Brain-on-Cloud for automatic diagnosis of Alzheimer's disease from 3D structural magnetic resonance whole-brain scans*

Selene Tomassini[a], Agnese Sbrollini[a], Giacomo Covella[a], Paolo Sernani[a], Nicola Falcionelli[a], Henning Müller[b,1], Micaela Morettini[a,1], Laura Burattini[a,*,1] and Aldo Franco Dragoni[a,1]

[a]Department of Information Engineering, Engineering Faculty, Università Politecnica delle Marche (UnivPM), Ancona, Italy
[b]Information Systems Institute, University of Applied Sciences Western Switzerland (HES-SO), Sierre, Switzerland

ARTICLE INFO

Keywords:
Alzheimer's disease
Cloud computing
Computer-aided diagnosis
Convolutional LSTM
Diagnostic performance
3D structural magnetic resonance

Abstract

Background and objective: Alzheimer's disease accounts for approximately 70% of dementia cases. Cortical and hippocampal atrophy caused by Alzheimer's disease can be appreciated easily from a T1-weighted structural magnetic resonance scan. Since a timely therapeutic intervention during the initial stages of the syndrome has a positive impact on both disease progression and quality of life of affected subjects, Alzheimer's disease diagnosis is crucial. Thus, this study relies on the development of a robust yet lightweight 3D framework, Brain-on-Cloud, dedicated to efficient learning of Alzheimer's disease-related features from 3D structural magnetic resonance whole-brain scans by improving our recent convolutional long short-term memory-based framework with the integration of a set of data handling techniques in addition to the tuning of the model hyper-parameters and the evaluation of its diagnostic performance on independent test data.
Methods: For this objective, four serial experiments were conducted on a scalable GPU cloud service. They were compared and the hyper-parameters of the best experiment were tuned until reaching the best-performing configuration. In parallel, two branches were designed. In the first branch of Brain-on-Cloud, train, validation and test were performed on OASIS-3. In the second branch, unenhanced data from ADNI-2 were employed as independent test set and the diagnostic performance of Brain-on-Cloud was evaluated to prove its robustness and generalization capability. The prediction scores were computed for each subject and stratified according to age, sex and mini mental state examination.
Results: In its best guise, Brain-on-Cloud is able to discriminate Alzheimer's disease with an accuracy of 92% and 76%, sensitivity of 94% and 82%, and area under the curve of 96% and 92% on OASIS-3 and independent ADNI-2 test data, respectively.
Conclusions: Brain-on-Cloud results to be a reliable, lightweight and easily-reproducible framework for automatic diagnosis of Alzheimer's disease from 3D structural magnetic resonance whole-brain scans, performing well without segmenting the brain into its portions. Preserving the brain anatomy, its application and diagnostic ability can be extended to other cognitive disorders. Due to its cloud nature, computational lightness and fast execution, it can also be applied in real-time diagnostic scenarios providing prompt clinical decision support.

1. Introduction

Dementia is affecting around fifty million people worldwide and Alzheimer's Disease (AD) is the most predominant form [1], contributing up to 70% of all dementia cases as reported by the World Health Organization (WHO)[2]. AD is a cognitive disorder that begins with mild memory losses and worsens progressively until death, as it damages brain cells irreversibly, ending up with the destruction of the brain area that controls cardiac and respiratory functions [1, 2]. Age may significantly affect the evolution of the syndrome: the prevalence of AD after 85 years of age (up to 35%) is estimated to be higher than in other age groups [3, 4]. However, AD is not an exclusive consequence of biological ageing. According to the WHO, the onset of symptoms before 65 years of age accounts for up to 9% of all AD cases. Anatomically, as neurons are injured, connections between neurons may break down and many brain regions begin to shrink dramatically [5, 6]. Even relatively early in its clinical expression, brain atrophy caused by AD targets mainly the cerebral cortex and the anterior hippocampal regions that are involved in thinking, reasoning and keeping new memories [4, 7, 8].

The diagnosis of AD requires careful medical evaluations, including anamnesis, neuropsychological tests, such as Mini Mental State Examination (MMSE), and other neurobiological exams [5]. In addition, neuroimaging data are extensively used as diagnostic support [6]. Structural Magnetic Resonance Imaging (sMRI) is widely exploited

*The research presented here was supported by the projects "Using 3D convolutional neural networks for the identification of lung cancer histological types directly from computed tomography scans" funded by the Cariverona Foundation, Italy, and "Deep learning for early medical diagnosis: A novel methodology for different clinical scenarios" funded by the Department of Information Engineering, Engineering Faculty, Università Politecnica delle Marche, Ancona, Italy

*Corresponding author
s.tomassini@pm.univpm.it (S. Tomassini); a.sbrollini@staff.univpm.it (A. Sbrollini); g.covella@studenti.univpm.it (G. Covella); p.sernani@staff.univpm.it (P. Sernani); n.falcionelli@staff.univpm.it (N. Falcionelli); henning.mueller@hevs.ch (H. Müller); m.morettini@univpm.it (M. Morettini); l.burattini@univpm.it (L. Burattini); a.f.dragoni@.univpm.it (A.F. Dragoni)
ORCID(s): 0000-0002-1087-7004 (S. Tomassini); 0000-0002-9152-7216 (A. Sbrollini); 0000-0001-7614-7154 (P. Sernani); 0000-0002-1312-6310 (N. Falcionelli); 0000-0001-6800-9878 (H. Müller); 0000-0002-8327-8379 (M. Morettini); 0000-0002-9474-7046 (L. Burattini); 0000-0002-3013-3424 (A.F. Dragoni)

[1]Equal contribution

[2]https://www.who.int/en/news-room/fact-sheets/detail/dementia

for the investigation of progressive neurological impairment [9, 10]. It offers a painless and non-invasive method of analyzing the anatomical changes of the brain, combining radio waves and strong magnetic fields, in order to guarantee high level of spatial resolution [11, 12]. The clinical utility of sMRI in distinguishing AD from other cognitive disorders is well established [13], especially when using T1-weighted (T1w) sMRI scans [14, 15]. T1w images are useful to analyze the brain structure from an anatomical point of view, reliably differentiating between the gray and white matters [16]. Since strong T1 contrast is present between fluid and more solid anatomical structures, T1w images are more suitable for the morphological assessment of the brain anatomy [16]. Thus, the presence of cortical and hippocampal atrophy caused by AD can be easily appreciated from a T1w sMRI scan [14].

Despite the available diagnostic tools, the diagnosis of AD is still very difficult due to similar symptoms with other cognitive disorders. In clinical practice, AD diagnosis can be confirmed only after the patient's death by means of a post-mortem examination of the brain tissue [17]. Furthermore, an effective cure to reverse or arrest AD progression has not been identified yet [17]. Nevertheless, given the disproportionate aging of the population, AD-related socio-economic impact is continuing to rise [18]. Hence, AD diagnosis is crucial, as a timely therapeutic intervention especially during the initial stages of the syndrome appears to have a positive impact on both the progression of symptoms and the quality of life of diseased subjects [1, 2, 19]. In this regard, Artificial Intelligence (AI)-guided computer-aided systems have been widely parsed to automatically diagnose AD, especially using Deep Learning (DL) algorithms [20].

DL has been one of the crucial factors for the success of AI in the medical field [21, 22]. Compared with conventional Machine Learning (ML) algorithms, DL has multiple advantages in analyzing medical images, presenting high power in identifying complex structures and in automatizing feature extraction [7]. Indeed, DL is able to adaptively learn from the data (*i.e.,* fully data-driven process), obtaining the optimal representation of the problem, without relying on handcrafted features [1, 21]. Handcrafted feature extraction is difficult and time-consuming, especially due to the complexity of the diagnostic problem and the difficulty to model prior knowledge completely [17, 23]. Moreover, handcrafted features potentially lead to non-optimal diagnostic results, as they may not be well coordinated with succeeding classifiers [10]. Therefore, DL algorithms are better suitable than ML ones for generalizing even under slight anatomical changes, like the ones caused by AD [13, 24].

To the best of our knowledge, our recent manuscript was the first and only to propose an end-to-end framework, named ConvLSTM4AD[3], leveraging exclusively on a Convolutional Long Short-Term Memory (ConvLSTM)-based neural network to investigate the presence of AD [1]. Despite the promising results, ConvLSTM4AD was designed to work with only 5 slices per scan. Furthermore, neither the impact of data handling techniques nor the impact of model parameter optimization were analyzed there. Thus, the motivation behind the hereby presented study is to find a more robust yet lightweight 3D framework, named Brain-on-Cloud, for automatically detecting AD from 3D sMRI whole-brain scans on cloud. In this regard, the focus is on improving the end-to-end ConvLSTM-based model dedicated to efficient learning of AD features while overcoming the main limitations raised in [1] by: increasing the cardinality of 3D sMRI scans used to feed the neural network; increasing the number of analyzed slices per scan; automatizing the entire workflow; and conducting in-depth studies, including the impact of different data handling techniques, the impact of hyper-parameter selection on the performance of the model, and the investigation of the diagnostic performance of Brain-on-Cloud in relation to age, sex and MMSE. Additionally, the entire source code of Brain-on-Cloud will be available on GitHub under copyright, to ensure its full reproducibility[4].

2. Literature review

Discovering an algorithm able to automatically classify the anatomical brain changes caused by AD is an interesting research topic for the scientific community.

DL algorithms for automatic AD diagnosis mainly focused on semisupervised learning algorithms to make full use of both labelled and unlabelled sMRI data, and supervised learning algorithms to make use of labelled sMRI data only. As for semisupervised learning algorithms, Yu et al. [15] were the first to propose a tensor-train, high-order pooling-based Generative Adversarial Network (GAN) to assess Mild Cognitive Impairment (MCI) and AD from T1w sMRI scans taken from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. Yu et al. [25] also proposed a multi-directional perception GAN to visualize the morphological features indicating the severity of AD for subjects of different stages, conducting extensive experiments on the same ADNI database. As for supervised learning algorithms, Convolutional Neural Network (CNN) has become the standard DL model for automatic diagnosis of AD from sMRI data, mainly thanks to the ability to automatically extract the most relevant features [7, 26, 27]. The majority of CNNs takes as input 2D data. Since the sMRI scan is inherently a volume, such architectures split the 3D information into 2D multi-channel data (*i.e.,* each slice is analyzed independently). However, analyzing the 3D sMRI scan slice by slice may lead to loss of valuable information, as the spatial correlation between adjacent slices is neglected, which can thus not be taken into account [1, 17, 24]. Integrating the volumetric information into the learning can help the process. CNN-based methods that take as input sMRI volumes, referred to as scan-based approaches, can possibly improve the model performance as they capture richer spatial information (*i.e.,* the entire brain or the most significant brain portions can be encapsulated) and extract more discriminative features than slice-based approaches [1, 28]. Thus, scan-based approaches have the potential to investigate the overall information in a more

[3] https://github.com/airtlab/ConvLSTM4AD

[4] https://github.com/airtlab/Brain-on-Cloud

detailed way, resulting in clinically more reliable judgments [29].

Focusing exclusively on supervised learning, scan-based approaches starting from 2017, Luo et al. [20] presented an automatic AD recognition system based on a 3D CNN fed with sMRI whole-brain scans, taken from a subset of the ADNI database. Specifically, 47 AD and 34 Cognitively Normal (CN) scans were taken into account. The input consisted of 5 neighbouring slices per scan, each having a post-processed resolution of 54 pixels × 54 pixels. Brain registration was accomplished through a series of pre-processing steps. Random zooming and in-and-out cropping were used to augment and balance the dataset. Once pre-processed, the dataset was split into 66% data for training and 33% for testing, without reserving a percentage for validation. Their framework gained a good AD detection, reaching a SPecificity (SP) of 93% and a SEnsitivity (SE) of 100% on test data. In the same year, Korolev et al. [30] compared a 3D CNN and a residual neural network, emphasizing to achieve high performance without incorporating any handcrafted feature extraction step. The input consisted of 110 slices per scan, each having a resolution of 110 pixels × 110 pixels. The first model, trained, validated and tested on a subset of sMRI whole-brain data (50 AD and 61 CN) of the ADNI database, that has been pre-processed with alignment and skull-stripping, gained a test ACCuracy (ACC) of 79% and an Area Under the Curve (AUC) of 88% in discriminating AD from CN subjects. In 2018, Bäckström et al. [31] proposed a 3D CNN for automatic feature learning and AD detection on a pre-processed sMRI whole-brain subset of the ADNI database, considering multiple scans per subject. Image pre-processing consisted of cortical reconstruction (performed by the dataset provider), trimming, resizing and normalization. The 3D CNN, made up of five convolutional layers for feature extraction and three fully-connected layers for AD/CN classification, was fed with 110 slices per scan, each with a resolution of 110 pixels × 110 pixels, and reached a test ACC of 98.73%. In 2019, Basaia et al. [13] developed a 3D CNN leveraged on a single axial sMRI whole-brain scan to discriminate AD from CN subjects. Data of 294 AD and 352 CN subjects were selected from three subsets of the ADNI database, and split into train and test sets according to a ratio of 9:1. They also reserved 10% data from the train set to perform validation. In AD versus CN classification, their transfer learning-based method reached an ACC of 99.20%, SP of 99.50% and SE of 98.90% on test data. In the same year, Jabason et al. [17] built a framework based on an ensemble of CNNs fed with the 100 most informative slices of each sMRI scan taken from the OASIS-3 subset of the Open Access Series of Imaging Studies (OASIS) database. By applying a combination of voting algorithms of the three pre-trained CNN pipelines, the best-performing experimental setup reached a test ACC of 95.23% without specifying how the data were pre-processed. In 2020, Xia et al. [24] designed a hybrid framework to separate AD from CN, consisting of a 3D CNN to learn low-level features, coupled with a ConvLSTM module to extract high-level spatial information. Experiments were performed on T1w sMRI whole-brain data (198 AD and 229 CN) taken from a subset of the ADNI database. All scans were pre-processed and reduced to 143 slices per scan, each with a resolution of 119 pixels × 119 pixels. The best experiment gained a test ACC of 94.19%, SP of 94.57%, SE of 93.75% and AUC of 96% for classifying AD. In 2021, Saratxaga et al. [32] implemented a DL-based method for AD automatic diagnosis from sMRI whole-brain scans, considering both state-of-the-art and customized architectures. Data were selected from two subsets (OASIS-1 and OASIS-2, separately used to train, validate and test the models) of the OASIS database. According to the input size requirements of each implemented model, two slice sizes were considered: 176 pixels × 176 pixels and 224 pixels × 224 pixels. Moreover, a data augmentation strategy (rotation, vertical flip and horizontal flip) was performed to avoid overfitting. Over the OASIS-1 subset, the best 3D model performing a subject-level classification resulted to be a customized architecture with five convolutional blocks, named BrainNet3D, as it achieved a test ACC of 80%. Over the OASIS-2 subset, the same model achieved a test ACC of 84% in discriminating AD from CN subjects.

3. Brain-on-Cloud

With the aim to improve our end-to-end ConvLSTM-based neural network, four serial experiments were conducted, adding in each experiment new improvements in relation to the previous one. In the first experiment, only intensity normalization and automated cropping were performed. In the second and third experiments, scan registration and brain extraction were respectively added, as they are the distinctive steps in the analysis of sMRI brain scans [26]. In the fourth experiment, data augmentation was performed on the train set. Then, the output scans of each experiment were used to feed the ConvLSTM-based neural network for feature extraction and classification. All experiments were compared in terms of classification metrics and the hyper-parameters of the best experiment were tuned until reaching the best-performing configuration. In parallel, two branches were designed. In the first branch of Brain-on-Cloud, train, validation and test were performed on the OASIS-3 dataset. In the second branch, unenhanced data from the ADNI-2 dataset were employed as independent test set and the diagnostic performance of Brain-on-Cloud was evaluated to prove its robustness and generalization capability. Also, prediction scores were computed for each subject and stratified according to age, sex and MMSE.

The described workflow is shown in Figure 1 and details are reported in the following subsections.

3.1. Data selection

For ensuring a less database-specific, thus more generalizable approach, two datasets (OASIS-3 and ADNI-2) of two different openly-available databases (OASIS[5] and

[5] https://www.oasis-brains.org/

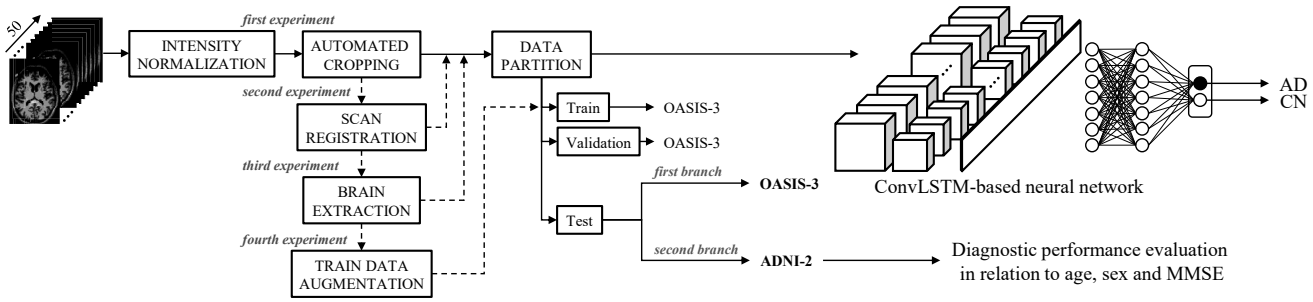


Figure 1: The workflow of Brain-on-Cloud. Four serial experiments were conducted. In the first experiment, intensity normalization and automated cropping were performed. In the second and third experiments, scan registration and brain extraction were respectively added. In the fourth experiment, train data augmentation was performed. Then, the output scans of each experiment were used to feed the ConvLSTM-based neural network for feature extraction and classification. In parallel, two branches were designed. In the first branch of Brain-on-Cloud, train, validation and test were performed on OASIS-3. In the second branch, unenhanced data from ADNI-2 were employed as independent test set and prediction scores were stratified according to age, sex and MMSE.

ADNI[6]) were chosen, as better detailed in subsection 3.1.1 and subsection 3.1.2.

For more reliable outcomes and also to lighten the overall computational effort, it is important to choose the most informative slices from the available data [17]. Thus, the slices corresponding to the anterior hippocampal regions were carefully selected, as affected early in AD progression [7]. To do so, the recommendations for anterior hippocampal area contouring from axial T1w sMRI proposed by Chera et al. [33] were followed. As the hippocampus is the gray matter located inside the temporal horn curve [34], this latter structure was identified first. After contouring the hippocampus inside the temporal horn curve, it was identified inferiorly from the level of the temporal horn, as its caudal portion is located at the level of the pituitary gland and pons [33]. Then, the hippocampus was identified superiorly from the level of the temporal horn, as its cranial portion is located at the level of the splenium of the corpus callosum [33]. Hence, of the totality of slices per scan, only slices comprised in the interval ranging from 85 to 135 were automatically selected for both OASIS-3 and ADNI-2, as that range perfectly encapsulated both anterior hippocampi. The slice selection procedure (Figure 2) led to a total of 50 whole-brain slices for each 3D sMRI scan.

3.1.1. OASIS-3

Data for the first branch (Figure 1) were selected from the OASIS-3 dataset, as it is the most recent subset of the OASIS database. Only one raw T1w 3D sMRI scan was considered for each subject, resulting in 275 scans, 145 AD and 130 CN. In case of multiple scans per subject, the first in chronological order was chosen to avoid an intra-subject bias. Each scan contains 256 stacked slices, having an original resolution of 176 pixels × 256 pixels, a thickness of 1 mm and a pixel size of 1 mm [35]. AD scans belong to 74 anonymized women and 71 anonymized men ranging from 52 to 95 years in age, whereas CN scans belong to 81 anonymized women and 49 anonymized men ranging from 45 to 86 years in age. All selected scans were acquired with 1.5 T and 3.0 T Siemens scanners and stored as Digital Imaging and Communications in Medicine (DICOM) files, then converted to compressed Neuroimaging Informatics Technology Initiative (NIFTI) files.

3.1.2. ADNI-2

Data for the second branch (Figure 1) were selected from the ADNI-2 dataset, which is a subset of the ADNI database[7]. Only one unenhanced T1w 3D sMRI scan was considered for each subject, resulting in 66 scans, 33 AD and 33 CN. In case of multiple scans per subject, the first in chronological order with at least one clinical assessment (*e.g.,* MMSE) was chosen to avoid an intra-subject bias. Each scan contains 256 stacked slices, having an original resolution of 176 pixels × 240 pixels, a thickness of 1.2 mm and a pixel size of 1 mm. AD scans belong to 8 anonymized women and 25 anonymized men ranging from 57 to 88 years in age, whereas CN scans belong to 23 anonymized women and 10 anonymized men ranging from 57 to 82 years in age. All selected scans were acquired with 1.5 T and 3.0 T Siemens scanners and stored as DICOM files, then converted to compressed NIFTI files.

3.2. Neural network architecture and setup

Seeking the best trade-off between the model computational cost, speed and performance, a 6-layered sequential neural network with 56,425,794 total parameters was designed. The first layer, a ConvLSTM layer, is the backbone of the model. It has 8 convolution filters with a kernel of 3 × 3. ConvLSTM is a variant of a recurrent neural network that exploits convolution filters and is able to model the long-term interactions while exploring the spatial information [1, 24].

[6] https://http://adni.loni.usc.edu/

[7] The investigators within ADNI, listed at http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf, contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this study

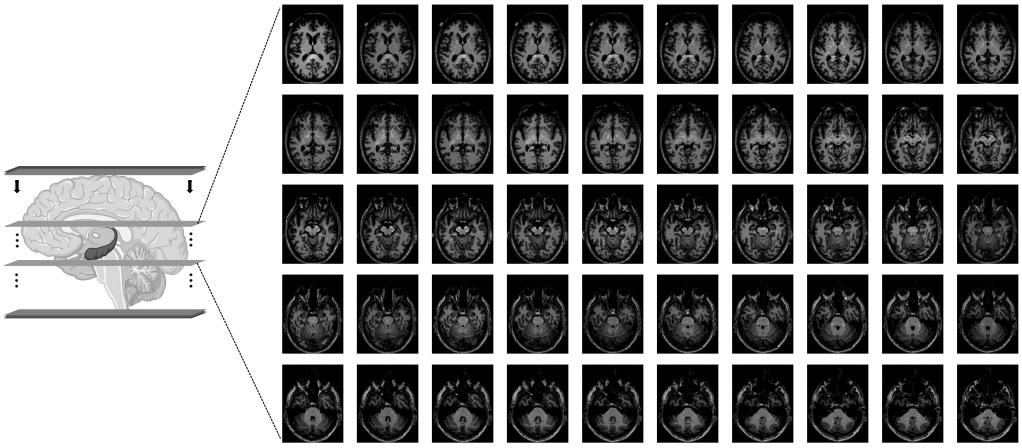


Figure 2: The slice selection procedure. Of the 256 slices per scan, 50 slices were automatically selected for both OASIS-3 and ADNI-2, perfectly encapsulating both anterior hippocampal regions.

Discriminative features are automatically extracted through convolution mechanisms different from the ones exploited by a CNN [24]. During convolution operations, the model learns which filters have to be activated when seeing a feature at a specific spatial position in the input. The main function of LSTM, instead, is to help the preservation of the error in order to be back-propagated [21, 26]. In a LSTM hidden unit (Figure 3), each sequence is treated entirely and the information is stored in a gated memory cell. The memory cell itself decides about what to store and when to allow the reading and updating of the information via three gates, which open and close whenever required [26]: the input gate transfers new input information to the memory cell, the forget gate selectively forgets non-valuable information, whereas the output gate allows the storage of relevant information [12, 24]. As pointed out in [36], the output h_t at time point t

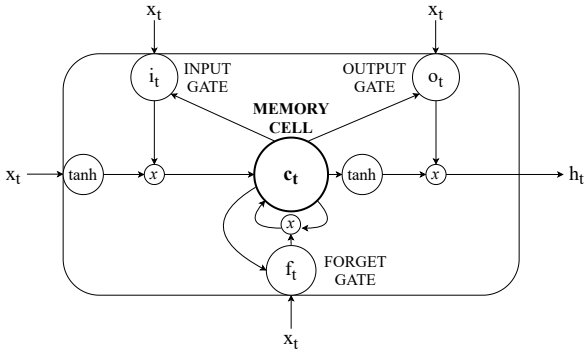


Figure 3: The LSTM hidden unit. The memory cell decides about what to store and when to allow the reading and updating of the information via three (input, forget and output) gates.

is regulated by (1),

$$\begin{aligned} i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \\ f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \\ c_t &= f_t c_{t-1} + i_t tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \\ o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \\ h_t &= o_t tanh(c_t) \end{aligned} \quad (1)$$

where i_t, f_t, o_t and c_t are respectively the activation vectors of the gates and of the memory cell at time point t, σ is the sigmoid activation function, $tanh$ is the hyperbolic tangent activation function, x_t denotes the current input, b denotes the bias of the memory cell and of each gate, and W are the weight matrices. ConvLSTM was selected because both spatial and temporal AD features take part in the classification [12]. The second layer is a Dropout layer. Thanks to dropout, which is a regularization technique, a few units are randomly removed from the model during the training phase, reducing the overall complexity of the neural network [1]. The dropout rate was initially set to 0.5 to smooth out overfitting. The third layer is a Flatten layer, which flattens all the extracted features into a big mono-dimensional tensor. The fourth layer is a Dense layer. It has 256 neurons and Rectified Linear Unit (ReLU) as activation function. ReLU helps the model consider non-linear effects and interactions, demonstrating faster training and better results than sigmoid [37]. The fifth layer is also a Dropout layer. The dropout rate was initially set to 0.7 to further reduce overfitting. The sixth layer is also a Dense layer. It has 2 neurons and Softmax as activation function. Softmax assigns probabilities to each class by outputting real values between 0 and 1 with a sum of 1 [1].

The Pro version of Google Colab cloud service was utilized to run the experiments, selecting the high RAM (34 GB) and GPU hardware acceleration setups. The Keras library built on a TensorFlow backend version 2.6.0 was used to train the model from scratch for 50 epochs, fixing initially the learning rate to 0.001 and the batch size to 10. The stochastic gradient descent was chosen as optimizer because, during training, the optimization based on a stochastic gradient is crucial to minimize the loss function while assuring higher

efficiency [1, 37]. The binary cross entropy was set as loss function. The early stopping callback with a patience of 5 was also used to further smooth out overfitting, as early stopping stops the training at the point where the validation loss, which captures exactly the divergence between the predicted output and the desired one, is minimal [6]. The training of the model took approximately 20 min for each experiment. Specifically, in the first branch of Brain-on-Cloud, a stratified shuffle-split cross validation was used in order to obtain a better approximation of the classification performance. In this regard, a randomized split of the OASIS-3 dataset into train (80%) and test (20%) sets was repeated five times, and 20% of the train set served as validation set in each data split. In the second branch, the model weights of the best split among the carried-out experiments were utilized to evaluate the performance of Brain-on-Cloud on all data (100%) of the ADNI-2 dataset, used as independent test set.

3.3. Experiments

For an in-depth investigation of the impact of data handling on the performance of Brain-on-Cloud, the following four experiments were conducted (Figure 1).

3.3.1. First experiment

In the first experiment, intensity normalization and automated cropping were added to the pipeline.

Since intensity normalization reduces the intensity variation caused by the use of different scanners or parameters for scanning different subjects [26], all raw sMRI data were normalized by subtracting the mean intensity and dividing the standard deviation. Next, each scan was automatically cropped by removing all black, non-informative voxels, ending up with different scan shapes. To have a dimensionally-uniform dataset, all scans were reshaped to 147 pixels × 192 pixels per slice, as it was the mean resolution of the cropped scans.

3.3.2. Second experiment

In the second experiment, the scan registration step was added to the pipeline.

This step is crucial for the alignment of multiple structures in order to verify their spatial correlation in anatomical terms [1, 26] by removing global differences and also resampling all scans to have 1-mm spatial resolution [10]. A common technique to perform scan registration is affine linear transformation [26]. For this, the linear image registration tool module of FMRIB Software Library (FSL)[8] was exploited. First, a dataset-specific template was generated by merging all AD and CN samples, respectively, and calculating their average. Then, all sMRI scans were registered using the average template, so that the spatial location of each anatomical structure of a sMRI scan perfectly matched with the one of another scan.

3.3.3. Third experiment

In the third experiment, the brain extraction step was added to the pipeline.

This step is fundamental to extract only brain voxels [26], which include the brain stem, cerebrospinal fluid, gray matter, white matter, sub-cortical structures and cerebellum [12]. For this, the brain extraction tool module of FSL was exploited, setting the fractional intensity threshold to 0.3 to reduce the bias without discarding any brain voxel [1].

3.3.4. Fourth experiment

In the fourth experiment, train data augmentation was added to the pipeline.

As AD affects both sides of the brain, augmentation via horizontal flipping was implemented only on the train set to exchange left and right symmetrical brain hemispheres [1, 24].

3.4. Classification and selection criteria

To evaluate the model performance, the following classification metrics were taken into account: ACC, SP, SE, F1-Score (F1-S), Receiver Operating Characteristic (ROC) curve and AUC. Specifically, a 0.5 discrimination threshold was chosen to compute ACC (2), SP (3), SE (4) and F1-S (5), regulated by:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

$$SP = \frac{TN}{TN + FP} \tag{3}$$

$$SE = \frac{TP}{TP + FN} \tag{4}$$

$$F1 - S = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \tag{5}$$

where TP stands for True Positive and it is the case where AD subjects are correctly classified as subjects affected by AD, TN stands for True Negative and it is the case where CN subjects are correctly classified as healthy subjects, FP stands for False Positive and it is the case where CN subjects are wrongly classified as subjects affected by AD, and FN stands for False Negative and it is the case where AD subjects are wrongly classified as healthy subjects.

To select the best experiment among the four experiments, all classification metrics were taken into account but more attention was paid to AUC and F1-S, as the first highlights the ability to discriminate between AD and CN without depending on the chosen discrimination threshold, whereas the second is the harmonic mean between precision and SE. It reflects how many AD cases were correctly identified. Specifically, FN must be as small as possible in medical tests, as a miss-classified pathology is the most dangerous problem in the clinical diagnosis [1, 20].

[8] https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSL

3.5. Hyper-parameter tuning

Since hyper-parameter selection has an high influence on the model performance, it needs to be carefully investigated. In this regard, a meticulous tuning of the hyper-parameters was performed for the best experiment in order to reach the best-performing configuration.

The dropout rate, learning rate and batch size are the hyper-parameters that impact the model performance most strongly [31]. Thus, all the possible combinations between the following values, chosen as they proved to guarantee both computational lightness and good speed of training in a preliminary experimental evaluation, were investigated:

- Dropout rate: [0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8];
- Learning rate: [0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05];
- Batch size: [1, 2, 3, 4, 5, 6, 7, 8, 9, 10].

To do so, the grid search algorithm was employed.

3.6. Diagnostic performance

To prove the robustness and generalization capability of Brain-on-Cloud, its diagnostic performance was evaluated on a cohort of 66 T1w 3D sMRI scans of 66 subjects belonging to an independent test set (ADNI-2) in relation to age, sex and MMSE. Specifically, subjects' prediction scores were categorized into:

- Four age groups, namely subjects < 65 years of age, $65 \leq$ subjects < 75 years of age, $75 \leq$ subjects < 85 years of age and subjects ≥ 85 years of age;
- Two sex groups, namely Female (F) and Male (M);
- Four MMSE groups, namely MMSE ≤ 18 (severe AD), $19 \leq$ MMSE ≤ 24 (from moderate to mild AD), MMSE $= 25$ (borderline) and $26 \leq$ MMSE ≤ 30 (CN).

Prediction score distributions were described in terms of $50^{th}[25^{th};75^{th}]$ percentiles for each group and compared with ground truth score distributions (1[1;1]) by means of the Wilcoxon Rank-Sum test [38], setting the statistical level of significance (P) to 0.05.

4. Results

Table 1 reports the performance in classifying AD of the four experiments and of the tuned best experiment, across the five splits of the stratified shuffle-split cross validation scheme, on OASIS-3 test data. Performance is given in terms of average ACC, SP, SE, F1-S and AUC, together with the respective standard deviation. Figure 4 focuses on the ROC curve and AUC values of the four experiments and of the tuned best experiment, across the five splits of the stratified shuffle-split cross validation scheme, on OASIS-3 test data.

The best experiment among the four experiments resulted to be the fourth. Regarding the best experiment hyper-parameter tuning, the best-performing combination for Brain-on-Cloud turned out to be the following:

Table 1
Performance in classifying AD of the four (1^{st}, 2^{nd}, 3^{rd} and 4^{th}) experiments and of the Tuned Best (TB) experiment, across the five splits of the stratified shuffle-split cross validation scheme, on OASIS-3 test data. Mean values of ACC, SP, SE, F1-S and AUC ± the respective standard deviation are reported.

	ACC (%)	SP (%)	SE (%)	F1-S (%)	AUC (%)
1st	75.27±6.15	68.46±11.77	81.38±8.04	77.64±5.07	80.63±7.80
2nd	77.09±9.54	75.38±16.25	78.62±6.69	78.68±7.89	88.08±10.26
3rd	85.45±7.54	84.62±12.64	86.21±3.78	86.46±6.34	92.09±8.33
4th	86.18±7.42	86.15±6.70	86.21±8.45	86.71±6.71	92.76±5.05
TB	**91.64**±6.57	**89.23**±14.06	**93.79**±4.02	**92.48**±5.34	**96.22**±5.06

- Best dropout rate value of the first Dropout layer: 0.6;
- Best dropout rate value of the second Dropout layer: 0.6;
- Best learning rate value: 0.005;
- Best batch size value: 6.

Table 2 reports the diagnostic performance of Brain-on-Cloud on independent ADNI-2 test data in relation to age, sex and MMSE, where AD and CN prediction score distributions are reported in terms of $50^{th}[25^{th};75^{th}]$ percentiles. It also reports the performance in classifying AD of Brain-on-Cloud in relation to age, sex and MMSE as well as its overall performance on independent ADNI-2 test data, both in terms of ACC, SP, SE, F1-S and AUC.

5. Discussion

This study proposed a robust yet lightweight 3D framework, Brain-on-Cloud, for automatically classifying AD from 3D sMRI whole-brain scans using a scalable GPU cloud service to ensure its public availability under copyright and, as consequence, a complete reproducibility of the entire algorithm. This outcome was obtained by improving our end-to-end ConvLSTM-based model dedicated to efficient learning of AD features by integration with a set of data handling techniques in addition to the tuning of the model hyper-parameters and the evaluation of its diagnostic performance on independent test data (Figure 1). In practice, four data handling pipelines were implemented for in-depth analysis of the impact of each step. In parallel, two branches were designed to investigate the diagnostic performance of Brain-on-Cloud. The feature extraction and classification tasks were pursued by performing end-to-end training, five-split cross validation and test on cloud, exploiting a fast, minutely-tuned neural network whose core layer is a ConvLSTM layer.

Results show how data handling techniques significantly affect the performance of Brain-on-Cloud, as reported in Table 1 and Figure 4. In particular, it was found that the addition of the brain extraction step affects the classification metrics more than all the other data-handling steps, increasing the average ACC of 9%, SP of 10%, SE of 7%, F1-S of 8%

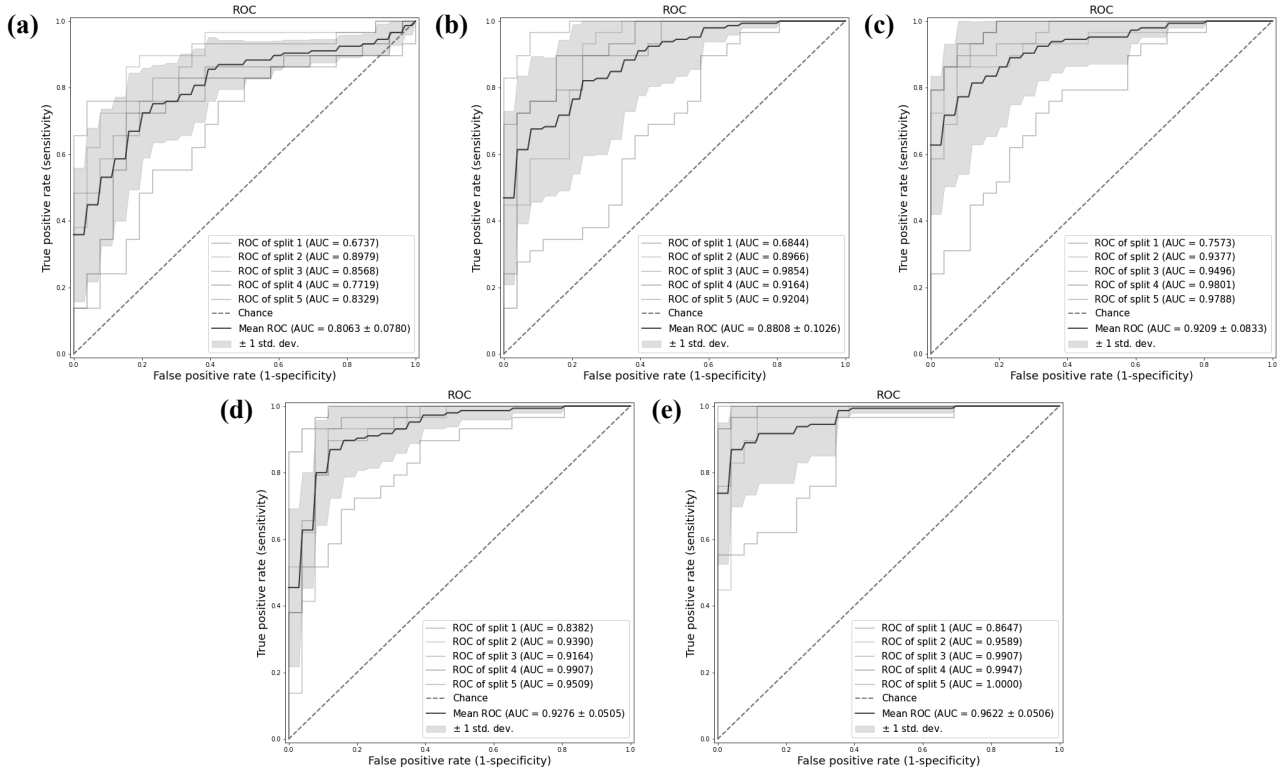


Figure 4: ROC and AUC values of the first experiment (**a**), second experiment (**b**), third experiment (**c**), fourth experiment (**d**) and tuned best experiment (**e**), across the five splits of the stratified shuffle-split cross validation scheme, on OASIS-3 test data. Mean ROC and AUC values of each experiment are reported in blue.

and AUC of 4%. The main reason behind this outcome is that brain extraction allowed the removal of all non-brain voxels (*i.e.,* noisy voxels), thus significantly reducing the confounders for the neural network in both feature extraction and classification tasks. Results highlight also the importance of hyper-parameter tuning, as reported in the last raw of Table 1 and in the panel (**e**) of Figure 4. With the grid search to find the best hyper-parameter combination for the best experiment among the carried-out experiments, Brain-on-Cloud registered a further increase of the average ACC by 6%, SP by 3%, SE by 8%, F1-S by 6% and AUC by 3%. Indeed, in its best guise, Brain-on-Cloud achieves an average ACC of 92%, SP of 89%, SE of 94%, F1-S of 93% and AUC of 96% on OASIS-3 test data. The diagnostic performance of Brain-on-Cloud was also investigated on an independent test set (ADNI-2). Using a more challenging evaluation protocol (*i.e.,* OASIS-3 as train and validation set, ADNI-2 as independent test set), Brain-on-Cloud obtained promising results. Indeed, it achieves an overall ACC of 76%, SP of 70%, SE of 82%, F1-S of 77% and AUC of 92% on independent ADNI-2 test data. Regarding the age groups, no statistically significant difference was observed between AD prediction score distributions and ground truth score distributions for ADNI-2 subjects younger than 65 and between both AD and CN prediction score distributions and ground truth score distributions for ADNI-2 subjects older than 75. Regarding the sex groups, no statistically significant difference was observed between both AD and CN prediction score distributions and ground truth score distributions for both females and males. Regarding the MMSE groups, no statistically significant difference was observed between CN prediction score distributions and ground truth score distributions for ADNI-2 subjects with a MMSE score higher than 26 as well as between AD prediction score distributions and ground truth score distributions for ADNI-2 subjects with a MMSE score lower than 24, whereas AD prediction score distributions were statistically different from ground truth score distributions for borderline subjects (MMSE = 25). These outcomes prove the robustness and generalization capability of Brain-on-Cloud in classifying severe AD and from moderate to mild AD, whereas it is still challenging to discriminate borderline AD cases from CN ones. Moreover, Brain-on-Cloud is robust enough to classify AD from both females and males as well as from subjects younger and older than respectively 65 and 75 years in age.

The strengths of Brain-on-Cloud with respect to the state-of-the-art frameworks are multiple, as reported in Table 3. Specifically, it summarizes the classification performance of Brain-on-Cloud and of the state-of-the-art methods cited in section 2 in terms of ACC, SE and AUC. It also reports the database(s) from which the datasets employed for training and testing belong to, the number of AD and CN subjects

Table 2
Diagnostic performance of Brain-on-Cloud on independent ADNI-2 test data in relation to age, sex and MMSE. Number of AD and CN subjects is given in percentage (%), whereas AD and CN prediction score distributions are reported in terms of $50^{th}[25^{th};75^{th}]$ percentiles. Performance in classifying AD is also reported in terms of ACC, SP, SE, F1-S and AUC.

		AD (%)	CN (%)	AD prediction ($50^{th}[25^{th};75^{th}]$)	CN prediction ($50^{th}[25^{th};75^{th}]$)	
Age	< 65	18	3	1[0.95;1]	0.64[0.64;0.64]*	
	65-75	27	73	0.77[0.50;0.94]*	0.88[0.60;0.99]	
	75-85	46	24	1[0.93;1]	0.87[0.61;0.99]	
	≥ 85	9	0	1[0.99;1]	n.a.	
Sex	F	24	70	0.97[0.75;1]	0.94[0.61;0.99]	
	M	76	30	1[0.70;1]	0.81[0.54;0.91]	
MMSE	≤ 18	21	0	1[1;1]	n.a.	
	19-24	49	0	1[0.94;1]	n.a.	
	= 25	30	0	0.64[0.42;0.95]*	n.a.	
	26-30	0	100	n.a.	0.93[0.66;0.99]	
		ACC (%)	SP (%)	SE (%)	F1-S (%)	AUC (%)
Age	< 65	83	100	80	89	100
	65-75	78	79	75	63	92
	75-85	86	75	93	90	98
	≥ 85	100	n.a.	100	n.a.	n.a.
Sex	F	80	78	86	67	97
	M	88	80	92	92	95
MMSE	≤ 18	100	n.a.	100	n.a.	n.a.
	19-24	93	n.a.	93	n.a.	n.a.
	= 25	67	n.a.	67	n.a.	n.a.
	26-30	75	75	n.a.	n.a.	n.a.
		ACC (%)	SP (%)	SE (%)	F1-S (%)	AUC (%)
Overall		76	70	82	77	92

*: $P < 0.05$, when comparing prediction vs. ground truth score distributions
n.a.: not applicable, if can not be computed

involved in the study, the learning algorithm, the input shape, the training type and the computing mode. The standard choice was to exploit a 3D CNN to automatically diagnose AD from sMRI scans. Although 3D CNN has the ability to preserve inter-slice context information, it comes with a high computational cost due to the high number of parameters [26]. Thus, it is not so clear how much performance is gained by using a 3D CNN to process volumetric data. Conversely, by using a ConvLSTM-based model like the one hereby proposed, it is possible to simultaneously process multiple slices of the same scan, preserving their spatial correlation in terms of anatomy, while ensuring high performance and low computational cost due to the reduced number of total parameters. Since ConvLSTM does not require an input tensor of a fixed shape, it is also possible to earn in scan resolution. Indeed, the resolution of 3D sMRI whole-brain scans used in this study is higher than the scan resolution of the other frameworks. In contrast to the state-of-the-art methods, which employed sMRI data selected from the same database, Brain-on-Cloud was also tested on subjects from two datasets of two different openly-available databases, which is more challenging but also more fair as it allows understanding the generalization capability of the proposed framework. Furthermore, the methods in Table 3 were designed without giving details on the sample IDentification numbers (IDs), whereas the complete list of the sample IDs used to feed Brain-on-Cloud will be provided on GitHub, together with the full source code of the implementation, in order to give the scientific community the possibility to exploit the same sMRI data and make objective comparisons with our framework. Another important difference from the state-of-the-art methods is that Brain-on-Cloud was developed with cloud computing, allowing a fast, easy and machine-independent reproducibility of the entire algorithm.

Brain-on-Cloud achieved competitive results in automatic diagnosis of AD, but its performance could be further improved. One limitation relies on the fact that only two classes were taken into account in this study, because the OASIS-3 dataset lacks labelled samples of all AD prodromal stages. Another limitation is that only one imaging modality was processed in this study. So, as part of future work, it is planned to extend the analysis to a multi-class/multi-modality classification, including from early-onset lo late-onset MCI and coupling sMRI with positron emission tomography. It is also planned to make the results interpretable for visually proving whether anterior hippocampal regions influenced the most the automatic decision-making process.

6. Conclusion

This study demonstrated that Brain-on-Cloud represents a reliable, efficient, lightweight and easily-reproducible method for automatic diagnosis of AD from 3D sMRI whole-brain scans. As Brain-on-Cloud performs well without segmenting the brain into its portions, it can also be applied to other neurological disorders using volumetric whole-brain data, as proven by Tomassini et al. [39]. There, the applicability of Brain-on-Cloud was extended to another neurodegenerative disorder (*i.e.,* Parkinson's disease) and to a psychiatric one (*i.e.,* schizophrenia), and its efficiency was confirmed. Furthermore, Brain-on-Cloud could find applicability on real-time diagnostic scenarios providing prompt clinical decision support, thanks to its cloud nature, computational lightness and fast execution.

CRediT authorship contribution statement

Selene Tomassini: Conceptualization, Data curation, Methodology, Software, Writing - Original draft preparation. **Agnese Sbrollini:** Methodology, Writing - Review and editing. **Giacomo Covella:** Software. **Paolo Sernani:** Writing - Review and editing. **Nicola Falcionelli:** Writing - Review and editing. **Henning Müller:** Project administration, Supervision. **Micaela Morettini:** Project administration, Supervision. **Laura Burattini:** Project administration, Supervision. **Aldo Franco Dragoni:** Project administration, Supervision.

Declaration of competing interest

The authors affirm that this manuscript is an honest, accurate and transparent account of the study being reported, that no important aspects of the study have been omitted, and that any discrepancies from the study as planned (and,

Table 3
Performance in classifying AD of Brain-on-Cloud and of the state-of-the-art methods doing a scan-based classification on 3D T1w sMRI whole-brain data. Values of ACC, SE and AUC are given in percentage (%). Database(s), AD and CN subjects, neural network, input shape, training type (scratch/transfer learning) and computing mode (cloud/remote) are also reported.

Framework	Database(s)	AD+CN	Model	Input shape	Training	Computing	ACC (%)	SE (%)	AUC (%)
Luo et al. [20]	Train: ADNI Test: ADNI	47+34	3D CNN	5×54×54	Scratch	Remote	-	100	-
Korolev et al. [30]	Train: ADNI Test: ADNI	50+61	3D CNN	110×110×110	Scratch	Remote	79	-	88
Bäckström et al. [31]	Train: ADNI Test: ADNI	-	3D CNN	110×110×110	Scratch	Remote	99	-	-
Basaia et al. [13]	Train: ADNI Test: ADNI	294+352	3D CNN	-	Transfer learning	Remote	99	99	-
Jabason et al. [17]	Train: OASIS Test: OASIS	-	Ensemble of CNNs	-	Transfer learning	Remote	95	-	-
Xia et al. [24]	Train: ADNI Test: ADNI	198+229	3D CNN+ConvLSTM	143×119×119	Scratch	Remote	94	94	96
Saratxaga et al. [32]	Train: OASIS Test: OASIS	128+177	3D CNN	176×176×176	Scratch	Remote	84	-	-
Brain-on-Cloud 1st branch	Train: OASIS Test: OASIS	145+130	ConvLSTM	50×147×192	Scratch	Cloud	92	94	96
Brain-on-Cloud 2nd branch	Train: OASIS Test: ADNI	149+137	ConvLSTM	50×147×192	Scratch	Cloud	76	82	92

if relevant, registered) have been explained. The authors declare no relationships with other people or organizations that could inappropriately bias the work. The authors confirm that neither the manuscript nor any parts of its content are currently under consideration or published elsewhere in any language.

Acknowledgment

The authors would like to thank the head of neurosurgery of Azienda Ospedaliero Universitaria Ospedali Riuniti di Ancona, dr. Roberto Trignani, for his clinical support in the development of the proposed framework.

References

[1] S. Tomassini, N. Falcionelli, P. Sernani, H. Müller, A. F. Dragoni, An end-to-end 3D convLSTM-based framework for early diagnosis of Alzheimer's disease from full-resolution whole-brain sMRI scans, in: IEEE International Symposium on Computer-Based Medical Systems, IEEE, 2021, pp. 74–78.

[2] J. Sun, S. Yan, C. Song, B. Han, Dual-functional neural network for bilateral hippocampi segmentation and diagnosis of Alzheimer's disease, International Journal of Computer Assisted Radiology and Surgery 15 (2020) 445–455.

[3] G. Folego, M. Weiler, R. F. Casseb, R. Pires, A. Rocha, Alzheimer's disease detection through whole-brain 3D-CNN MRI, Frontiers in Bioengineering and Biotechnology 8 (2020).

[4] J. Islam, Y. Zhang, Early diagnosis of Alzheimer's disease: A neuroimaging study with deep learning architectures, in: IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2018, pp. 1881–1883.

[5] S. Sarraf, G. Tofighi, A. D. N. Initiative, et al., DeepAD: Alzheimer's disease classification via deep convolutional neural networks using MRI and fMRI, BioRxiv (2016) 070441.

[6] Z. Guan, R. Kumar, Y. R. Fung, Y. Wu, M. Fiterau, A comprehensive study of Alzheimer's disease classification using convolutional neural networks, arXiv preprint arXiv:1904.07950 (2019).

[7] J. Wen, E. Thibeau-Sutre, M. Diaz-Melo, J. Samper-González, A. Routier, S. Bottani, D. Dormont, S. Durrleman, N. Burgos, O. Colliot, et al., Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation, Medical Image Analysis 63 (2020) 101694.

[8] L. G. Apostolova, P. M. Thompson, Mapping progressive brain structural changes in early Alzheimer's disease and mild cognitive impairment, Neuropsychologia 46 (2008) 1597–1612.

[9] S. Afzal, M. Maqsood, F. Nazir, U. Khan, F. Aadil, K. M. Awan, I. Mehmood, O.-Y. Song, A data augmentation-based framework to handle class imbalance problem for Alzheimer's stage detection, IEEE Access 7 (2019) 115528–115539.

[10] C. Lian, M. Liu, J. Zhang, D. Shen, Hierarchical fully convolutional network for joint atrophy localization and alzheimer's disease diagnosis using structural MRI, IEEE Transactions on Pattern Analysis and Machine Intelligence 42 (2018) 880–893.

[11] E. U. Haq, J. Huang, L. Kang, H. U. Haq, T. Zhan, Image-based state-of-the-art techniques for the identification and classification of brain diseases: A review, Medical & Biological Engineering & Computing (2020) 1–18.

[12] S. V. Yakkundi, D. Subha, Convolutional LSTM: A deep learning approach for dynamic MRI reconstruction, in: IEEE International Conference on Trends in Electronics and Informatics, IEEE, 2020, pp. 1011–1015.

[13] S. Basaia, F. Agosta, L. Wagner, E. Canu, G. Magnani, R. Santangelo, M. Filippi, A. D. N. Initiative, et al., Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks, NeuroImage: Clinical 21 (2019) 101645.

[14] N. Goenka, S. Tiwari, Volumetric convolutional neural network for Alzheimer detection, in: IEEE International Conference on Trends in Electronics and Informatics, IEEE, 2021, pp. 1500–1505.

[15] W. Yu, B. Lei, M. K. Ng, A. C. Cheung, Y. Shen, S. Wang, Tensorizing GAN with high-order pooling for Alzheimer's disease assessment, IEEE Transactions on Neural Networks and Learning Systems (2021).

[16] D. Le Bihan, How MRI makes the brain visible, in: Make Life Visible, Springer, 2020, pp. 201–212.

[17] E. Jabason, M. O. Ahmad, M. Swamy, Classification of Alzheimer's disease from MRI data using an ensemble of hybrid deep convolutional neural networks, in: IEEE International Midwest Symposium on Circuits and Systems, IEEE, 2019, pp. 481–484.

[18] P. D. Sloane, S. Zimmerman, C. Suchindran, P. Reed, L. Wang, M. Boustani, S. Sudha, The public health impact of Alzheimer's disease, 2000–2050: Potential implication of treatment advances, Annual Review of Public Health 23 (2002) 213–231.

[19] S. Todd, P. Passmore, Alzheimers disease, the importance of early detection, European Neurological Review 110 (2009).

[20] S. Luo, X. Li, J. Li, Automatic Alzheimer's disease recognition from MRI data using deep learning method, Journal of Applied Mathematics and Physics 5 (2017) 1892–1898.

[21] D. Ravì, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, G.-Z. Yang, Deep learning for health informatics, IEEE Journal of Biomedical and Health Informatics 21 (2016) 4–21.

[22] J. Ma, Y. Song, X. Tian, Y. Hua, R. Zhang, J. Wu, Survey on deep learning for pulmonary medical imaging, Frontiers of Medicine (2019) 1–20.

[23] S. Wang, D. M. Yang, R. Rong, X. Zhan, J. Fujimoto, H. Liu, J. Minna, I. I. Wistuba, Y. Xie, G. Xiao, Artificial intelligence in lung cancer pathology image analysis, Cancers 11 (2019) 1673.

[24] Z. Xia, G. Yue, Y. Xu, C. Feng, M. Yang, T. Wang, B. Lei, A novel end-to-end hybrid network for Alzheimer's disease detection using 3D CNN and 3D CLSTM, in: IEEE International Symposium on Biomedical Imaging, IEEE, 2020, pp. 1–4.

[25] W. Yu, B. Lei, S. Wang, Y. Liu, Z. Feng, Y. Hu, Y. Shen, M. K. Ng, Morphological feature visualization of Alzheimer's disease via Multidirectional Perception GAN, IEEE Transactions on Neural Networks and Learning Systems (2022).

[26] M. B. T. Noor, N. Z. Zenia, M. S. Kaiser, S. Al Mamun, M. Mahmud, Application of deep learning in detecting neurological disorders from magnetic resonance images: A survey on the detection of Alzheimer's disease, Parkinson's disease and schizophrenia, Brain Informatics 7 (2020) 1–21.

[27] N. Yamanakkanavar, J. Y. Choi, B. Lee, MRI segmentation and classification of human brain using deep learning for diagnosis of Alzheimer's disease: A survey, Sensors 20 (2020) 3243.

[28] S. Arukonda, S. Sountharrajan, Investigation of lung cancer detection using 3D convolutional deep neural network, in: IEEE International Conference on Advances in Computing, Communication Control and Networking, IEEE, 2020, pp. 763–768.

[29] J. Zhao, C. Zhang, D. Li, J. Niu, Combining multi-scale feature fusion with multi-attribute grading, a CNN model for benign and malignant classification of pulmonary nodules, Journal of Digital Imaging 33 (2020) 869–878.

[30] S. Korolev, A. Safiullin, M. Belyaev, Y. Dodonova, Residual and plain convolutional neural networks for 3D brain MRI classification, in: IEEE International Symposium on Biomedical Imaging, IEEE, 2017, pp. 835–838.

[31] K. Bäckström, M. Nazari, I. Y.-H. Gu, A. S. Jakola, An efficient 3D deep convolutional network for Alzheimer's disease diagnosis using MR images, in: IEEE International Symposium on Biomedical Imaging, IEEE, 2018, pp. 149–153.

[32] C. L. Saratxaga, I. Moya, A. Picón, M. Acosta, A. Moreno-Fernandez-de Leceta, E. Garrote, A. Bereciartua-Perez, MRI deep learning-based solution for alzheimer's disease prediction, Journal of Personalized Medicine 11 (2021) 902.

[33] B. S. Chera, R. J. Amdur, P. Patel, W. M. Mendenhall, A radiation oncologist's guide to contouring the hippocampus, American Journal of Clinical Oncology 32 (2009) 20–22.

[34] H. M. Duvernoy, The human hippocampus: Functional anatomy, vascularization and serial sections with MRI, Springer Science & Business Media, 2005.

[35] P. J. LaMontagne, T. L. Benzinger, J. C. Morris, S. Keefe, R. Hornbeck, C. Xiong, E. Grant, J. Hassenstab, K. Moulder, A. Vlassenko, et al., OASIS-3: Longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease, MedRxiv (2019).

[36] C. Savur, F. Sahin, Real-time american sign language recognition system using surface EMG signal, in: IEEE International Conference on Machine Learning and Applications, IEEE, 2015, pp. 497–502.

[37] H. Allioui, M. Sadgal, A. Elfazziki, Deep MRI segmentation: A convolutional method applied to Alzheimer disease detection, International Journal of Advanced Computer Science and Applications 10 (2019) 365–371.

[38] H. B. Mann, D. R. Whitney, On a test of whether one of two random variables is stochastically larger than the other, The Annals of Mathematical Statistics (1947) 50–60.

[39] S. Tomassini, P. Sernani, N. Falcionelli, A. F. Dragoni, CASPAR: Cloud-based Alzheimer's, schizophrenia and Parkinson's automatic recognizer, in: IEEE International Conference on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering, IEEE, 2022, p. in press.