



HAL
open science

Attention-Guided Neural Network for Early Dementia Detection Using MRS data

Anouar Kherchouche, Olfa Ben Ahmed, Carole Guillevin, Benoit Tremblais, Adrien Julian, Christine Fernandez-Maloigne, Rémy Guillevin

► **To cite this version:**

Anouar Kherchouche, Olfa Ben Ahmed, Carole Guillevin, Benoit Tremblais, Adrien Julian, et al.. Attention-Guided Neural Network for Early Dementia Detection Using MRS data. Computerized Medical Imaging and Graphics, 2022, 99, pp.102074. 10.1016/j.compmedimag.2022.102074 . hal-03684564

HAL Id: hal-03684564

<https://hal.science/hal-03684564>

Submitted on 22 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Attention-Guided Neural Network for Early Dementia Detection Using MRS data

Anouar Kherchouche^{a,c,*}, Olfa Ben-Ahmed^{a,c}, Carole Guillevin^{b,c}, Benoit Tremblais^{a,c}, Julian Adrien^{a,c}, Christine Fernandez-Maloigne^{a,c}, Rémy Guillevin^{b,c}

^a*xlim research institute, umr cnrs 7252 university of poitiers, poitiers, france.*

^b*dactim-mis, lma, umr cnrs 7348, university and hospital of poitiers, france.*

^c*i3m, common laboratory cnrs-siemens, university and hospital of poitiers, france.*

Abstract

Imaging bio-markers have been widely used for Computer-Aided Diagnosis (CAD) of Alzheimer's Disease (AD) with Deep Learning (DL). However, the structural brain atrophy is not detectable at an early stage of the disease (namely for Mild Cognitive Impairment (MCI) and Mild Alzheimer's Disease (MAD)). Indeed, potential biological bio-markers have been proved their ability to early detect brain abnormalities related to AD before brain structural damage and clinical manifestation. Proton Magnetic Resonance Spectroscopy (¹H-MRS) provides a promising solution for biological brain changes detection in a no invasive manner. In this paper, we propose an attention-guided supervised DL framework for early AD detection using ¹H-MRS data. In the early stages of AD, features may be closely related and often complex to delineate between subjects. Hence, we develop a 1D attention mechanism that explicitly guides the classifier to focus on diagnostically relevant metabolites for classes discrimination. Synthetic data are used to tackle the lack of data problem and to help in learning the feature space. Data used in this paper are collected in the University Hospital of Poitiers, which contained 111 ¹H-MRS samples extracted from the Posterior Cingulate Cortex (PCC) brain region. The data contain 33 Normal Control (NC), 49 MCI due to AD, and 29 MAD subjects. The proposed model achieves

*Corresponding author

Email address: anouar.kherchouche@univ-poitiers.fr (Anouar Kherchouche)

an average classification accuracy of 95.23%. Our framework outperforms state of the art imaging-based approaches, proving the robustness of learning metabolites features against traditional imaging bio-markers for early AD detection.

Keywords: Computer-Aided Diagnosis, Deep learning, Attention mechanism, Feature refinement, Alzheimer’s disease, Magnetic resonance spectroscopy

1. Introduction

Alzheimer’s Disease (AD) is the most frequently occurring progressive neurodegenerative disease that is increasing in prevalence and cost at unsustainable levels Mayeda et al. (2016). Patients with AD suffer from the cognitive decline that leads to brain neurons and synaptic loss (i.e. memory loss, difficulty with problem-solving, etc). Although there is currently no cure for AD, there are available medications that can slow down disease progression and improve the patient lifestyle López et al. (2020). Recent studies on bio-markers research have demonstrated that the AD pathology is now suspected to start a long time before the manifestation of the clinical symptoms and even before brain damage Morley et al. (2018). Hence, diagnosis of AD at earlier stages is of great clinical importance so that cognitive functions would be improved by medications and the spread of the disease would be prevented. Mild Cognitive Impairment (MCI) is an intermediary stage condition between healthy people and AD Gallaway et al. (2017). Detecting MCI subjects provide a potential window for early AD detection. However, MCI subjects detection remain a challenging clinical problem as it lies on a spectrum between NC and manifest AD Honig et al. (2018). Before the confirmed AD, Mild Alzheimer’s Disease (MAD) is the next stage after the MCI condition due to AD. Compared to the moderate and severe stages of AD, MAD patients retain independence in simpler activities Knopman & Petersen (2014). Therefore, identifying efficient bio-markers for early AD stages detection helps in establishing diagnosis and treatment strategies without delay.

Over the last decades, imaging-based bio-markers have been extensively in-

investigated for AD detection through Magnitude Resonance Imaging (MRI) techniques Young et al. (2020); Ben Ahmed et al. (2015). These approaches are mainly based on the quantification of brain volume loss and the neuronal degeneration process. However, these structural atrophies occur once brain damage has already begun and hence cannot provide diagnosis at an early stage. Therefore, imaging-based bio-markers failed to accurately distinguish between patients with MCI and MAD. An efficient bio-marker that could detect early brain changes before neuronal death would be valuable. In this context, several studies in the medical domain have identified biological bio-markers that could detect early brain changes before neuronal death Bateman et al. (2012); Beason-Held et al. (2013). These studies are mainly based on Cerebrospinal Fluid (CSF) which can detect brain damage before 15 to 20 years of the first clinical signs of AD Andreasen et al. (2001). However, CSF bio-markers are collected by lumbar puncture and require an invasive procedure that is not recommended for routine evaluation of early dementia Menéndez-González (2014).

Magnetic Resonance Spectroscopy (MRS) is a non-invasive technique providing a complementary approach to brain metabolism in vivo, during conventional MRI examination. MRS provides biological information of brain tissues at the molecular level allowing detecting brain abnormalities while MRI remains normal. Indeed, studies on brain metabolism have demonstrated differential metabolite measurements across early AD stages Graff-Radford & Kantarci (2013); Maul et al. (2020). This ability to safely assess brain biological information made MRS an efficient tool for virtual biopsy of early AD detection. Although, the importance of MRS in the detection of brain chemical components, their investigation for CAD of AD remains scarce. Indeed, DL is a powerful paradigm for medical data analysis Bhatt et al. (2021) and MRS data opens up new perspectives for DL-based CAD systems development of early AD.

However, compared to imaging data, MRS data are more challenging. First, MRS are complex and high dimensional data with a high correlation between metabolite spectral patterns. Moreover, MRS data are scarce and hence require designing specific deep architectures and efficient features learning strategies.

To the best of our knowledge, there is no work that has investigated DL for early AD detection with MRS data. Yet, there is no public MRS data for CAD.

In this paper, inspired by the clinician diagnosis process, we propose a self-guided attention DL framework for CAD of early AD. The contributions of the paper can be summarized as follows:

- Biological bio-markers (brain metabolites) from non-invasive technique with (^1H -MRS) data are investigated for early AD detection.
- Unlike most existing works for binary classification, we address the problem of multi-class classification for MCI, MAD, and NC categories discrimination. Indeed, the multi-classification strategy is more suitable in clinical practice.
- We propose a method that guides the network to learn discriminative class-related metabolites for fine-grained classification. Hence, we develop a 1D attention mechanism for disease-related features learning. Indeed, the latter is designed to build an attention map that represents the important metabolites in spectrum recognition and uses it to refine the features space in a supervised fashion.
- We tackle the problem of the lack of MRS data by simulating MRS data from real spectra. Indeed, synthetic data help in learning the latent features space and the attention mechanism guides the model to refine these learned features.
- Finally, we give an interpretation and explanation of the proposed model using a 1D attention map designed for MRS data.

The rest of this paper is organized as follows: First, Section 2 reviews State Of The Art (SOTA) deep learning based approaches for AD detection. Second, Section 3 describes the proposed subject classification framework. Then, experiments and results are presented and analyzed in Section 4. Finally, Section 5 concludes the paper and opens new perspectives.

2. Related Work

During the last years, DL techniques have been widely investigated for AD detection using medical imaging data. In this section, we review some SOTA deep learning-based approaches for early AD detection. Then, we present SOTA contributions for CAD using MRS data.

2.1. Deep learning for early CAD of AD

Two-Dimensional (2D) Convolutional Neural Network (CNN) models have been widely used for feature learning from 2D MRI slices for AD detection. Suk et al. (2015) proposed to use a stack sparse auto-encoder for feature extraction along with a Support Vector Machine (SVM) for classification. Martinez-Murcia et al. (2019), the authors proposed a deep convolutional auto-encoder to compress MRI slices into low-dimensional representations and use SVM and Neural Network (NN) for AD subjects classification. Moreover, 2D CNN models have been used with transfer learning to identify AD Ebrahimi-Ghahnavieh et al. (2019) Aderghal et al. (2017). However, the shortcomings of the 2D CNN-based methods are that the spatial and the inter-slice information of the 3D images are not fully explored. 3D CNN has been used to address this issue. For instance, Hosseini-Asl et al. (2016), utilized a pre-trained 3D Adaptive CNN classifier for the classification of AD vs NC. Korolev et al. (2017) used 3D residual NN architecture together with several regularization techniques for AD classification. Recently, Dufumier et al. (2021) proposed a 3D contrastive learning framework for AD classification. The proposed contrasting scheme is based on the assumption that similar metadata share similar discriminative features. However, 3D CNNs are compute and memory-intensive limiting their use in clinical practice. Other studies focused on combining 2D CNNs with Recurrent Neural Networks (RNNs) to aggregate the intra-slice information. Liu et al. (2018) proposed a classification framework based on 2D CNN and RNN from PET scan. They used 2D CNN to capture the slice features with RNN to learn and integrate the inter-slice features. Li & Liu (2019) combined CNN with RNN based on

internal and external hippocampal to perform binary distinction, i.e., (AD vs NC), (MCI vs AD), etc.

In order to boost the performance of early AD detection, information fusion have been explored by either combining several imaging modalities Feng et al. (2019) or by adding demographic features (cognitive scores, Age, etc) Khatri & Kwon (2020). However, the necessity of balanced data made fusion approaches not generic, since the model needs to use all modalities/information to give a prediction. Most DL models proposed for AD diagnosis are formulated as binary classification problems (AD vs NC, NC vs MCI, and MCI vs AD) far from meeting the clinical demand. Although CNN learns meaningful and high-level features, they are known to learn redundant features leading to poor classification performance. Therefore, learning to suppress irrelevant features and keep only the important ones is very needed for early AD diagnosis. Based on the idea that pertinent information in medical imaging are located in some regions of interest, attention mechanism has been investigated for CAD Schlemper et al. (2019) in 2D images. Indeed, attention mechanisms improve the classification performance and provide decision interpretation which is important in the medical domain Bourdon et al. (2021). For instance, Zhang et al. (2021), proposed a 3D explainable network improved by a self-attention network with residual connection to automatically extract discriminative localization on sMRI for efficient diagnosis.

2.2. CAD approaches using MRS data and deep learning

Deep learning has been recently explored for MRS data analysis Chen et al. (2020), especially for data denoising Hu et al. (2021) and metabolites quantification Lee & Kim (2020) Lee & Kim (2019) Javid & Buzdar (2021). More recent DL-based approaches have been proposed for brain tumor detection using MRS data Menze et al. (2021) Lu et al. (2021a). Lu et al. (2021b) proposed a DL framework to detect brain tumors with data augmentation and distillation. Distilling and raising the number of data improved the performance of the model and outperformed the human expert-level prediction. Jang et al. (2021)

proposed to train a Generative Adversarial Network (GAN) in an unsupervised manner using ^1H -MRS of the brain, which is capable of filtering out abnormal spectra with different range of abnormalities from the normal ones. Dandil & Karaca (2021) proposed a DL model based on Long Short-Term Memory (LSTM) and Bidirectional LSTM for binary classification able to distinguish between brain tumors and normal brain tissue with pseudo-brain tumors. They augmented their training data using the normal distribution to generate new samples. Previous works proved that MRS data opens up new perspectives in CAD, especially with the use of Deep Neural Networks (DNNs) models.

Although, the importance of MRS in the detection of brain chemical components, their investigation for an individual diagnosis of AD remains scarce. To the best of our knowledge, there are no studies that have investigated DL to detect AD at an early phase using MRS data. Only one recent work Ahmed et al. (2020) has used ^1H -MRS data for binary classification of AD and NC subjects. This work outperformed SOTA structural MRI-based approaches for AD vs NC detection. Indeed, the authors proposed to learn a 1D-CNN model using the real and the imaginary parts of the ^1H -MRS signal. However, including the imaginary part in the model negatively impacts the interpretation of the classification results because clinical information is located only in the real part. Moreover, the authors used Global Average Pooling (GAP) layer to build a Class Activation Maps (CAM) for model interpretation. However, the GAP layer ignores the local spectrum details which are important for fine-grained classification (early AD stages distinguishing).

3. Methodology

3.1. Medical background and research hypotheses

In the case of AD, the entorhinal and hippocampal cortices show some differences at the beginning of this dementia Huang et al. (2002). Furthermore, recent brain imaging studies Lee et al. (2020) indicate that the Posterior Cingulate Cortex (PCC) is more closely associated with early AD compared to other

regions. In fact, the damage of the PCC region is frequently present before the hippocampus neurodegeneration Leech & Sharp (2014). In addition, the position of the PCC region and its clear structure (compared to the hippocampus) permits obtaining efficient ^1H -MRS data.

Indeed, the PCC region includes the connectivity between these regions and the whole brain regions Leech & Sharp (2014). Yet, abnormal metabolic ratios in the PCC have been identified with MRS in early AD Kantarci et al. (2007). In addition, the position of the PCC region and its clear structure (compared to the hippocampus) permits obtaining efficient ^1H -MRS data. Figure 1 shows three projections of a brain MRI of a MAD subject. Blue boxes indicate the position of the PCC region.

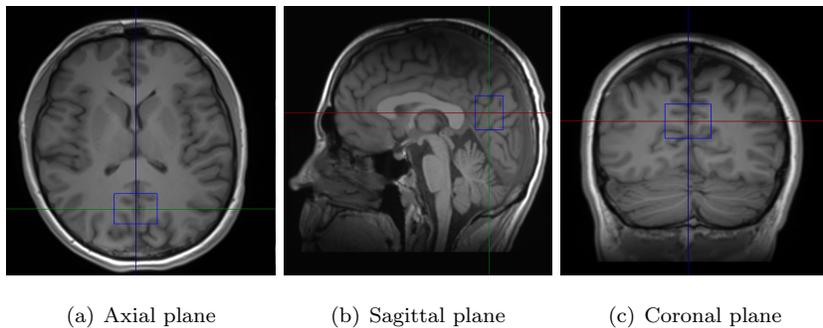


Figure 1: PCC region placement on brain MRI ((a) Axial plane, (b) Sagittal plane and (c) Coronal plane) of a MAD subject from Poitiers University Hospital’s dataset

3.2. Self-guided attention deep network for AD diagnosis

In this work, we propose a 1D CNN model for MCI, MAD, and NC subjects detection using ^1H -MRS data. The proposed framework is presented in Figure 2. As we can see, the model is composed of three main sub-networks. A features extractor block, an attention network block, and a classification block. The attention block is used to capture and select relevant features during the training. It guides the classifier to focus on the most important features for groups classification to ignore the non-relevant ones. In our case, we propose a 1D attention mechanism adapted for MRS data, (i.e., allows our model to focus only on the

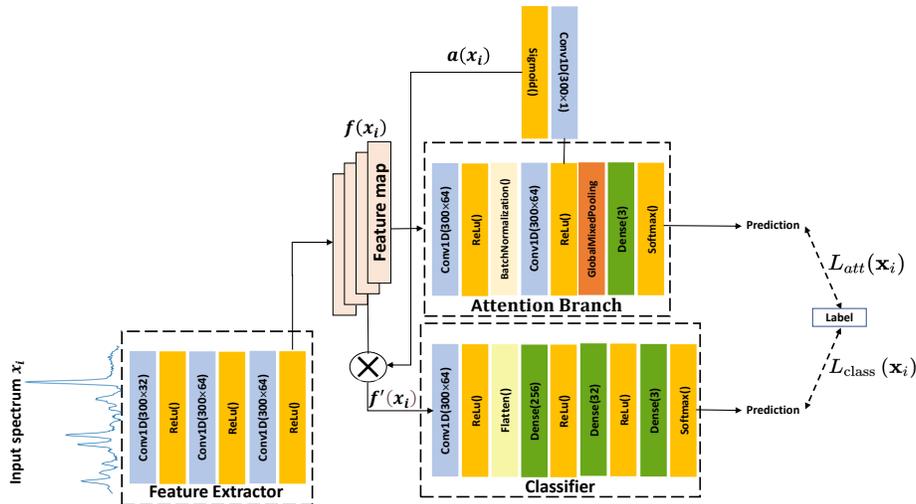


Figure 2: Self-Guided attention deep network.

most relevant metabolites in order to distinguish between categories). To tackle the problem of small MRS data for features space learning, we synthesize MRS data using a Deep Neural Network (DNN). Then, learned features are refined using the attention mechanism. The model is trained in an end-to-end manner by minimizing the sum of the respectively classifier and the attention network losses:

$$L_{total}(\mathbf{x}_i) = L_{att}(\mathbf{x}_i) + L_{class}(\mathbf{x}_i)$$

$$\text{with } L_{att} = L_{class} = - \sum_{i=1}^3 t_i \log(p_i) \quad (1)$$

where t_i is the truth label and p_i is the softmax probability for the i^{th} class.

3.2.1. Features extractor block

This block is used to extract the learned feature maps $f(x_i)$ for each spectrum input x_i . Based on medical domain knowledge, we used only the real part of the spectral signal. The architecture of this network contains three layers of 1D convolution (with ReLU as activation function) using respectively 32, 64 and 64 filters with kernels size of 8, 5 and 8. We use padding for each convolution layer to keep the same input size. Hence, we use the same padding type,

which appends zero values in the extremities of the spectrum without affecting the semantic information. Thus, the convolution filters can cover the borders of the spectrum for further processing. The feature extractor is optimized by passing through the gradients of the attention and classification branches during back-propagation. From one training epoch to another, the features extracted became more discriminative.

3.2.2. Attention block for pertinent metabolites learning

The attention block uses an attention mechanism that teaches the model to focus on important metabolites and ignores irrelevant ones for classes discrimination in a supervised fashion. Indeed, the attention block is designed to take as input the feature maps $f(x_i)$ to build an attention map $a(x_i)$ which is then used to focus on discriminative parts of signal in an end-to-end learning process. This attention map is equivalent to a mask that filters only the important information of each feature map from the features extractor block. Yet, this map allows giving interpretation and explanation of the pertinent features for each spectrum. The attention block architecture is composed of two convolution layers using 64 filters and kernel sizes of 5 and 8 respectively. This block includes a batch normalization layer after the first convolution layer to stabilize the learning to reduce model over-fitting. In order to build $a(x_i)$, we rely on the last layer with a convolution layer of 1 filter, this layer is followed by a sigmoid function for normalization.

Instead of using the traditionally used Global Average Pooling (GAP) for attention mechanisms Zhou et al. (2016) on 2D images, we propose to adapt this layer to our MRS data characteristics. Indeed, the GAP lets the model focus on the globality of the spectrum to distinguish the classes, which is not sufficient for clinicians that need also to get the importance of each metabolite per class. In fact, in clinical practice, clinicians focus on an important peaks in the signal while keeping the global context. In this work, we argue that attention can benefit from both GAP and Global Max Pooling (GMP). Hence, we propose to design a Global Mixed Pooling G_{mix} layer that would investigate

the most prominent features by integrating the advantages of both GAP (for global context) and GMP (for local context). The G_{mix} layer is a weighted linear combination of GMP and GAP layers as shown in equation 2.

$$G_{mix}(\mathcal{F}) = \alpha.G_{max}(\mathcal{F}) + (1 - \alpha).G_{avg}(\mathcal{F}) \quad (2)$$

where $\mathcal{F} = \{F_1, F_2, \dots, F_k\}$, represents the feature maps of the last convolution layer containing k feature map F_i , $\alpha \in \mathbf{R}$ is a none trainable parameter, and G_{max} , G_{avg} are the (GMP) and (GAP) layers that calculate respectively the global maximum and average for each F_i and then append all the values to get a vector. These layers are defined as follows:

$$G_{max}(\mathcal{F}) = [m_1, m_2, \dots, m_k]^\top \quad (3)$$

with $m_i = \max F_i$, $i \in \{1, 2, \dots, k\}$

and

$$G_{avg}(\mathcal{F}) = [v_1, v_2, \dots, v_k]^\top \quad (4)$$

with $v_i = \frac{1}{W} \sum_{i=1}^W (F_i)$, $i \in \{1, 2, \dots, k\}$

where W is the length of F_i . Thus, we get an output vector which has a length of $k = 64$ constituted from \mathcal{F} . The G_{mix} layer adapts the attention branch network to help the classifier block. In fact, the G_{mix} is expected to build a CAM for model interpretation. However, traditional CAM cannot generate attention maps in the training stage. Thus, we hypothesize that using attention map for features filtering in the training process would help the classifier to filter the non-pertinent information for each class and hence learning discriminative class-related features.

This layer is followed by a dense layer of 3 neurons and a softmax activation function to get prediction from the attention block.

3.2.3. Classification network for subject class detection

Finally the classification block outputs the final probability of each class. The input of the classifier block $f'(\mathbf{x}_i)$ is fed by the multiplication of the feature

maps $f(\mathbf{x}_i)$ that were come out from the extractor block with the attention map $a(\mathbf{x}_i)$ as illustrated in equation 5. The attention map reflects the metabolites on input spectrum which support the network’s prediction.

$$f'(\mathbf{x}_i) = f(\mathbf{x}_i) \cdot a(\mathbf{x}_i) \quad (5)$$

The classifier block contains one convolution layer of 64 filters and 5 as kernel size, followed by three dense layers of 256, 32, and 3 neurons followed by a softmax function to get the final predictions.

3.2.4. Features space augmentation using synthetic data

To deal with the challenges of learning from small MRS data, we propose to augment the features space by learning from synthetic data. In order to generate synthetic ^1H -MRS spectra for each class, we use a DNN model composed of two fully-connected layers followed by a (LeakyReLU = $\max(0.1c, c)$, where c is the input) and a linear activation function respectively. The first fully connected layer is fed by a random noise input of dimension 1024 (size of the spectrum) with 128 neurons. For the remaining dense layer (output), we use 1024 neurons.

We employ the available ^1H -MRS data of each class to generate artificial spectra. Thus, we train one model for each class, trying to minimize the Mean Squared Error (MSE) (Equation 6) between a real spectrum x_i which is the ground truth and the generated spectrum \tilde{x}_i .

$$\text{MSE} = \frac{1}{n - \kappa} \sum_{i=1}^{n-\kappa} (x_i - \tilde{x}_i)^2 \quad (6)$$

where n is the total number of spectra for a specific class and κ is a reserved number of spectra for each class for the final framework test. These reserved spectra are independent from the synthetic spectra used to train the classifier model. Thus, we hypothesize that correctly classifying these κ spectra in the test phase would design a generalized method.

4. Experiments and Results

In this section, we present the experimental setup and classification results. First, we present the used material for our MRS data acquisition and collecting procedure. Then, we showcase the model parameters and the training settings as well as the obtained results. After that, an ablation study is performed to study the efficiency of the attention mechanism and the coherence of the model prediction with medical domain knowledge in AD. Finally, we compare our framework with SOTA methods.

4.1. MR Spectroscopy data acquisition, collection and pre-processing

4.1.1. Data acquisition and preparation

¹H-MR spectra were acquired with a single voxel sequence, TR = 1700 ms, TE= 135 ms, number of excitation ((NEX)=160) with a voxel size of $28 \times 28 \times 20$ mm³. To facilitate voxel placement, high-resolution MR images were acquired with a T1-weighted Magnetization-Prepared Rapid Gradient-Echo (MPRAGE) with the following parameters: TR = 18000 ms, TE= 2.24 ms, TI= 900 ms, 176 axial slices and a spatial resolution of 1 mm³. The jMRUI plugin ¹ is used to process the MRS data. The data used in this work contain 3 Tesla 111 ¹H-MRS spectra divided into three groups namely MCI, MAD, and NC with an average Mini-Mental State Examination (MMSE) score of 26.73 ± 2.00 , 20.54 ± 3.06 and 29.5 ± 1.00 respectively.

The individual labels of the data were achieved via the diagnosis of a practiced physician in which the clinical and radiological information of each patient were incorporated. Knowing that all the subjects go through a brain examination by a conventional MRI, where the anatomical T1-weighted MRIs were recorded for each patient. This exam included also a hydrogen-based MRS sequence with the use of a Siemens 64-channel 3-Tesla "Skyra" whole-body MRI system (Siemens Healthcare, Erlangen, Germany).

¹<http://www.jmrui.eu/plugins-overview/>

4.1.2. MRS extraction from PCC

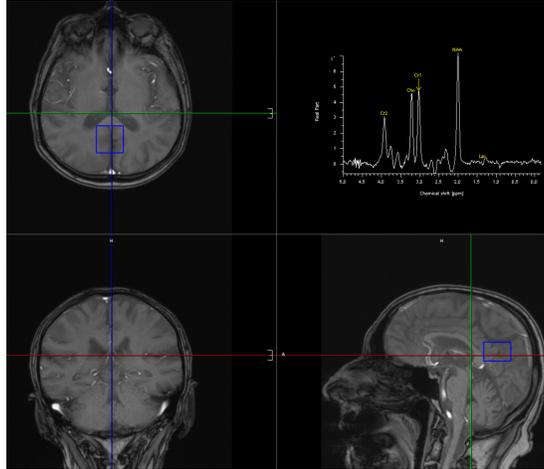


Figure 3: T1-weighted MRI for a random patient from Poitiers University Hospital’s dataset with the three projection and its corresponding ^1H -MRS spectrum from the PCC region.

Spectroscopy acquisitions were performed in the PCC region. Figure 3 shows the placement of the PCC region (blue box) and its corresponding spectrum in top right. The extraction of the ^1H -MRS was in a Single Voxel Sequence (SVS) manner, i.e., acquiring directly our volume of interest which is the PCC. After obtaining all the PCC spectra, a Fourier Transform with a shift was applied to move from the spatial to the frequency domain. Finally, each patient has its ^1H -MRS complex spectrum. Based on clinical domain knowledge Öz et al. (2021); Near et al. (2021), we use the spectrum’s real part for AD stages discrimination. Moreover, we select the range of 300 frequencies after the water peak to learn the classification model. Indeed, relevant metabolites such as N-acetyl-aspartate+N-acetyl-aspartyl-glutamate (NAA), Creatine/phosphocreatine (Cr), Choline (Cho), and myo-Inositol (mI) are only in the real part situated on the right side of the water peak.

4.2. Experimental setup

For model training, the sum of two categorical loss functions for both attention and classification blocks (see equation 1) is minimized using Adam opti-

mizer with learning rate = 0.001, $\beta_1 = 0.9$ and $\beta_2 = 0.999$.

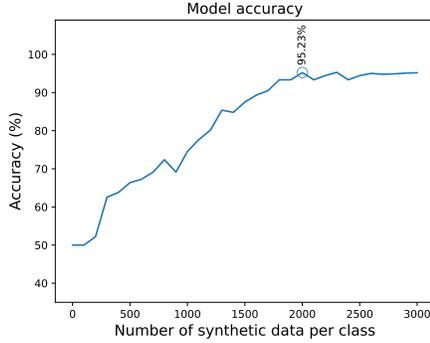


Figure 4: Variation the number of training synthetic spectra per class and their corresponding classification accuracy.

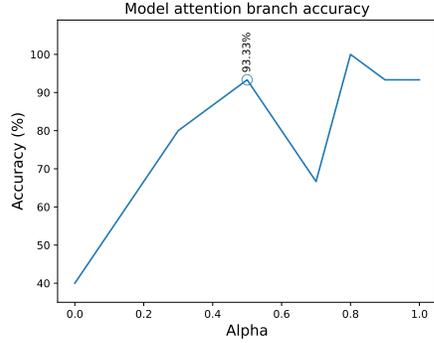


Figure 5: Variation of α value for the Global Mixed Pooling layer and its corresponding attention branch accuracy.

The model is trained using randomly select 80% samples from the generated synthetic spectra, while keeping the remaining 20% for the validation set. We train the model with a batch size of 128 for 150 epochs. We recall here that when generating synthetic data, we left 5 spectra per class for model testing. In order to select the optimal number of generated spectra for model training, we vary the number of training synthetic spectra per class from 0 to 3000 samples per class and we compute the corresponding classification accuracy with test data. As shown in Figure 4, the accuracy increases proportionally with the number of used data, until the value of 2000 samples per class, where the accuracy starts to stabilize.

In order to find the optimal value of parameter α for the proposed Global Mixed Pooling layer, we varied α from 0 to 1. Figure 5 represents the variation of α over the training accuracy of the attention block. The best accuracy value is obtained with $\alpha = 0.8$, where we get nearly 100% prediction accuracy. However, with this value of α the model produces much sparser attention maps which affect the classifier block inputs and thereby reduce the accuracy of this block. The second-best accuracy is when $\alpha = 0.5$, where we obtained an accuracy of 93.33%. The involvement of the GAP decreases the performance of

the attention branch. However, the attention maps are less dense which encourages the classifier block to be stable and focus on features with respect to the distinction between the three classes.

4.3. Results and evaluation

4.3.1. Synthetic data generation

Data are generated from 1024-D spectra with the water peak following the data generation scheme described in (Section 3.2.4) where a set of spectra is used for training. For instance, given the 33 spectra of the NC class, we iteratively generate synthetic ^1H -MRS data from 28 spectra.

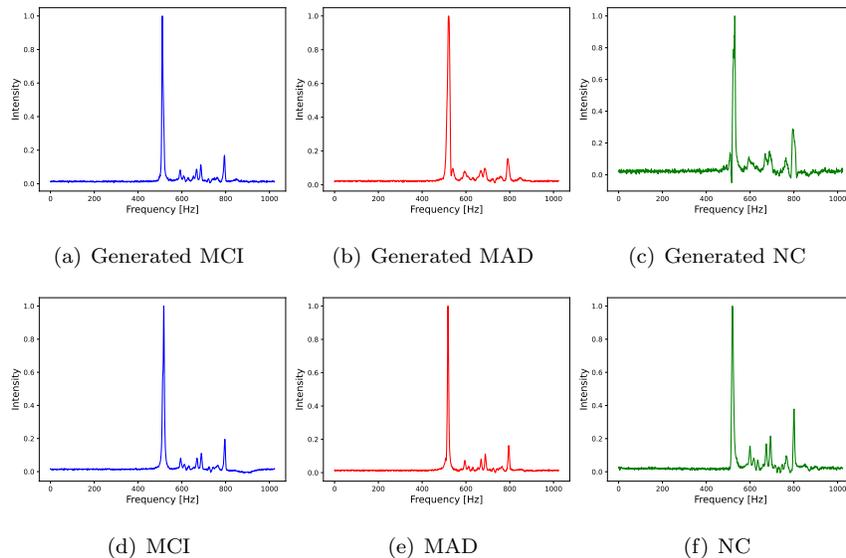


Figure 6: Synthetic generated MRS data of MCI (blue), MAD (red) and NC (green) classes 6(a), 6(b), 6(c) compared with spectra from the ground truth dataset 6(d), 6(e), 6(f).

Thus, after deploying the model, we can generate as many spectra as we want. The reserved 5 spectra will be used for testing the proposed classification model, i.e., we test on the signals 1 to 5 in the first iteration, 6 to 10 in the second, and so on. By this way, we are performing cross-validation to prove that our proposed method is generic and does not depend on training data. Figure

6 shows some examples of real spectra for the three classes with the synthetic ones, which are very close to the original data.

4.3.2. Classification results

We compute the accuracy, sensitivity, and specificity of the model. In these experiments, the strategy “One-Versus-Rest” (OVR) is adopted to expand the model for multi-class classification. For class C, we compute “C versus the rest of the classes” metrics. We compute these metrics for respectively MCI versus (NC and MAD), MAD versus (NC and MCI), and NC versus (MCI and MAD).

Table 1: Evaluation of the data augmentation performance using a cross-validation method.

Fold	Accuracy	Sensitivity			Specificity		
		MCI vs rest	MAD vs rest	NC vs rest	MCI vs rest	MAD vs rest	NC vs rest
1	100%	100%	100%	100%	100%	100%	100%
2	93.33%	100%	80%	100%	90%	100%	100%
3	93.33%	100%	80%	100%	90%	100%	100%
4	93.33%	80%	100%	100%	100%	90%	100%
5	86.66%	80%	100%	80%	90%	90%	100%
6	100%	100%	100%	100%	100%	100%	100%
7	100%	100%	100%	100%	100%	100%	100%
8	93.33%	100%	80%	100%	90%	100%	100%
	95.23% ± 5.04%	95.00% ± 9.25%	93.33% ± 10.00%	97.50% ± 7.07%	95.00% ± 5.34%	97.50% ± 4.62%	100.00% ± 0.00%

Table 1 gives the performance of our method in terms of accuracy, sensitivity, and specificity. Each row in this table represents the test performances of the reserved five spectra for each class. Thus, each row contains 15 spectra as a data test. The average accuracy is 95.23% with a standard deviation of 5% that was calculated with the equation 7.

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (acc_i - \bar{acc})^2}{n - 1}} \quad (7)$$

where acc_i is the accuracy of the i^{th} row in the table, \bar{acc} is the mean accuracy and n is the number of the used iterations. We report also the sensitivity for each class, where we get an average of 95.00%, 93.33%, and 97.5% as sensitivity for the three classes MCI, MAD, and NC respectively. These results show the robustness of the proposed method, even between the similar two classes (i.e.,

MCI and MAD), our model can distinguish the two classes. The results of the specificity confirm the performance of the framework and how well our model recognizes the negative samples for a given class, where we obtain an average specificity of 95.00%, 97.50%, and 100% for MCI, MAD, and NC respectively. We can observe that the specificity of the MCI class is relatively less than the rest of both classes, which is very logical since this class is between the two phases. Therefore, the misled subjects from the NC and MAD classes are generally classified as MCI. Figure 7 represents the mean ROC of each class

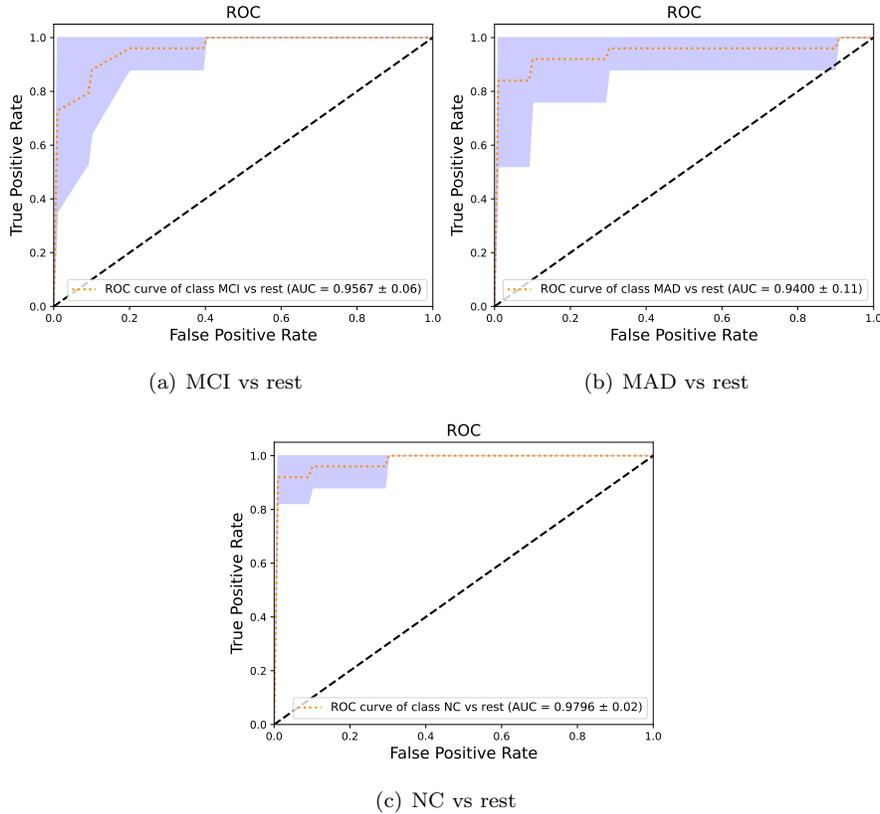


Figure 7: Mean Receiver operating characteristic (ROC) curve of each class with its standard deviation.

with respect to the remaining ones using the 8 models. We reported an average Area Under the Curve (AUC) of 0.957 ± 0.06 , 0.940 ± 0.11 , and 0.980 ± 0.02 for

MCI vs rest, MAD vs rest, and NC vs rest respectively.

In order to show the model learning performances in terms of loss and accuracy convergences during the training, we present in Figure 8 and Figure 9 the learning curves of a random model from Table 1 including the accuracy and the loss respectively. We observe in Figure 8 that the accuracies of attention and the classifier blocks converge after the 50th epoch without over-fitting. In parallel, in Figure 9 the total training and validation losses decrease and converge after the same epoch.

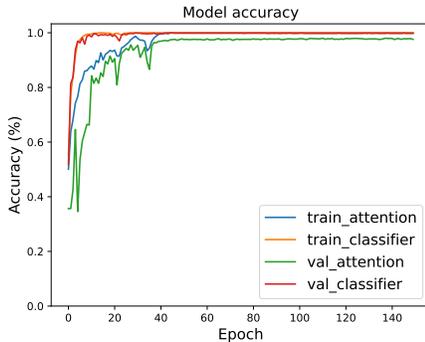


Figure 8: Learning curves (accuracy) for training and validation data for attention and classifier blocks.

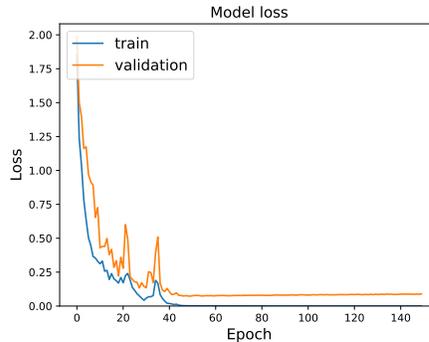


Figure 9: Learning curves (total loss) for training and validation data of the proposed framework.

4.3.3. Ablation study

a. Data generation. In order to study the performance of the proposed data generation method, we evaluate its effectiveness compared to widely used 1D data generation methods in terms of data quality generation criteria (i.e., Peak Signal-to-Noise Ratio (PSNR)) and AD stages classification performance.

We compare the proposed spectra generation method with GAN, recently used for MRS data generation for glioma detection Jang et al. (2021), Wasserstein GAN Gradient Penalty (WGAN-GP) Gulrajani et al. (2017) used for spectra augmentation Liu et al. (2019) and Conditional GAN (CGAN) Mirza & Osindero (2014). It is to note that all studied models have the same generator architectures and the same training hyper-parameters as the proposed DNN-

based data generation method. While for the discriminator, we use 03 fully connected layers of 32, 16, and 1 neurons.

Table 2 presents PSNR values by using different data generation methods and settings. The PSNR values are used as quality measurements between the original and synthesized data and computed as follows:

$$\text{PSNR}_{avg} = \mathbb{E}_{\tilde{\mathbf{x}}_i \sim \mathbb{P}_c} \frac{1}{n} \sum_{i=1}^n \text{PSNR}(\boldsymbol{\mu}_c, \tilde{\mathbf{x}}_i). \quad (8)$$

where \mathbb{P}_c is the model distribution for a class c of n spectra. \mathbb{P}_c is defined by $\tilde{\mathbf{x}} = G(\mathbf{z})$ with G is the data generator and \mathbf{z} is a random noise input. $\tilde{\mathbf{x}}_i$ is a generated spectrum, and $\boldsymbol{\mu}_c$ is the average of the original data belonging to the same class.

Also, we study the effect of synthesizing from 300-Dimensional (D)(without water peak) and 1024-Dimensional spectra (with water peak). First, Table 2 shows that the PSNR values of the data generated by the proposed method are higher than the studied SOTA data generation methods. Indeed, the GAN model gives the worst PSNR results. Yet, the class-oriented CGAN model generates noisy spectra compared to the WGAN-GP method, especially for the MAD class, where there are fewer spectra than NC and MCI classes. Moreover, we can see that our model generates clean data distribution close to the original data (row 6 from table 2) and can generalize especially for the MCI and NC class. Second, experimental results in Table 2 show that learning from 1024-Dimensional spectra (with water peak) generates better spectra quality than directly learning from 300-D spectra.

Table 2: PSNR by using different data generation methods and settings

Spectra generation method	MCI group	MAD group	NC group
GAN Jang et al. (2021)	24.01	20.33	23.44
WGAN-GP Liu et al. (2019)	29.12	23.65	28.74
Conditional GAN Mirza & Osindero (2014)	26.21	21.64	25.72
Ours (generating from 300-D spectra (no water peak))	30.57	24.11	28.78
Ours (generating from 1024-D spectra (with water peak))	30.62	23.98	29.23
PSNR values of the original spectra	29.14	25.85	28.65

Table 3, represents the model classification performance using synthetic data obtained by different aforementioned data generation techniques. The best clas-

sification accuracy for the three GANs-based methods is about 59.21% which is very low compared to the classification results when using data generated by our proposed method (rows 4 and 5). In fact, training such a class-oriented data augmentation with the few available MRS data leads to overfitting. Moreover, the discriminator block forces the model to generate data with high similarity and small diversity. Thus, testing the classifier model on the kept κ spectra per class (see section 3.2.4), which are independent from the training synthetic data, would give a random classification, especially between MCI and MAD classes.

Furthermore, classification accuracy using data generated from 1024-D spectra with water peak gives an average accuracy of 78.46%, which is better than the results obtained with the three GANs-based methods. However, as we can see the water peak which is higher than the other metabolites peaks impacts the classification results, especially between MCI and MAD classes, where we report a sensitivity of 77.50% for both classes. Finally, using 300-D data from 1024-D generated spectra, gives the best average classification accuracy (95.23%).

Table 3: Classification results using data generated with different data augmentation methods/settings.

Method	Accuracy	Sensitivity			Specificity		
		MCI vs rest	MAD vs rest	NC vs rest	MCI vs rest	MAD vs rest	NC vs rest
GAN Jang et al. (2021)	38.75% \pm 22.04%	36.00% \pm 8.94%	32.00% \pm 17.88%	84.00% \pm 16.73%	56.00% \pm 5.47%	70.00% \pm 12.24%	90.00% \pm 7.07%
WGAN+GP Gulrajani et al. (2017)	42.98% \pm 26.06%	40.00% \pm 14.14%	48.00% \pm 17.88%	70.00% \pm 10.00%	60.00% \pm 12.24%	74.00% \pm 16.73%	90.00% \pm 7.07%
CGAN Mirza & Osindero (2014)	59.21% \pm 13.18%	43.33% \pm 15.05%	53.33% \pm 20.65%	68.33% \pm 9.83%	63.33% \pm 12.11%	83.33% \pm 12.11%	90.00% \pm 8.94%
Ours (1024-D with water peak)	78.46% \pm 6.66%	77.5% \pm 12.82%	77.50% \pm 16.69%	87.50% \pm 14.88%	88.75% \pm 11.26%	88.75% \pm 9.91%	93.75% \pm 9.16%
Ours (300-D/without water peak)	95.23% \pm 5.04%	95.00% \pm 9.25%	93.33% \pm 10.00%	97.50% \pm 7.07%	95.00% \pm 5.34%	97.50% \pm 4.62%	100.00% \pm 0.00%

b Attention mechanism . In order to demonstrate the effectiveness of the learned features with and without the attention mechanism, we use the t-distributed Stochastic Neighbor Embedding (t-SNE) technique der Maaten & Hinton (2008) for features visualization. Figure 10, plots the feature t-SNE visualization of the used samples in the first column, with the last hidden fully-connected layer without and with the attention block in the second and the third columns respectively. Before using the model (i.e., the first column), the samples are randomly distributed after applying the t-SNE. This random distribution is diminished, when we use our framework as shown in the rest of the columns (i.e., second and

third columns). Moreover, when we use the attention block, the model shows its capability to discriminate the MCI samples from the MAD ones which is not the case without using it from both training and test samples. These results demonstrate that the attention block allows to the exploitation of the similarities and differences between classes via analyzing the learned feature maps from the extractor block. Thus, adapting the network to better generalize and thereby improves the robustness of the model. Therefore, eliminating this block causes instability of learning and generalization.

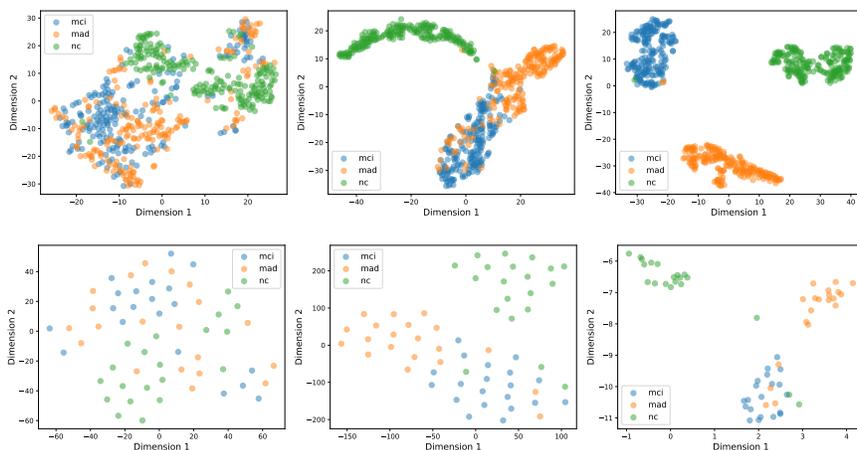


Figure 10: Feature visualization via t-SNE (t-distributed stochastic neighbor embedding) of synthetic training data and Poitiers university hospital data in the first and second row respectively. Column1: shows the representations of original data. Column2: shows the feature representations without attention mechanism. Column3: shows the feature representations using the attention mechanism.

In addition to its role in guiding the network for learning and refining features. The attention mechanism allows to interpret and validate the model with domain knowledge the model with domain knowledge for AD diagnosis using MRS data. Hence, we generate an attention map that highlights the relevant metabolites for classes discrimination from ^1H -MRS data. Such representation would help radiologists to understand the model predictions and provide corrections if necessary. Figure 11 represents attention maps (heat-maps) overlaid

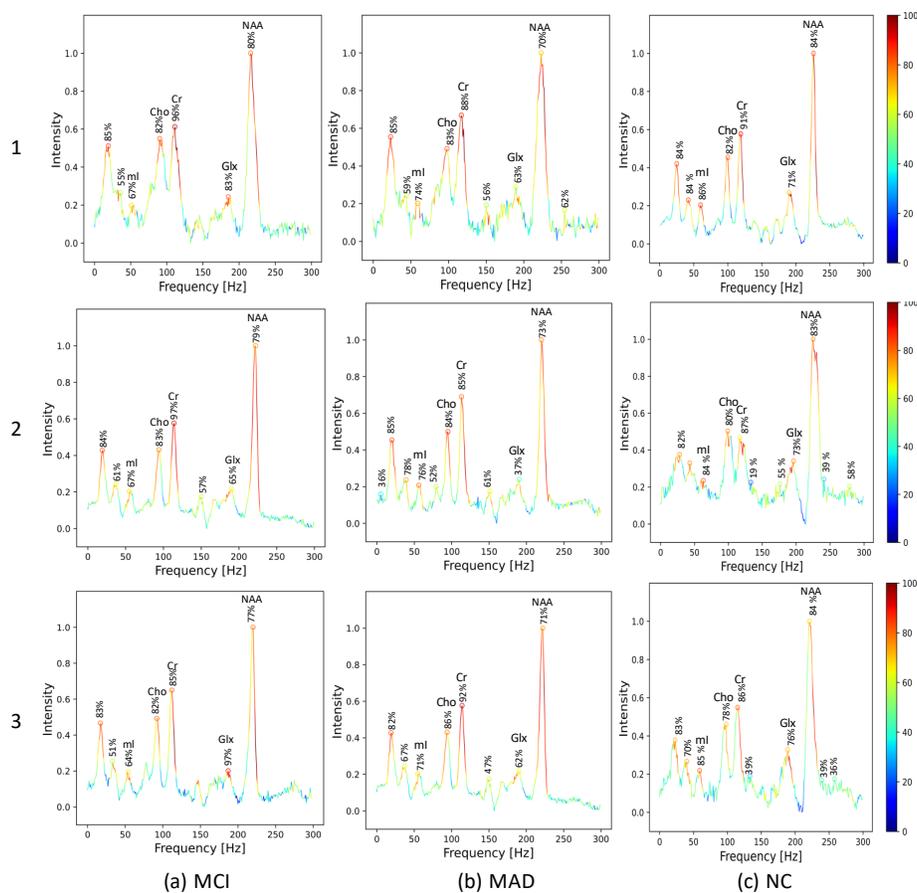


Figure 11: Heat-maps visualization for each class from 03 testing spectra samples (real part).

on 03 random spectrum of a MCI, MAD and NC subjects respectively. The importance of the metabolites grows as the colors vary from blue to brown color. Also, we report the percentage importance of each peak having an intensity above 0.15, in order to include all the relevant metabolites such as N-acetyl-aspartate+N-acetyl-aspartyl-glutamate (NAA), Creatine/phosphocreatine (Cr), Choline (Cho), and myo-Inositol (mI). Indeed, the sigmoid activation function is used to build an attention map with values ranging between 0 and 1. Therefore, the percentage importance of each peak is illustrated by overlaying the position of each peak on the corresponding value of the attention map at this position.

All these metabolites are situated around the frequencies $230Hz$, $120Hz$, $90Hz$ and $50Hz$. As shown in Li et al. (2003), the level of Cr is considered relatively stable during the degenerative pathology. The Cr is used as a reference value for the clinical diagnostic criteria computation namely the NAA/Cr, Cho/Cr, and mI/Cr ratios Kantarci & Jicha (2019). In Figure 11, we present the attention map for 03 examples per class. As we can see, the most important metabolite is the Cr for all the classes, i.e., MCI, MAD, and NC (e.g., a percentage of 96%, 88%, and 91% respectively for the first row), which demonstrate the stability of this metabolite. In the same figures, the percentages of neuronal integrity markers NAA (allows for an accurate assessment of neuronal functional loss) and NAA/Cr decrease moving from NC to MAD groups (eg, an NAA percentage of 84%, 80%, and 70% respectively for the first row), which is in conformity with Salem et al. (2008); Adalsteinsson et al. (2000). More additionally, the ratio Cho/Cr is higher in MAD classes compared to the MCI classes with a ratio of $83/88 = 0.94$ and $82/96 = 0.85$ respectively for the first row. Moreover, the mI/Cr ratio also increases moving from MCI to MAD classes with a ratio of $67/96 = 0.67$ and $74/88 = 0.84$ respectively for the first row. These results are in concordance with SOTA bio-medical works on early AD detection Li et al. (2003); Kantarci et al. (2007); Schott et al. (2010). Furthermore, we note that the importance order of the metabolites is unchanged for all testing spectra with respect to a specific class. For example, for the class NC, the order is as follows: Cr, mI, NAA, and Cho is in the last position. The same conclusion is for the MCI and MAD classes. Thus, the proposed classification model uses prominent metabolites in the MRS data that are commonly used in clinical practice but also considers metabolites such as glutamate and glutamine (Glx) peaks. These metabolites have been recently investigated for neural brain damage detection Zeydan & Kantarci (2021); Oeltzschner et al. (2019), pointing out new directions for future clinical research on early AD detection.

4.3.4. Comparison with SOTA

In order to evaluate the performance of the proposed framework against SOTA works. First, we compare our approach to Ahmed et al. (2020) which is the only work that has tackled the detection of AD vs NC using ^1H -MRS data using an end to end 1D-CNN approach. Indeed, we use their proposed model to perform multi-class classification using our ^1H -MRS dataset. However, we use three instead of two fully-connected neurons since we have three categories.

Table 4: Evaluation of our proposed approach with different settings compared to (a) DeepMRS method Ahmed et al. (2020) (b) Hipp (MRI) Li & Liu (2019). (c) Whole brain + meta-data Dufumier et al. (2021). (d) our proposed approach.

Method	Accuracy	Sensitivity			Specificity		
		MCI vs rest	MAD vs rest	NC vs rest	MCI vs rest	MAD vs rest	NC vs rest
DeepMRS Ahmed et al. (2020)	67.5% \pm 9.72%	67.5% \pm 26.05%	42.5% \pm 19.82%	92.5% \pm 10.35%	71.25% \pm 12.46%	86.25% \pm 10.60%	85.00% \pm 7.56%
Hipp (MRI) Li & Liu (2019)	61.73% \pm 6.06%	72.00% \pm 30.33%	12.00% \pm 17.88%	100.00% \pm 0.00%	56.00% \pm 8.94%	88.00% \pm 16.43%	98.00% \pm 4.47%
Whole brain + meta-data Dufumier et al. (2021)	78.66% \pm 5.57%	61.00% \pm 8.94%	72.00% \pm 17.88%	100.00% \pm 0.00%	86.00% \pm 8.94%	90.00% \pm 10.00%	96.00% \pm 8.94%
Ours	95.23% \pm 5.04%	95.00% \pm 9.25%	93.33% \pm 10.00%	97.50% \pm 7.07%	95.00% \pm 5.34%	97.50% \pm 4.62%	100.00% \pm 0.00%

Second, we evaluate the proposed MRS-based CAD approach for early AD against recent imaging-based approaches. Hence, we use the corresponding T1-weighted MRI of our MRS data for approaches evaluation and we perform multi-classification by updating the last model layer. For a fair comparison, we follow the same experimental protocol and training parameters

We start by comparing the obtained classification results with those reported in Li & Liu (2019). The latter used the hippocampus region as imaging-based bio-markers to learn relevant features for AD detection in deep learning framework. Then, we perform comparison with a recently proposed approach Dufumier et al. (2021) based on contrastive learning. The authors investigated patients meta-data and the whole brain structure as bio-imaging bio-markers for AD detection.

Table 4 reports the comparison results. We can see that our method outperforms the DeepMRS model in all the proposed metrics, especially in destining MCI samples from MAD ones, which demonstrates the robustness of our framework and the efficiency of the attention mechanism to distinct AD early stage subjects. Moreover, contrary to Ahmed et al. (2020) in which the authors

quantify the contribution of each metabolite separately, we highlight the importance of clinical features in terms of metabolite ratios (NAA/Cr, Cho/Cr and mI/Cr) which correspond to the diagnostic criteria adapted in clinical practice.

Furthermore, compared to the structural approaches, our model outperforms both frameworks proposed by Li & Liu (2019) and Dufumier et al. (2021) by respectively 27% and 14% difference in term of classification accuracy. This is due to the highly challenging issue, where it produced a random prediction between MCI and MAD classes, which is not the case for our method. These obtained results proved the effectiveness of using MRS metabolites with DL for early AD detection compared to structural MRI.

5. Conclusion

In this paper, we proposed a multi-class classification framework for early AD detection from ^1H -MRS data. The proposed classification is explicitly guided by an attention mechanism that helps the classifier to focus on diagnostically relevant metabolites for classes discrimination. Synthetic data are used to tackle the problem of lack of data and to help in learning the features space. The obtained results show the efficiency of the approach under extensive metrics. Our framework outperforms the SOTA imaging bio-markers-based approach which demonstrates the robustness of selectively learning metabolites features against traditional imaging bio-markers for early AD detection. Using ^1H -MRS data with 1D-CNN opens a promising direction to create a new virtual biopsy for early AD detection. For future works, we will investigate additional medical imaging modalities (PET, FLAIR, and DTI) for proposed method evaluation. Yet, we plan to use multi-modal MRS including ^1H -MRS, MRS of GABA, Phosphorus MRS (^{31}P -MRS) and high magnetic field imaging such as 7 Tesla MRI.

References

Adalsteinsson, E., Sullivan, E., Kleinhans, N., Spielman, D., & Pfefferbaum, A. (2000). Longitudinal decline of the neuronal marker n-acetyl aspartate in

- alzheimer’s disease. *The Lancet*, 355, 1696–1697.
- Aderghal, K., Benois-Pineau, J., & Afdel, K. (2017). Classification of smri for alzheimer’s disease diagnosis with cnn: single siamese networks with 2d+? approach and fusion on adni. In Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval (pp. 494–498).
- Ahmed, O. B., Fezzani, S., Guillevin, C., Fezai, L., Naudin, M., Gianelli, B., & Fernandez-Maloigne, C. (2020). Deepmrs: An end-to-end deep neural network for dementia disease detection using mrs data. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)* (pp. 1459–1463). IEEE.
- Andreasen, N., Minthon, L., Davidsson, P., Vanmechelen, E., Vanderstichele, H., Winblad, B., & Blennow, K. (2001). Evaluation of csf-tau and csf-a β 42 as diagnostic markers for alzheimer disease in clinical practice. *Archives of neurology*, 58, 373–379.
- Bateman, R. J., Xiong, C., Benzinger, T. L. S., Fagan, A. M., Goate, A., Fox, N. C., Marcus, D. S., Cairns, N. J., Xie, X., & Blazey, T. M. (2012). et al., “clinical and biomarker changes in dominantly inherited alzheimer’s disease,”. *N Engl J Med*, 367, 795–804.
- Beason-Held, L. L., Goh, J. O., An, Y., Kraut, M. A., O’Brien, R. J., Ferrucci, L., & Resnick, S. M. (2013). Changes in brain function occur years before the onset of cognitive impairment. *Journal of Neuroscience*, 33, 18008–18014.
- Ben Ahmed, O., Mizotin, M., Benois-Pineau, J., Allard, M., Catheline, G., & Ben Amar, C. (2015). Alzheimer’s disease diagnosis on structural mr images using circular harmonic functions descriptors on hippocampus and posterior cingulate cortex. *Computerized Medical Imaging and Graphics*, 44, 13–25.
- Bhatt, C., Kumar, I., Vijayakumar, V., Singh, K. U., & Kumar, A. (2021). The state of the art of deep learning models in medical science and their challenges. *Multimedia Systems*, 27, 599–613.

- Bourdon, P., Ahmed, O. B., Urruty, T., Djemal, K., & Fernandez-Maloigne, C. (2021). Explainable ai for medical imaging: Knowledge matters. In *Multi-faceted Deep Learning* (pp. 267–292).
- Chen, D., Wang, Z., Guo, D., Orekhov, V., & Qu, X. (2020). Review and prospect: deep learning in nuclear magnetic resonance spectroscopy. *arXiv preprint arXiv:2001.04813*, .
- Dandil, E., & Karaca, S. (2021). Detection of pseudo brain tumors via stacked lstm neural networks using mr spectroscopy signals. *Biocybernetics and Biomedical Engineering*, *41*, 173–195.
- Dufumier, B., Gori, P., Victor, J., Grigis, A., Wessa, M., Brambilla, P., Favre, P., Polosan, M., Mcdonald, C., Piguët, C. M. et al. (2021). Contrastive learning with continuous proxy meta-data for 3d mri classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 58–68). Springer.
- Ebrahimi-Ghahnavieh, A., Luo, S., & Chiong, R. (2019). Transfer learning for alzheimer’s disease detection on mri images. In *2019 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)* (pp. 133–138). IEEE.
- Feng, C., Elazab, A., Yang, P., Wang, T., Zhou, F., Hu, H., Xiao, X., & Lei, B. (2019). Deep learning framework for alzheimer’s disease diagnosis via 3d-cnn and fsbi-lstm. *IEEE Access*, *7*, 63605–63618.
- Gallaway, P. J., Miyake, H., Buchowski, M. S., Shimada, M., Yoshitake, Y., Kim, A. S., & Hongu, N. (2017). Physical activity: a viable way to reduce the risks of mild cognitive impairment, alzheimer’s disease, and vascular dementia in older adults. *Brain sciences*, *7*, 2.
- Graff-Radford, J., & Kantarci, K. (2013). Magnetic resonance spectroscopy in alzheimer’s disease. *Neuropsychiatric disease and treatment*, *9*.

- Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. C. (2017). Improved training of wasserstein gans. *CoRR*, *abs/1704.00028*. [arXiv:1704.00028](#).
- Honig, L. S., Vellas, B., Woodward, M., Boada, M., Bullock, R., Borrie, M., Hager, K., Andreasen, N., Scarpini, E., & Liu-Seifert, H. (2018). et al, “trial of solanezumab for mild dementia due to alzheimer’s disease,”. *New England Journal of Medicine*, *378*, 321–330.
- Hosseini-Asl, E., Gimel’farb, G., & El-Baz, A. (2016). Alzheimer’s disease diagnostics by a deeply supervised adaptable 3d convolutional network. *arXiv preprint arXiv:1607.00556*, .
- Hu, W., Chen, D., Qiu, T., Chen, H., Chen, X., Yang, L., Yan, G., Guo, D., & Qu, X. (2021). Denoising single voxel magnetic resonance spectroscopy with deep learning on repeatedly sampled in vivo data. *arXiv preprint arXiv:2101.11442*, .
- Huang, C., Wahlund, L.-O., Svensson, L., Winblad, B., & Julin, P. (2002). Cingulate cortex hypoperfusion predicts alzheimer’s disease in mild cognitive impairment. *BMC neurology*, *2*, 1–6.
- Jang, J., Lee, H. H., Park, J.-A., & Kim, H. (2021). Unsupervised anomaly detection using generative adversarial networks in 1h-mrs of the brain. *Journal of Magnetic Resonance*, *325*.
- Javid, M. A., & Buzdar, S. A. (2021). et al., “a novel computer aided diagnostic system for quantification of metabolites in brain cancer,”. *Biomedical Signal Processing and Control*, *66*.
- Kantarci, K., & Jicha, G. A. (2019). Development of 1h mrs biomarkers for tracking early predementia alzheimer disease.
- Kantarci, K., Weigand, S. D., Petersen, R. C., Boeve, B. F., Knopman, D. S., Gunter, J., Reyes, D., Shiung, M., O’Brien, P. C., & Smith, G. E. (2007). et

- al., “longitudinal 1h mrs changes in mild cognitive impairment and alzheimer’s disease,”. *Neurobiology of aging*, 28, 1330–1339.
- Khatri, U., & Kwon, G.-R. (2020). An efficient combination among smri, csf, cognitive score, and apoe ϵ 4 biomarkers for classification of ad and mci using extreme learning machine. *Computational intelligence and neuroscience*, 2020.
- Knopman, D. S., & Petersen, R. C. (2014). Mild cognitive impairment and mild dementia: a clinical perspective. In *Mayo Clinic Proceedings* (pp. 1452–1459). vol. 89: Elsevier.
- 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.
- Lee, H. H., & Kim, H. (2019). Intact metabolite spectrum mining by deep learning in proton magnetic resonance spectroscopy of the brain. *Magnetic resonance in medicine*, 82, 33–48.
- Lee, H. H., & Kim, H. (2020). Deep learning-based target metabolite isolation and big data-driven measurement uncertainty estimation in proton magnetic resonance spectroscopy of the brain. *Magnetic resonance in medicine*, 84, 1689–1706.
- Lee, P.-L., Chou, K.-H., Chung, C.-P., Lai, T.-H., Zhou, J. H., Wang, P.-N., & Lin, C.-P. (2020). Posterior cingulate cortex network predicts alzheimer’s disease progression. *Frontiers in aging neuroscience*, 12.
- Leech, R., & Sharp, D. J. (2014). The role of the posterior cingulate cortex in cognition and disease. *Brain*, 137, 12–32.
- Li, B. S. Y., Wang, H., & Gonen, O. (2003). Metabolite ratios to assumed stable creatine level may confound the quantification of proton brain mr spectroscopy. *Magnetic resonance imaging*, 21, 923–928.

- Li, F., & Liu, M. (2019). Alzheimer’s disease neuroimaging initiative, et al., “a hybrid convolutional and recurrent neural network for hippocampus analysis in alzheimer’s disease,”. *Journal of neuroscience methods*, *323*, 108–118.
- Liu, M., Cheng, D., & Yan, W. (2018). Alzheimer’s disease neuroimaging initiative, et al., “classification of alzheimer’s disease by combination of convolutional and recurrent neural networks using fdg-pet images,”. *Frontiers in neuroinformatics*, *12*.
- Liu, Y., Zhou, Y., Liu, X., Dong, F., Wang, C., & Wang, Z. (2019). Wasserstein gan-based small-sample augmentation for new-generation artificial intelligence: a case study of cancer-staging data in biology. *Engineering*, *5*, 156–163.
- López, C., Sánchez, J. L., & Martín, J. (2020). The effect of cognitive stimulation on the progression of cognitive impairment in subjects with alzheimer’s disease. *Applied Neuropsychology: Adult*, (pp. 1–10).
- Lu, D., Kurz, G., Polomac, N., Gacheva, I., Hattingen, E., & Triesch, J. (2021a). Multiple instance learning for brain tumor detection from magnetic resonance spectroscopy data. [arXiv:2112.08845](https://arxiv.org/abs/2112.08845).
- Lu, D., Polomac, N., Gacheva, I., Hattingen, E., & Triesch, J. (2021b). Human-expert-level brain tumor detection using deep learning with data distillation and augmentation. In *ICASSP IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 2021–2021). IEEE pp. 3975–3979.
- der Maaten, L. V., & Hinton, G. (2008). Visualizing data using t-sne. *Journal of machine learning research*, *9*, 11.
- Martinez-Murcia, F. J., Ortiz, A., Gorriz, J.-M., Ramirez, J., & Castillo-Barnes, D. (2019). Studying the manifold structure of alzheimer’s disease: A deep learning approach using convolutional autoencoders. *IEEE Journal of Biomedical and Health Informatics*, *24*, 17–26.

- Maul, S., Giegling, I., & Rujescu, D. (2020). Proton magnetic resonance spectroscopy in common dementias—current status and perspectives. *Frontiers in Psychiatry, 11*.
- Mayeda, E. R., Glymour, M. M., Quesenberry, C. P., & Whitmer, R. A. (2016). Inequalities in dementia incidence between six racial and ethnic groups over 14 years. *Alzheimer's & Dementia, 12*, 216–224.
- Menéndez-González, M. (2014). Routine lumbar puncture for the early diagnosis of alzheimer's disease. is it safe? *Frontiers in aging neuroscience, 6*.
- Menze, B., Isensee, F., Wiest, R., Wiestler, B., Maier-Hein, K., Reyes, M., & Bakas, S. (2021). Analyzing magnetic resonance imaging data from glioma patients using deep learning. *Computerized medical imaging and graphics, 88*.
- Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. *CoRR, abs/1411.1784*. [arXiv:1411.1784](https://arxiv.org/abs/1411.1784).
- Morley, J. E., Farr, S. A., & Nguyen, A. D. (2018). Alzheimer disease. *Clinics in geriatric medicine, 34*, 591–601.
- Near, J., Harris, A. D., Juchem, C., Kreis, R., Marjańska, M., Öz, G., Slotboom, J., Wilson, M., & Gasparovic, C. (2021). Preprocessing, analysis and quantification in single-voxel magnetic resonance spectroscopy: experts' consensus recommendations. *NMR in Biomedicine, 34*, e4257.
- Oeltzschner, G., Wijtenburg, S. A., Mikkelsen, M., Edden, R. A., Barker, P. B., Joo, J. H., Leoutsakos, J.-M. S., Rowland, L. M., Workman, C. I., & Smith, G. S. (2019). Neurometabolites and associations with cognitive deficits in mild cognitive impairment: a magnetic resonance spectroscopy study at 7 tesla. *Neurobiology of aging, 73*, 211–218.
- Öz, G., Deelchand, D. K., Wijnen, J. P., Mlynárik, V., Xin, L., Mekle, R., Noeske, R., Scheenen, T. W., Tkáč, I., on Advanced Single Voxel 1H MRS,

- E. W. G. et al. (2021). Advanced single voxel 1h magnetic resonance spectroscopy techniques in humans: Experts' consensus recommendations. *NMR in Biomedicine*, *34*, e4236.
- Salem, D. B., Walker, P. M., Bejot, Y., Aho, S. L., Tavernier, B., Rouaud, O., Ricolfi, F., & Brunotte, F. (2008). N-acetylaspartate/creatine and choline/creatine ratios in the thalami, insular cortex and white matter as markers of hypertension and cognitive impairment in the elderly. *Hypertension Research*, *31*, 1851–1857.
- Schlemper, J., Oktay, O., Schaap, M., Heinrich, M., Kainz, B., Glocker, B., & Rueckert, D. (2019). Attention gated networks: Learning to leverage salient regions in medical images. *Medical image analysis*, *53*, 197–207.
- Schott, J. M., Frost, C., MacManus, D. G., Ibrahim, F., Waldman, A. D., & Fox, N. C. (2010). Short echo time proton magnetic resonance spectroscopy in alzheimer's disease: a longitudinal multiple time point study. *Brain*, *133*, 3315–3322.
- Suk, H.-I., Lee, S.-W., & Shen, D. (2015). Latent feature representation with stacked auto-encoder for ad/mci diagnosis. *Brain Structure and Function*, *220*, 841–859.
- Young, P. N. E., Estarellas, M., Coomans, E., Srikrishna, M., Beaumont, H., Maass, A., Venkataraman, A. V., Lissaman, R., Jiménez, D., & Betts, M. J. (2020). et al., “imaging biomarkers in neurodegeneration: current and future practices,”. *Alzheimer's research & therapy*, *12*, 1–17.
- Zeydan, B., & Kantarci, K. (2021). Decreased glutamine and glutamate: an early biomarker of neurodegeneration. *International psychogeriatrics*, *33*, 1–2.
- Zhang, X., Han, L., Zhu, W., Sun, L., & Zhang, D. (2021). An explainable 3d residual self-attention deep neural network for joint atrophy localization and

alzheimer's disease diagnosis using structural mri. *IEEE Journal of Biomedical and Health Informatics*, (pp. 1–1).

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning deep features for discriminative localization. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2921–2929).