parameters in the network, some are redundant and do not contribute considerably to the output. Pruning is a promising way for eliminating parameters based on a cost function. Generally, pruning is applied to reduce computational costs (Molchanov, Tyree, Karras, Aila, & Kautz, 2016), but they can also be used to deal with over-fitting (Barbu, She, Ding, & Gramajo, 2016; Bartoldson, Barbu, & Erlebacher, 2018).

In this paper, the proposed GABM method is applied to discover most important brain regions and discarding the redundant part of the brain MRI scans according to the disease under study. The test accuracy of 3D-CNN + GABM method on the ADNI dataset, was 0.85 when $\alpha = 0.03$ and $\beta = 0.97$. This accuracy was obtained using only 41 brain regions, which is equal to the obtained accuracy using all 96 brain regions. This experiment proved that, some brain regions may be redundant and the proposed GABM can find them properly. For the ABIDE dataset, the best test accuracy was 0.73 for $\alpha = 0.03$ and $\beta = 0.97$. This result shows that using the proposed GABM improves



19=>Left Lateral Occipital Cortex, inferior division

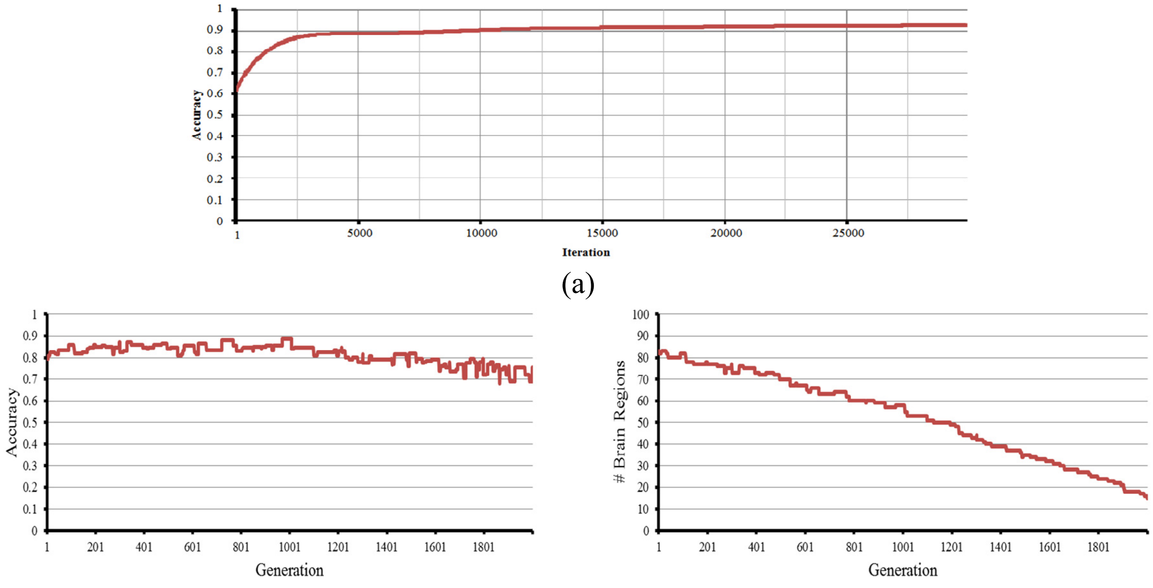
46=>Left Temporal Fusiform Cortex, posterior division

53=>Right Cuneal Cortex (↑)

70=>Right Middle Frontal Gyrus

72=>Right Middle Temporal Gyrus, posterior division

Fig. 10. The obtained mask, its corresponding chromosome, and the names of discovered knowledgeable brain regions from the ADNI dataset ("↑" means dilation).



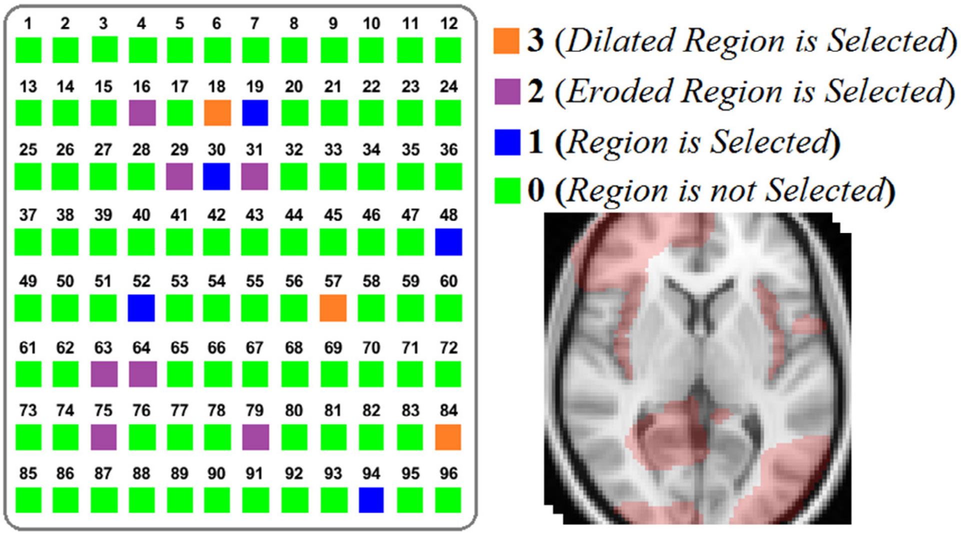
(a)

(b) α=0.02 , β=0.98 , test accuracy=0.61

Fig. 11. Results on the ABIDE dataset: (a) classification accuracy on training data during the 3D-CNN learning phase. In (b) *left column* shows classification accuracy on the masked training data using best solution, and *right column* shows the number of involved brain regions during GABM optimization, for $\alpha = 0.99$ and $\beta = 0.01$.

the test accuracy about 0.03 by selecting only 62 brain regions. As mentioned, the GABM method is similar to pruning methods. If we carefully select the pruned parameters according to a suitable measure, performance improvement may be occurred (He et al., 2019). This is the main reason for accuracy improvement by the proposed GABM method.

The comparison of the proposed method with a list of recent works on ADNI and ABIDE datasets is presented in Table 4. For



16=>Left Insular Cortex (↓)

18=>Left Juxtapositional Lobule Cortex (formerly Supplementary Motor Cortex) (↑)

19=>Left Lateral Occipital Cortex, inferior division

29=>Left Parahippocampal Gyrus, anterior division (↓)

30=>Left Parahippocampal Gyrus, posterior division

31=>Left Parietal Operculum Cortex (↓)

48=>Left Temporal Pole

52=>Right Cingulate Gyrus, posterior division

57=>Right Frontal Pole (↑)

63=>Right Inferior Temporal Gyrus, temporooccipital part (↓)

64=>Right Insular Cortex (↓)

75=>Right Occipital Pole (↓)

79=>Right Parietal Operculum Cortex (↓)

84=>Right Precuneous Cortex (↑)

94=>Right Temporal Fusiform Cortex, posterior division

Fig. 12. The obtained mask, its corresponding chromosome, and names of discovered knowledgeable brain regions from the ABIDE dataset ("↑" means dilation and "↓" means erosion).

Table 3
Summarized results of all experiments. Includes parameters values, number of selected regions and accuracy (using 5-fold cross validation) on train and test data.

Method	Parameters		ADNI			ABIDE		
	α	β	# regions	Train	Test	# regions	Train	Test
3D-CNN	––	––	96	1.00	0.85	96	0.93	0.70
3D-CNN + GABM	0.99	0.01	65	1.00	0.80	75	0.94	0.70
3D-CNN + GABM	**0.03**	**0.97**	**41**	**0.96**	**0.85**	**62**	**0.92**	**0.73**
3D-CNN + GABM	0.025	0.975	31	0.92	0.83	53	0.89	0.67
3D-CNN + GABM	0.02	0.98	6	0.75	0.63	15	0.76	0.61

ADNI dataset, three studies have been reported which applied a 3D-CNN model on full MRI scans. Besides, for ABIDE dataset we just reported a set of papers which worked on a large sample size. The classification accuracy for each dataset is reported in two cases, for the 3D-CNN model alone, and for the combination of 3D-CNN and GABM method (3D-CNN + GABM). According to this table, the proposed method outperforms other approaches in terms of accuracy. For ADNI dataset, the classification accuracy

Table 4
Comparison of the proposed method with a list of previous research with the data and the performances obtained. CC stands for Computational Complexity.

Reference	Method	Modal	CC	# Subject	Accuracy
Alzheimer Classification					
Korolev et al. (2017)	3D-CNN	MRI	$O(n^5)$	111	0.80
Rieke et al. (2018)	3D-CNN (5-fold)	MRI	$O(n^5)$	969	0.77
Yang et al. (2018)	3D-CNN (5-fold)	MRI	$O(n^5)$	103	0.79
Proposed Method	**3D-CNN (5-fold)**	**MRI**	$O(n^5)$	**140**	**0.85**
Proposed Method	**3D-CNN + GABM (5-fold)**	**MRI**	$O(n^6)$	**140**	**0.85**
Autism Classification					
Sabuncu et al. (2015)	MVPA (5-fold)	MRI	$O(n^4)$	935	0.60
Monté-Rubio et al. (2018)	SVM + Bayesian (5-fold)	MRI	$O(n^4)$	1102	0.62
Dvornek et al. (2017)	Deep Learning (10-fold)	R-fMRI	$O(n^6)$	1100	0.68
Dvornek et al. (2018)	Deep Learning (10-fold)	R-fMRI	$O(n^6)$	1100	0.70
Heinsfeld et al. (2018)	Deep Learning (10-fold)	R-fMRI	$O(n^6)$	964	0.70
Li et al. (2018)	Deep Learning (5-fold)	R-fMRI	$O(n^6)$	1100	0.70
Proposed Method	**3D-CNN (5-fold)**	**MRI**	$O(n^5)$	**1000**	**0.70**
Proposed Method	**3D-CNN + GABM (5-fold)**	**MRI**	$O(n^6)$	**1000**	**0.73**

using all brain regions is equal to the classification accuracy using only the knowledgeable brain regions. For ABIDE dataset the classification accuracy of the proposed 3D-CNN is 0.70 and the classification accuracy of 3D-CNN + GABM is 0.73, which shows about 0.03 improvement.

4.2. Discussion on computational complexity

For analyzing the computational complexity of neural networks, it will be useful to separate the training and inference phases, because we do not have back propagation in inference phase which has a very high computational complexity. By fixing an architecture of a neural network (underlying graph and activation functions), each network is parameterized by a weight vector $w \in R^d$. The computational complexity of a network is highly dependent on the number of its weights. In CNN models, the fully connected layers typically contain more than 90% of the CNN weights (Shen, Ferdman, & Milder, 2017). For a fully connected network, the computational complexity of forward pass is of order $O(n^4)$ and the computational complexity of back propagation is of order $O(n^5)$ (Zhang & Leatham, 2019). Table 4 shows the computational complexity of the proposed methods in comparison with other state of the art methods. According to this table, we can see that the computational complexity of the proposed GABM method is $O(n^6)$ which has been obtained according to the fact that the trained 3D-CNN should be employed (with the order $O(n^4)$) within the second internal loop of the fitness function calculation of the presented GA. Table 4 confirms that the proposed methods in this paper are not only more accurate and explainable (capable of discovering knowledgeable brain regions), but also computationally competitive compared to several state of the art methods.

4.3. Discussion on discovered knowledgeable brain regions

In order to validate the discovered knowledgeable brain regions, recent papers have been investigated in both Alzheimer and Autism research fields. In Alzheimer domain, 5 knowledgeable brain regions have been discovered by the proposed GABM method (see Fig. 10). Several papers also reported the high importance of these discovered brain regions in Alzheimer disease analysis (see Table 5). Wang, Wilson, and Hancock (2017) reported top 10 regions of interest (ROIs) with significant differences between AD patients and normal cases. These ROIs include the *Left Temporal Fusiform Cortex* and the *Middle Temporal Gyrus*, which have been also discovered by the proposed GABM method. These results are consistent with previous studies (Khazaee, Ebrahimzadeh, Babajani-Feremi, & Initiative, 2017; Rubinov & Sporns, 2010) which suggested that the *Middle Temporal Gyrus* is an important region in AD pathology. Dillen et al. (2016) reported prominent group differences localized in the *Left Lateral Occipital Cortex* for AD cases, healthy subjects and SCI patients. Similarly, *Right Cuneal Cortex* and *Right Middle Frontal Gyrus* have been reported as important brain regions for AD diagnosis in (Hafkemeijer et al., 2015) and (Guo et al., 2016; Jung et al., 2017), respectively.

In Autism domain, the proposed GABM method has discovered 15 knowledgeable brain regions. All of these regions were previously studied by other researchers to discover their involvement in ASD. Mengotti et al. (2011) reported that compared to normal children, individuals with ASD had significantly increased white matter volumes and decreased gray matter volumes in the *left Juxtapositional Lobule Cortex* (supplementary motor area). They also reported that children with autism compared to normally developing subjects had significantly increased in gray matter volumes in the *Right Inferior Temporal Gyrus*. Goddard, Swaab, Rombouts, and van Rijn (2016) reported significant gray matter volume differences between normal, ASD, and Klinefelter syndrome groups in the *Left Insular Cortex*. Furthermore, the effects of ASD on the *Right Inferior Temporal Gyrus, Left Juxtapositional Lobule Cortex, Left Parahippocampal Gyrus, Right Occipital Pole,* and *Right Temporal Fusiform Cortex* are studied by (Alvarez-Jimenez, Múnera-Garzón, Zuluaga, Velasco, & Romero, 2019), which have been also discovered by the proposed GABM method. Similarly, other discovered knowledgeable brain regions for ASD were studied in this research field. The *Left Lateral Occipital Cortex* has been studied by (Mueller et al., 2013), *Left Parietal Operculum Cortex* by (Rosenblau, Kliemann, Dziobek, & Heekeren, 2017), *Left Temporal Pole* by (Pua, Malpas, Bowden, & Seal, 2018), *Right Cingulate Gyrus* by (Green et al., 2013), *Right Frontal Pole* by (Pua et al., 2018), *Right Insular Cortex* by (Goddard et al., 2016), *Right Parietal Operculum Cortex* by (Nickl-Jockschat et al., 2012), and *Right Precuneus Cortex* by (Bonilha et al., 2008). The list of all knowledgeable brain regions discovered by the proposed GABM method, and the references which studied them are provided in Table 5.

To discover more efficient brain regions using GABM method, it is possible to use multiple brain atlases instead of a single one. Several researchers reported that, by employing more than one atlas for brain image analysis, an accuracy improvement can be achieved (Alvén, Norlén, Enqvist, & Kahl, 2016; Iglesias & Sabuncu, 2015; Sun, Shao, Wang, Zhang, & Liu, 2019). When multiple brain atlases are used in a given model, establishing a reliable fusion method is of great importance in giving accurate results (Yang, Jia, & Yang, 2019). Using multiple atlases inside GABM method would be an interesting idea. The majority voting

Table 5
List of related works, which have been previously studied, the discovered knowledgeable brain regions.

Region ID	Region name	Reference
Alzheimer Classification		
19	Left Lateral Occipital Cortex	McLachlan et al. (2018) Dillen et al. (2016) Hafkemeijer et al. (2015)
46	Left Temporal Fusiform Cortex	Wang et al. (2017) Irish et al. (2016)
53	Right Cuneal Cortex	Hafkemeijer et al. (2015)
70	Right Middle Frontal Gyrus	Ortiz et al. (2017) Guo et al. (2016)
72	Right Middle Temporal Gyrus	Liu et al. (2018) Wang et al. (2017)
Autism Classification		
16	Left Insular Cortex	Goddard et al. (2016)
18	Left Juxtapositional Lobule Cortex	Alvarez-Jimenez et al. (2019) Mengotti et al. (2011)
19	Left Lateral Occipital Cortex	Mueller et al. (2013)
29	Left Parahippocampal Gyrus, anterior division	Alvarez-Jimenez et al. (2019)
30	Left Parahippocampal Gyrus, posterior division	Alvarez-Jimenez et al. (2019)
31	Left Parietal Operculum Cortex	Rosenblau et al. (2017)
48	Left Temporal Pole	Boddaert et al. (2009)
52	Right Cingulate Gyrus	Green et al. (2013)
57	Right Frontal Pole	Pua et al. (2018)
63	Right Inferior Temporal Gyrus	Alvarez-Jimenez et al. (2019) Mengotti et al. (2011)
64	Right Insular Cortex	Goddard et al. (2016)
75	Right Occipital Pole	Alvarez-Jimenez et al. (2019)
79	Right Parietal Operculum Cortex	Nickl-Jockschat et al. (2012)
84	Right Precuneus Cortex	Bonilha et al. (2008)
94	Right Temporal Fusiform Cortex	Alvarez-Jimenez et al. (2019)

method, union or intersection of different brain masks obtained by different brain atlases, are the most straightforward fusion methods, which can be used in multi atlas version of GABM method.

5. Conclusion

This paper proposed a 3D-CNN model to classify brain MRI scans into predefined groups. Furthermore, a GABM method was proposed as a visualization technique which gives insights into the function of the classifier. First, a set of preprocessed MRI scans have been used to train the proposed 3D-CNN. Then, a GA based method was introduced to discover knowledgeable regions of the brain. The knowledgeable regions are those areas of the brain which are important for classification. The proposed method was evaluated using two brain MRI datasets. The obtained classification accuracy was 0.85 for the ADNI dataset and 0.70 for the ABIDE dataset. Finally, the proposed GABM method has found 6 to 65 brain regions in ADNI dataset and 15 to 75 brain regions in ABIDE dataset with respect to the model parameters. The results shown that besides the model interpretability, the proposed GABM method has increased the final performance of the classifier in some cases.

As future works, we can aim to use multiple brain atlases rather than a single one to identify the knowledgeable brain regions. Also, the brain masks obtained by different brain atlases can be combined to achieve better results. Other significant point of the proposed GABM method is that it can be applied on other neuroimaging data (e.g. f-MRI data or multimodal data).

Declaration of competing interest

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

References

Alvarez-Jimenez, C., Múnera-Garzón, N., Zuluaga, M. A., Velasco, N. F., & Romero, E. (2019). Autism spectrum disorder characterization in children by capturing local-regional brain changes in MRI. *Medical Physics*.

Alvén, J., Norlén, A., Enqvist, O., & Kahl, F. (2016). Überatlas: fast and robust registration for multi-atlas segmentation. *Pattern Recognition Letters*, *80*, 249–255.

Anavi, Y., Kogan, I., Gelbart, E., Geva, O., & Greenspan, H. (2016). Visualizing and enhancing a deep learning framework using patients age and gender for chest x-ray image retrieval. In *Medical imaging 2016: Computer-aided diagnosis* (p. 978510). International Society for Optics and Photonics.

Barbu, A., She, Y., Ding, L., & Gramajo, G. (2016). Feature selection with annealing for computer vision and big data learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*, 272–286.

Bartoldson, B., Barbu, A., & Erlebacher, G. (2018). Enhancing the regularization effect of weight pruning in artificial neural networks. arXiv Prepr. arXiv:1805.01930.

Beheshti, I., Demirel, H., & Initiative, A. D. N. (2016). Feature-ranking-based Alzheimer's disease classification from structural MRI. *Magnetic Resonance Imaging*, *34*, 252–263.

Bernas, A., Aldenkamp, A. P., & Zinger, S. (2018). Wavelet coherence-based classifier: a resting-state functional MRI study on neurodynamics in adolescents with high-functioning autism. *Computer Methods and Programs in Biomedicine*, *154*, 143–151.

Blumberg, Stephen J., Bramlett, Matthew D., Kogan, Michael D., Schieve, Laura A., Jones, Jessica R., & Lu, M. C. (2013). *Changes in prevalence of parent-reported autism spectrum disorder in school-aged U.S. children: 2007 to 2011-2012. national center for health statistics reports. number 65. natl. cent. heal. stat.*

Boddaert, N., Zilbovicius, M., Philipe, A., Robel, L., Bourgeois, M., Barthélemy, C., et al. (2009). MRI Findings in 77 children with non-syndromic autistic disorder. *PLoS One*, *4*(e4415).

Bonilha, L., Cendes, F., Rorden, C., Eckert, M., Dalgalarrondo, P., Li, L. M., et al. (2008). Gray and white matter imbalance–typical structural abnormality underlying classic autism?. *Brain and Development*, *30*, 396–401.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3D U-net: learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention* (pp. 424–432). Springer.

Cheng, X., Zhang, L., & Zheng, Y. (2016). Deep similarity learning for multimodal medical images. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 1–5.

de Vos, B. D., Wolterink, J. M., de Jong, P. A., Viergever, M. A., & Isgum, I. (2016). 2d image classification for 3D anatomy localization: employing deep convolutional neural networks. *Medical Imaging: Image Processing*, 97841Y.

Dillen, K. N. H., Jacobs, H. I. L., Kukolja, J., von Reutern, B., Richter, N., Onur, Ö. A., et al. (2016). Aberrant functional connectivity differentiates retrosplenial cortex from posterior cingulate cortex in prodromal Alzheimer's disease. *Neurobiology of Aging*, *44*, 114–126.

Dvornek, N. C., Ventola, P., & Duncan, J. S. (2018). Combining phenotypic and resting-state fmri data for autism classification with recurrent neural networks. In *Biomedical imaging (ISBI 2018), 2018 IEEE 15th international symposium on* (pp. 725–728). IEEE.

Dvornek, N. C., Ventola, P., Pelphrey, K. A., & Duncan, J. S. (2017). Identifying autism from resting-state fmri using long short-term memory networks. In *International workshop on machine learning in medical imaging* (pp. 362–370). Springer.

Ecker, C., Marquand, A., Mourão Miranda, J., Johnston, P., Daly, E. M., Brammer, M. J., et al. (2010). Describing the brain in autism in five dimensions—magnetic resonance imaging-assisted diagnosis of autism spectrum disorder using a multiparameter classification approach. *Journal of Neuroscience*, *30*, 10612–10623.

Fein, D., Barton, M., Eigsti, I.-M., Kelley, E., Naigles, L., Schultz, R. T., et al. (2013). Optimal outcome in individuals with a history of autism. *Journal of Child Psychology and Psychiatry*, *54*, 195–205. http://dx.doi.org/10.1111/jcpp.12037.

García, E., Diez, Y., Diaz, O., Lladó, X., Gubern-Mérida, A., Martí, R., et al. (2019). Breast MRI and X-ray mammography registration using gradient values. *Medical Image Analysis*.

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 580–587).

Goddard, M. N., Swaab, H., Rombouts, S. A. R. B., & van Rijn, S. (2016). Neural systems for social cognition: gray matter volume abnormalities in boys at high genetic risk of autism symptoms, and a comparison with idiopathic autism spectrum disorder. *European Archives of Psychiatry and Clinical Neuroscience*, *266*, 523–531.

Gonzalez, R. C., & Woods, R. E. (2002). *Digital image processing*. Prentice hall Englewood Cliffs.

Green, S. A., Rudie, J. D., Colich, N. L., Wood, J. J., Shirinyan, D., Hernandez, L., et al. (2013). Overreactive brain responses to sensory stimuli in youth with autism spectrum disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, *52*, 1158–1172.

Guo, Z., Liu, X., Hou, H., Wei, F., Liu, J., & Chen, X. (2016). Abnormal degree centrality in Alzheimer's disease patients with depression: A resting-state functional magnetic resonance imaging study. *Experimental Gerontology*, *79*, 61–66.

Hafkemeijer, A., Möller, C., Dopper, E. G. P., Jiskoot, L. C., Schouten, T. M., van Swieten, J. C., et al. (2015). Resting state functional connectivity differences between behavioral variant frontotemporal dementia and Alzheimer's disease. *Frontiers Human Neuroscience*, *9*(474).

He, Y., Dong, X., Kang, G., Fu, Y., Yan, C., & Yang, Y. (2019). Asymptotic soft filter pruning for deep convolutional neural networks. *IEEE Transactions on Cybernetics*.

Heinsfeld, A. S., Franco, A. R., Craddock, R. C., Buchweitz, A., & Meneguzzi, F. (2018). Identification of autism spectrum disorder using deep learning and the ABIDE dataset. *NeuroImage: Clinical*, *17*, 16–23. http://dx.doi.org/10.1016/J.NICL.2017.08.017.

Hosseini-Asl, E., Ghazal, M., Mahmoud, A., Aslantas, A., Shalaby, A. M., Casanova, M. F., et al. (2018). Alzheimer's disease diagnostics by a 3D deeply supervised adaptable convolutional network. *Frontiers in Bioscience (Landmark Ed*, *23*, 584–596.

Huang, Y., Cao, X., Wang, Q., Zhang, B., Zhen, X., & Li, X. (2018). Long-short-term features for dynamic scene classification. *IEEE Transactions on Circuits and Systems for Video Technology*, *29*, 1038–1047.

Iglesias, J. E., & Sabuncu, M. R. (2015). Multi-atlas segmentation of biomedical images: a survey. *Medical Image Analysis*, *24*, 205–219.

Irish, M., Bunk, S., Tu, S., Kamminga, J., Hodges, J. R., Hornberger, M., et al. (2016). Preservation of episodic memory in semantic dementia: the importance of regions beyond the medial temporal lobes. *Neuropsychologia*, *81*, 50–60.

Jack, C. R., Jr., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., et al. (2008). The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetics Resonance Imaging An Official Journal of International Society of Magnetic Resonance in Medicine*, *27*, 685–691.

Jung, M., Tu, Y., Lang, C. A., Ortiz, A., Park, J., Jorgenson, K., et al. (2017). Decreased structural connectivity and resting-state brain activity in the lateral occipital cortex is associated with social communication deficits in boys with autism spectrum disorder. *Neuroimage*.

Khazaee, A., Ebrahimzadeh, A., Babajani-Feremi, A., & Initiative, A. D. N. (2017). Classification of patients with MCI and AD from healthy controls using directed graph measures of resting-state fmri. *Behavioural Brain Research*, *322*, 339–350.

Khedher, L., Ramírez, J., Górriz, J. M., Brahim, A., Segovia, F., & Initiative, A. s. D. N. (2015). Early diagnosis of Alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images. *Neurocomputing*, *151*, 139–150.

Khosla, M., Jamison, K., Kuceyeski, A., & Sabuncu, M. (2018). 3D Convolutional neural networks for classification of functional connectomes.

Klöppel, S., Abdulkadir, A., Jack, C. R. B., Koutsouleris, N., Mourão Miranda, J., & Vemuri, P. (2012). Diagnostic neuroimaging across diseases. http://dx.doi.org/10.1016/j.neuroimage.2011.11.002.

Kong, Y., Gao, J., Xu, Y., Pan, Y., Wang, J., & Liu, J. (2018). Classification of autism spectrum disorder by combining brain connectivity and deep neural network classifier. *Neurocomputing*, http://dx.doi.org/10.1016/J.NEUCOM.2018.04.080.

Korolev, S., Safiullin, A., Belyaev, M., & Dodonova, Y. (2017). Residual and plain convolutional neural networks for 3D brain MRI classification. In *2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017)* (pp. 835–838). IEEE.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*, 436–444.

Li, D., Karnath, H. O., & Xu, X. (2017). Candidate biomarkers in children with autism spectrum disorder: A review of MRI studies. *Neuroscience Bulletin*, http://dx.doi.org/10.1007/s12264-017-0118-1.

Li, H., Parikh, N. A., & He, L. (2018). A novel transfer learning approach to enhance deep neural network classification of brain functional connectomes. *Frontiers in Neuroscience*, *12*(491).

Li, F., Tran, L., Thung, K.-H., Ji, S., Shen, D., & Li, J. (2015). A robust deep model for improved classification of AD/MCI patients. *IEEE Journal of Biomedical Healing Informatics*, *19*, 1610–1616.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., et al. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, *42*, 60–88. http://dx.doi.org/10.1016/J.MEDIA.2017.07.005.

Liu, X., Chen, W., Tu, Y., Hou, H., Huang, X., Chen, X., et al. (2018). The abnormal functional connectivity between the hypothalamus and the temporal gyrus underlying depression in Alzheimer's disease patients. *Frontiers Aging Neuroscience*, *10*(37).

Liu, J., Li, M., Lan, W., Wu, F.-X., Pan, Y., & Wang, J. (2018). Classification of Alzheimer's disease using whole brain hierarchical network. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *15*, 624–632. http://dx.doi.org/10.1109/TCBB.2016.2635144.

Liu, J., Li, M., Pan, Y., Wu, F.-X., Chen, X., & Wang, J. (2017). Classification of schizophrenia based on individual hierarchical brain networks constructed from structural MRI images. *IEEE Transactions of Nanobioscience*, *16*, 600–608. http://dx.doi.org/10.1109/TNB.2017.2751074.

Liu, J., Pan, Y., Li, M., Chen, Z., Tang, L., Lu, C., et al. (2018). Applications of deep learning to MRI images: A survey. *Big Data Mining and Analytics*, *1*, 1–18.

Liu, X., Tizhoosh, H. R., & Kofman, J. (2016). Generating binary tags for fast medical image retrieval based on convolutional nets and radon transform. In *Neural networks (IJCNN), 2016 international joint conference on* (pp. 2872–2878). IEEE.

Liu, J., Wang, J., Tang, Z., Hu, B., Wu, F., & Pan, Y. (2017). Improving alzheimeres disease classification by combining multiple measures. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 1–1. http://dx.doi.org/10.1109/TCBB.2017.2731849.

Liu, J., Wang, Xiang, Zhang, X., Pan, Y., Wang, Xiaosheng, & Wang, J. (2017). MMM: classification of schizophrenia using multi-modality multi-atlas feature representation and multi-kernel learning. *Multimedia Tools and Applications*, 1–17.

Long, X., Chen, L., Jiang, C., Zhang, L., & Initiative, A. D. N. (2017). Prediction and classification of alzheimer disease based on quantification of MRI deformation. *PLoS One*, *12*, e0173372.

Lu, D., Heisler, M., Lee, S., Ding, G. W., Navajas, E., Sarunic, M. V., et al. (2019). Deep-learning based multiclass retinal fluid segmentation and detection in optical coherence tomography images using a fully convolutional neural network. *Medical Image Analysis*.

McLachlan, E., Bousfield, J., Howard, R., & Reeves, S. (2018). Reduced parahippocampal volume and psychosis symptoms in Alzheimer's disease. *International Journal of Geriatric Psychiatry*, *33*, 389–395.

Mengotti, P., D'Agostini, S., Terlevic, R., De Colle, C., Biasizzo, E., Londero, D., et al. (2011). Altered white matter integrity and development in children with autism: a combined voxel-based morphometry and diffusion imaging study. *Brain Research Bulletin*, *84*, 189–195.

Molchanov, P., Tyree, S., Karras, T., Aila, T., & Kautz, J. (2016). Pruning convolutional neural networks for resource efficient inference. arXiv Prepr. arXiv:1611.06440.

Monté-Rubio, G. C., Falcón, C., Pomarol-Clotet, E., & Ashburner, J. (2018). A comparison of various MRI feature types for characterizing whole brain anatomical differences using linear pattern recognition methods. *Neuroimage*.

Mueller, S., Keeser, D., Samson, A. C., Kirsch, V., Blautzik, J., Grothe, M., et al. (2013). Convergent findings of altered functional and structural brain connectivity in individuals with high functioning autism: a multimodal MRI study.

Nickl-Jockschat, T., Habel, U., Maria Michel, T., Manning, J., Laird, A. R., Fox, P. T., et al. (2012). Brain structure anomalies in autism spectrum disorder—a meta-analysis of VBM studies using anatomic likelihood estimation. *Human Brain Mapping*, *33*, 1470–1489.

Ortiz, A., Munilla, J., Martínez-Murcia, F. J., Górriz, J. M., Ramírez, J., & Initiative, A. D. N. (2017). Learning longitudinal MRI patterns by SICE and deep learning: assessing the Alzheimer's disease progression. In *Annual conference on medical image understanding and analysis* (pp. 413–424). Springer.

Payan, A., & Montana, G. (2015). Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks. arXiv Prepr. arXiv:1502.02506.

Plitt, M., Barnes, K. A., & Martin, A. (2015). Functional connectivity classification of autism identifies highly predictive brain features but falls short of biomarker standards. *NeuroImage Clin*, *7*, 359–366.

Prasoon, A., Petersen, K., Igel, C., Lauze, F., Dam, E., & Nielsen, M. (2013). Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. In *International conference on medical image computing and computer-assisted intervention* (pp. 246–253). Springer.

Pua, E. P. K., Malpas, C. B., Bowden, S. C., & Seal, M. L. (2018). Different brain networks underlying intelligence in autism spectrum disorders. *Human Brain Mapping*, *39*, 3253–3262.

Rieke, J., Eitel, F., Weygandt, M., Haynes, J.-D., & Ritter, K. (2018). Visualizing convolutional networks for MRI-based diagnosis of Alzheimer's disease. In *Understanding and interpreting machine learning in medical image computing applications* (pp. 24–31). Springer.

Rosenblau, G., Kliemann, D., Dziobek, I., & Heekeren, H. R. (2017). Emotional prosody processing in autism spectrum disorder. *Social Cognitive and Affective Neuroscience*, *12*, 224–239.

Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, *52*, 1059–1069.

Sabuncu, M. R., Konukoglu, E., & Initiative, A. D. N. (2015). Clinical prediction from structural brain MRI scans: a large-scale empirical study. *Neuroinformatics*, *13*, 31–46.

Sarraf, S., DeSouza, D. D., Anderson, J., & Tofighi, G. (2017). Deepad: Alzheimer′ s disease classification via deep convolutional neural networks using MRI and fMRI. BioRxiv 70441.

Shen, Y., Ferdman, M., & Milder, P. (2017). Escher: A CNN accelerator with flexible buffering to minimize off-chip transfer. In *2017 IEEE 25th annual international symposium on field-programmable custom computing machines (FCCM)* (pp. 93–100). IEEE.

Sun, L., Shao, W., Wang, M., Zhang, D., & Liu, M. (2019). High-order feature learning for multi-atlas based label fusion: Application to brain segmentation with MRI. *IEEE Transactions on Image Processing*.

Tan, F., Fu, X., Zhang, Y., & Bourgeois, A. G. (2008). A genetic algorithm-based method for feature subset selection. *Soft Computing*, *12*, 111–120.

Tejwani, R., Liska, A., You, H., Reinen, J., & Das, P. (2017). Autism classification using brain functional connectivity dynamics and machine learning. arXiv Prepr. arXiv:1712.08041.

Ventura, C., Masip, D., & Lapedriza, A. (2017). Interpreting CNN models for apparent personality trait regression. In *2017 IEEE conference on computer vision and pattern recognition workshops (CVPRW)* (pp. 1705–1713). IEEE, http://dx.doi.org/10.1109/CVPRW.2017.217.

Vieira, S., Pinaya, W. H. L., & Mechelli, A. (2017). Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications. *Neuroscience and Biobehavioral Reviews*, *74*, 58–75. http://dx.doi.org/10.1016/J.NEUBIOREV.2017.01.002.

Wang, Q., He, X., & Li, X. (2018). Locality and structure regularized low rank representation for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, *57*, 911–923.

Wang, Q., Liu, S., Chanussot, J., & Li, X. (2018). Scene classification with recurrent attention of VHR remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, *57*, 1155–1167.

Wang, J., Wilson, R. C., & Hancock, E. R. (2017). Detecting Alzheimer's disease using directed graphs. In *International workshop on graph-based representations in pattern recognition* (pp. 94–104). Springer.

Yang, Yunyun, Jia, W., & Yang, Yunna (2019). Multi-atlas segmentation and correction model with level set formulation for 3D brain MR images. *Pattern Recognition*, *90*, 450–463.

Yang, C., Rangarajan, A., & Ranka, S. (2018). Visual explanations from deep 3D convolutional neural networks for Alzheimer's disease classification. arXiv Prepr. arXiv:1803.02544.

Yang, D., Zhang, S., Yan, Z., Tan, C., Li, K., & Metaxas, D. (2015). Automated anatomical landmark detection ondistal femur surface using convolutional neural network. In *Biomedical imaging (ISBI), 2015 IEEE 12th international symposium on, vol. 1* (pp. 7–21). IEEE.

Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. In *European conference on computer vision* (pp. 818–833). Springer.

Zhang, Q., Cao, R., Shi, F., Wu, Y. N., & Zhu, S.-C. (2017). *Interpreting CNN Knowledge Via an Explanatory Graph*.

Zhang, N., & Leatham, K. (2019). A neurodynamics-based nonnegative matrix factorization approach based on discrete-time projection neural network. *Journal of Ambient Intelligence and Humanized Computing*, 1–9.

Zhang, J., Xie, Y., Wu, Q., & Xia, Y. (2019). Medical image classification using synergic deep learning. *Medical Image Analysis*, *54*, 10–19.

Zhang, Q., & Zhu, S. (2018). Visual interpretability for deep learning: a survey. *Frontiers of Information Technology & Electronic Engineering*, *19*, 27–39. http://dx.doi.org/10.1631/FITEE.1700808.