



The impact of the orientation of MRI slices on the accuracy of Alzheimer's disease classification using convolutional neural networks (CNNs)

Bruno A. C. Ramalho^{1^}, Lara R. Bortolato², Naomy D. Gomes^{3^}, Lauro Wichert-Ana^{4^},
Fernando Eduardo Padovan-Neto^{5^}, Marco Antonio A. da Silva^{6^}, Kleython José C. C. de Lacerda^{1,4^}

¹Department of Physics, Faculty of Philosophy, Sciences and Letters of Ribeirão Preto, University of São Paulo, Ribeirão Preto, Brazil; ²Departament of Pharmaceutical Sciences, Faculty of Pharmaceutical Sciences of Ribeirão Preto, University of São Paulo, Ribeirão Preto, Brazil; ³São Carlos Institute of Physics, University of São Paulo, São Carlos, Brazil; ⁴Nuclear Medicine & Molecular Imaging Section, Department of Medical Imaging, Hematology and Clinical Oncology, Hospital das Clínicas, School of Medicine, University of São Paulo, Ribeirão Preto, Brazil; ⁵Department of Psychology, Faculty of Philosophy, Sciences and Letters of Ribeirão Preto, University of São Paulo, Ribeirão Preto, Brazil; ⁶Department of Biomolecular Sciences, Faculty of Pharmaceutical Sciences of Ribeirão Preto, University of São Paulo, Ribeirão Preto, Brazil

Contributions: (I) Conception and design: KJCC de Lacerda, BAC Ramalho, MAA da Silva; (II) Administrative support: KJCC de Lacerda; (III) Provision of study materials or patients: KJCC de Lacerda; (IV) Collection and assembly of data: KJCC de Lacerda, BAC Ramalho; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Kleython José C. C. de Lacerda, BSC, MBA, DSC. Postdoctoral Fellow in Nuclear Medicine & Molecular Imaging Section, Department of Medical Imaging, Hematology and Clinical Oncology, Hospital das Clínicas, School of Medicine, University of São Paulo, Av. Bandeirantes 3900, Monte Alegre, Ribeirão Preto, SP 14049-900, Brazil; Department of Physics, Faculty of Philosophy, Sciences and Letters of Ribeirão Preto, University of São Paulo, Ribeirão Preto, Brazil. Email: kleython_lacerda@usp.br.

Background: Alzheimer's disease (AD) is the leading cause of major neurocognitive disorders, affecting approximately 50 million people worldwide. Due to its high prevalence, AD significantly impacts patients' quality of life and poses a substantial challenge to healthcare systems. Diagnosis is intricate, with specificity and sensitivity rates falling below the ideal. Early identification of AD is essential to increase the effectiveness of pharmacotherapeutic treatment and improve quality of life. Consequently, there is a quest for innovative methods, such as machine learning and deep learning, to automate the diagnosis of AD in its early stages.

Methods: We developed and validated a convolutional neural network (CNN) algorithm using the Keras Sequential API in Python to investigate the impact of slicing T1-weighted magnetic resonance images on the classification of patients with mild cognitive impairment (MCI) and healthy patients (NC), grouped based on scores on the Mini-Mental State Examination (MMSE). We selected 318 patients (250 healthy and 68 MCI) with a minimum of 16 years of education (equivalent to a completed undergraduate degree). The training, testing, and validation datasets were split in a 70/15/15 ratio for each slice.

Results: The CNN achieved high accuracy values in classifying healthy and MCI groups, ranging between 97% and 99% depending on the slice, the number of training epochs, and batch size. In addition to precision, the F1-score, recall, and precision parameters were also evaluated, with values above 91%. Generally, the coronal slice produced the best results, followed by the axial and the sagittal slices, which nevertheless showed high performance, standing out individually in different evaluation parameters. Notably, the choice of batch size and the number of epochs also influenced the network's classification.

Conclusions: Our study findings indicate that utilizing CNN in conjunction with selecting a coronal slice proves to be a promising tool for facilitating the early-stage diagnosis of neurodegenerative diseases, such

[^] ORCID: Bruno A. C. Ramalho, 0009-0007-7110-9395; Naomy D. Gomes, 0000-0002-0409-461X; Lauro Wichert-Ana, 0000-0003-2059-5120; Fernando Eduardo Padovan-Neto, 0000-0002-5529-3168; Marco Antonio A. da Silva, 0000-0002-6720-5073; Kleython José C. C. de Lacerda, 0000-0002-9542-6885.

as AD, through magnetic resonance imaging analysis, enabling more effective treatments and appropriate future planning. Moving forward, we aim to investigate whether these results replicate across other imaging modalities, such as positron emission tomography, and explore additional datasets.

Keywords: Alzheimer's disease (AD); convolutional neural network (CNN); magnetic resonance imaging slice orientation (MRI slice orientation); early diagnosis

Received: 21 February 2024; Accepted: 20 May 2024; Published online: 28 June 2024.

doi: 10.21037/jmai-24-51

View this article at: <https://dx.doi.org/10.21037/jmai-24-51>

Introduction

Background

The aging of the population has led to an increase in the prevalence of chronic diseases, particularly among those aged 65 years and older. One class of diseases that is particularly prominent is neurodegenerative diseases, especially Alzheimer's disease (AD) (1-3), a progressive neurodegenerative disorder that results in progressive neuronal death. Currently, AD is identified as the leading cause of major neurocognitive disorder (MNCDD), or

dementia, worldwide (4), accounting for 50–70% of all cases (5). AD symptoms typically begin with memory issues, primarily affecting recent memory. As the disease progresses, patients start experiencing more pronounced cognitive difficulties, such as problems with speech, comprehension, and decision-making, ultimately leading to the inability to perform basic daily activities (6).

It is estimated that up to 75% of patients affected by dementia have not received a proper diagnosis (7). This fact underscores the need to develop a highly accurate methodology to diagnose these conditions quicker and more precisely, capable of distinguishing similar presentations. The pharmacological approach to AD demonstrates greater efficacy when implemented in the early stages of the disease, where patients exhibit only mild cognitive impairment (MCI). Such interventions significantly contribute to delaying disease progression and mitigating the emergence of more severe symptoms. Given this, and the aim to improve the patient's quality of life, developing a tool that assists in early-stage AD diagnosis is imperative today (8-12).

The diagnosis of AD is complex. It necessitates a highly qualified neurologist who must comprehend both the patient's clinical and family history, conduct appropriate neuropsychological tests such as the Mini-Mental State Examination (MMSE), and interpret imaging tests like magnetic resonance imaging (MRI), which can reveal characteristic disease-related findings (13,14). Due to its reliance on multiple interpretative factors, the diagnostic methodology contends with wide sensitivity intervals and a specificity typically below an ideal value (15). Clinical and pathological studies suggest that physicians have a probability ranging from 70.9% to 87.3% accuracy in diagnosing AD (15). Unfortunately, professionals sufficiently qualified to precisely diagnose AD in its various progression stages are scarce.

Given the imprecision in AD diagnosis, the use of advanced machine learning and deep learning tools emerge as promising alternatives for achieving highly

Highlight box

Key findings

- The use of convolutional neural networks (CNNs) has significant potential to aid in the early diagnosis of neurodegenerative diseases such as Alzheimer's disease.
- Our study explored how artificial intelligence (AI) detects mild cognitive impairment (MCI) in magnetic resonance imaging (MRI) scans with varying slice orientations. Surprisingly, we found that AI could detect MCI more accurately based on slice orientation, suggesting its crucial role in improving AI-based disease detection.

What is known and what is new?

- CNNs are currently utilized in Alzheimer's diagnosis, but a major limitation is the number of intermediate layers, leading to long processing times and hardware demands.
- Our approach aims to optimize efficiency and accuracy by analyzing CNN metrics using the Alzheimer's Disease Neuroimaging Initiative (ADNI) 1 dataset to assess patients with a level of education equal to or greater than a graduate degree and understanding how MRI slice orientation influences network classification.

What is the implication, and what should change now?

- The synergistic use of CNNs, particularly in conjunction with coronal slice selection, presents a promising and effective tool for early neurodegenerative disease diagnosis, such as Alzheimer's, through T1-weighted MRI analysis.

precise classification regarding the disease's development and stage using image data generated by various types of examinations, such as MRI and positron emission tomography (PET)-scan (16-19). Convolutional neural networks (CNNs) represent an artificial neural network (ANN) architecture specialized in extracting complex features from images, thereby detecting patterns related to disease progression. For instance, they can identify degenerative changes in brain morphology (particularly in the cortex) caused by neuronal loss in AD (20-24). A CNN consists of layers that function like filters, traversing the entire image to extract relevant characteristics.

CNNs require image data for training to learn to characterize specific factors and generate classification. Hence, using data from open-source repositories like the Alzheimer's Disease Neuroimaging Initiative (ADNI) proves advantageous in developing this category of networks. ADNI is an initiative that aims to collect and provide quantitative clinical data and images of patients with AD and healthy individuals for researchers worldwide. These data contribute to advancing the understanding of the disease and the pursuit of treatments and diagnostic methodologies (25,26).

Related studies

Accurate diagnosis in the early stages of AD is crucial for initiating more effective treatment (27-30). Consequently, there is a continuous search for diagnostic strategies that employ machine learning tools, such as CNNs (17,18,31,32). In the study conducted by Basaia *et al.* (17), a total of 1,409 individuals were monitored, which included 294 probable AD patients, 763 patients with MCI, and 352 normal controls. An additional 229 individuals (comprising 124 probable AD patients, 50 patients with MCI, and 55 healthy controls) were included in an independent database. The study utilized 3D T1-weighted MRI images and employed a CNN architecture consisting of 12 repeated blocks of convolutional layers (2 blocks with 50 kernels of size 5×5×5 alternating steps 1 and 2 and 10 blocks with 100 to 1,600 kernels of size 3×3×3 alternating steps). The activation function used was ReLU (rectified linear unit). The network also included a fully connected layer and an output layer (logistic regression). Notably, this work replaced Max-Pooling layers with standard convolutional layers with step 2 (a fully convolutional network). The network successfully discriminates patients with MCI from those without, achieving accuracy,

sensitivity, and specificity values exceeding 86%.

The study by Alsaed and Omar in 2022 (33) used CNNs in conjunction with different traditional classifiers used in machine learning techniques: support vector machine (SVM), random forest (RF), and Softmax. All these methods were done while still using MRI images from an open database. The dataset from ADNI included 741 individuals, comprising 314 with AD and 427 healthy individuals. The study also incorporated images from the MIRIAD database, obtained from 46 AD patients and 23 normal controls. The researchers employed data augmentation techniques to increase the dataset size by making minor modifications to the original images (e.g., rotation and mirroring). The approach utilized a CNN architecture named ResNet-50, which consists of five stages of Conv blocks, pooling layers, and a fully connected layer. Both datasets were divided into 60% for training, 20% for validation, and 20% for testing. The results indicate that, with the ADNI dataset, ResNet-50 with Softmax achieved 99% accuracy, ResNet-50 with SVM obtained 92%, and ResNet-50 with Random Forest achieved 85.7%. For the MIRIAD dataset, the accuracy was 96% for ResNet-50 with Softmax, 90% for ResNet-50 with SVM, and 84.4% for ResNet-50 with Random Forest. It's important to note that the accuracy values obtained in this study refer to the classification of healthy individuals versus patients with more advanced stages of AD.

Other research groups have conducted studies (18,34,35) utilizing another commonly used imaging technique for AD diagnosis, which is positron emission tomography-computed tomography (PET-CT) using the radiopharmaceutical ¹⁸F-fluorodeoxyglucose (¹⁸F-FDG). The study (18) employed combinations of 2D CNNs and recurrent neural networks (RNNs) for AD classification. This developed architecture learns intra- and inter-slice aspects for classification after decomposing the 3D PET image into a sequence of 2D slices obtained from 339 individuals, including 93 AD patients, 146 patients with MCI, and 100 normal controls. The results of this study showed an accuracy of 78.9% for classifying patients with MCI and healthy individuals, which was the best result. In the future, with further implementations and enhancements, our CNN could be employed in the task of AD classification, utilizing images obtained from other types of examinations besides MRI.

Objective

This study's primary objective is to comprehensively

Table 1 Demographic characteristics of studied subjects from ADNI database

Classification	n	Gender (female/male)	Age [†] , years	MMSE [†]	Education [†]
MCI	68	20/48	75.9±0.9	22.9±0.1	17.3±0.2
NC	250	84/166	75.6±0.4	28.1±0.1	17.6±0.1

[†], data are presented as mean ± standard error. NC, normal control; MCI, mild-cognitive impairment; MMSE, Mini-Mental State Examination.

analyze how the choice of orientation in anatomical slices on magnetic resonance images influences the accuracy of AD classification performed through CNNs created with the Sequential Keras API in Python. In addition, the study considers patient screening from the ADNI database based on individuals' educational levels and scores obtained in the MMSE. The underlying hypothesis posits that these factors, particularly the orientation of anatomical slices, could significantly impact the accuracy of AD classification by the network. Therefore, this study represents an endeavor to deepen the understanding of how these elements interact and contribute to the efficacy of CNNs in accurately classifying AD.

Methods

Dataset, ADNI data collection, and participants selection

All data utilized in this study were obtained from the ADNI. The ADNI is a public-private partnership database originating from longitudinal multicenter studies to develop clinical, imaging, genetic, and biochemical biomarkers for the early detection and tracking of AD. Presently, it serves as a significant information source for research in the field of AD and other neurodegenerative diseases (25,26).

The images utilized in this study were selected from the ADNI 1 databases. This initial phase commenced in October 2004 and lasted for 5 years. The T1-weighted MRI images were taken 6 months after the patients began their study follow-up. The patient selection criterion was a minimum study duration of 16 years, corresponding to a complete graduation. After this initial selection, patients were categorized into two groups based on their scores on the MMSE: healthy [normal control (NC)] (scores 25–30) and MCI (scores 21–24) (36). More information about the groups is presented in *Table 1*.

MRI acquisition protocol

Our study utilized T1-weighted MRI scans featuring axial,

coronal, and sagittal slices acquired with both GE Medical Systems and SIEMENS Symphony scanners at a magnetic field strength of 1.5 Tesla. The data originated from the ADNI 1 database.

Detailed information about acquiring the MRI images used from the ADNI can be found on the official study page. The imaging acquisition protocol provides specifications such as the acquisition plane, type, magnetic field strength used, image matrix, pixel spacing, pulse sequence, slice thickness, and weighting, among other details. An example of the acquisition of a sagittal slice available in ADNI is: “acquisition plane = SAGITTAL; acquisition type = 3D; coil = 8HRBRAIN; field strength = 1.5 Tesla; flip angle = 8.0 degree; manufacturer = GE medical systems; matrix X = 256.0 pixels; matrix Y =256.0 pixels; matrix Z =166.0; Mfg model = SIGNA EXCITE; pixel spacing X =0.9 mm; pixel spacing Y =0.9 mm; pulse sequence = RM; slice thickness =1.2 mm; echo time (TE) =4.0 ms; inversion time (TI) =1,000.0 ms; repetition time (TR) =9.1 ms; weighting = T1”.

MRI analysis and data separation

During the analysis stage of the MRI images, modifying the original images supplied by ADNI was necessary. The image files from ADNI are primarily DICOM files captured in a sagittal plane, adhering to the standard image acquisition orientation. However, while the analysis of sagittal plane images is pertinent for numerous clinical applications, it is not the only feasible method for diagnosing Alzheimer's, as elaborated in this study. Consequently, it was essential to derive axial and coronal slices from the original patient images to extract valuable information from diverse viewpoints. To accomplish this, a Python algorithm was devised to extract axial and coronal slices from the sagittal plane images provided (see *Figure 1*). This algorithm is accessible as [Appendix 1](#) and utilizes popular libraries for medical image processing, such as PyDicom, employing interpolation and resizing techniques to create the

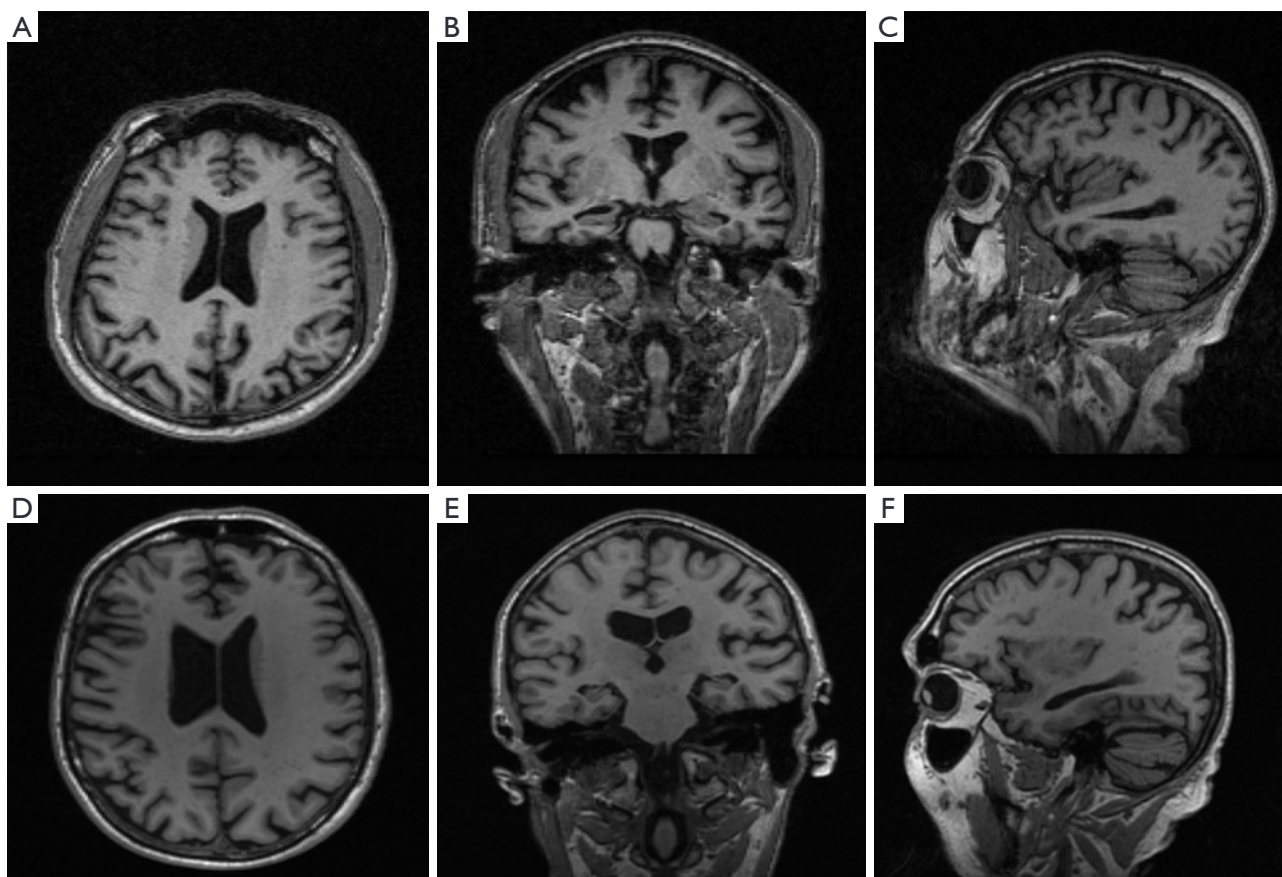


Figure 1 Comparison of axial (A,D), coronal (B,E), and sagittal (C,F) slices in T1-weighted MRI: (A-C) patient with mild cognitive impairment (ADNI subject ID: 067_S_0336, MMSE: 21); (D-F) healthy patient (ADNI subject ID: 033_S_0734, MMSE: 29). ADNI, Alzheimer's Disease Neuroimaging Initiative; MRI, magnetic resonance imaging; MMSE, Mini-Mental State Examination.

required slices.

The magnetic resonance images of the patients were carefully shuffled using a Python algorithm, available in the [Appendix 2](#), and then distributed among training, validation, and test sets, adhering to a standard split ratio of 70% for training, 15% for validation, and 15% for testing. This deliberate approach ensures that the CNN model undergoes training, validation, and testing on distinct and independent datasets. Such a practice not only mitigates the risk of overfitting but also facilitates an accurate assessment of the model's performance across both coronal and axial slices, thus enhancing the reliability and generalizability of our findings. By shuffling the images before the division into sets, we guarantee that each patient's images are randomly distributed among the sets, minimizing the possibility of any data leakage. This rigorous process enables an accurate assessment of the model's performance on coronal and axial slices, as it ensures

that the model is not biased by any unintended patterns or correlations present in the data. The distribution of images in each subdivision is presented in [Table 2](#).

This approach ensures that the CNNs model is trained across a variety of perspectives, thereby comprehensively testing its generalization capabilities. The number of images and their distribution in each subset were strategically determined to balance the size of each set while maintaining the representativeness and robustness of the data. This preprocessing strategy and data subdivision are crucial in ensuring the validity and reliability of the results obtained from our CNN analysis for Alzheimer's diagnosis based on magnetic resonance images.

CNNs

ANNs are mathematical models developed to enable

Table 2 Number of magnetic resonance images obtained in each of the groups for each available slicing plane in the dataset

Classification	Phase	Axial slice	Coronal slice	Sagittal slice
MCI	Training	3,987	6,619	4,935
	Validation	854	1,418	1,057
	Test	855	1,419	1,059
	Total	5,695	9,455	7,050
NC	Training	15,008	24,248	17,969
	Validation	3,216	5,196	3,850
	Test	3,217	5,197	3,852
	Total	21,440	34,640	25,670

NC, normal control; MCI, mild-cognitive impairment.

machines to solve problems like those of the human brain, mimicking information processing in the brain. Neural networks consist of processing units called ‘artificial neurons’. These units are organized into interconnected layers so that each neuron in one layer communicates with neurons in the next layer through weighted connections. The weight of these connections varies based on the importance of each input.

To achieve accurate and optimized results, networks must undergo extensive training and adjustments (17,24,37). The steps to reach the desired accuracy can be described as follows: (I) learning from a series of inputs representing the model of the data to be inserted; (II) propagating information through the neural network, traversing all layers from the input layer to the output layer; (III) calculating the error by comparing the obtained output with the desired output; (IV) propagating the error signal back to the input layer (backpropagation process); and (V) adjusting the weights of the network’s connections based on the previously measured error and repeating the process for various inputs until the network gradually improves its performance over epochs.

CNNs are feedforward neural network architecture used for image pattern recognition (38-41). CNNs can take images as input, assign weights to different image components, and thereby differentiate complex patterns within the input image. Additionally, CNNs are significant for reducing images to simpler formats, allowing for lighter processing without significant loss of essential information required for classification tasks. CNNs primarily consist of two types of layers to extract crucial features from the input image: the convolutional layer and the pooling layer, which

are successively repeated. In the Convolutional Layer, multiple filters (or kernels) are applied to the input image to extract important features and generate a feature map. This representation is achieved through matrix operations, as both the image and the kernel are matrices. The kernels convolve over the image during this process, traversing all pixels. The result is a new image that preserves essential information (high-level features) from the original, useful for tasks like classification, segmentation, and object detection (42,43). The pooling layer, typically following the convolutional layer, aims to reduce the spatial dimension of the processed image. It replaces blocks of pixels with a representative value (usually the maximum) in a process called MaxPooling. These layers are crucial for improving computational efficiency, suppressing noise, and maintaining important information in a condensed form.

In addition to these layers, a flattening layer is added to modify the data structure. Up to this point, we had images as two-dimensional matrices that needed to be transformed into a one-dimensional vector to ensure the continuity of the flow of information to the subsequent layers of the network, which can only process this type of structure. Finally, the fully connected (or dense) layers are added at the network’s end, following the convolutional, pooling, and flattening layers. Their objective is classification based on interpreting high-level features generated by the previous layers through complex reasoning. Thus, we can say that this layer functions as a nonlinear classifier composed of artificial neurons (44).

In this study, we constructed a CNN using the Sequential API from Keras in Python (45). The initial layer of our network is a convolutional layer, named Conv2D,

Table 3 Architecture and parameters of the CNNs described in this work

Layer ID	Layer name	Kernel number	Kernel size	Output image size
1	Input	–	–	128×128
2	Conv2D + ELU	32	3×3	128×128
3	MaxPooling2D	–	2×2	64×64
4	Conv2D_1 + ELU	32	3×3	64×64
5	MaxPooling2D_1	–	2×2	32×32
6	Conv2D_2 + ELU	64	3×3	32×32
7	MaxPooling2D_2	–	2×2	16×16
8	Flatten	–	–	256
9	Dense (Elu)	1,024	–	1,024
10	Dense_1 (Elu)	512	–	512
11	Dense_2 (Softmax)	2	–	2
12	Binary_Cross-Entropy	–	–	–

CNNs, convolutional neural networks.

which takes input images and uses a convolution kernel (3×3) that convolves with the input to produce an output tensor. We utilized 32 filters in this layer. Following each convolutional layer, we applied the Exponential Linear Unit (ELU) activation function (46) to expedite the network's learning. We also employed “padding = same” to ensure the output image maintains the same size as the input. After the Conv2D layer, we added a MaxPooling2D layer to reduce the dimensionality of the feature maps generated by the preceding Conv2D layer, enhancing the network's efficiency and reducing overfitting potential. We used a pool_size of [2, 2] in the MaxPooling2D layer, indicating that the maximum pooling operation would be applied to a 2×2 pixel window in each feature map. We introduced an additional pair of Conv2D and MaxPooling2D layers with identical characteristics to the initial ones. The final pair of layers featured a convolutional layer with 64 filters while maintaining the other parameters unchanged. The subsequent pooling layer remained the same as the previous ones. Finally, we added a flattening layer that converts the output into a one-dimensional vector for the subsequent network layers. From this stage, CNN behaves like a traditional network. Three dense layers composed of densely connected multilayer perceptrons were added for classification purposes, aiming to minimize errors. We utilized the ELU activation function for the first two layers and Softmax for the last layer and applied binary cross-entropy (Log Loss) as the loss function to

assess the network's performance, minimize mean squared error, and consequently optimize the system. The network architecture and other information are described in *Table 3*.

For the optimal training of this network, different numbers of epochs of training (20, 50, 80, and 100) and batch sizes (8, 16, 32, and 64) were tested to find the best possible configuration for each type of anatomical slice. Various parameters, such as accuracy, sensitivity, specificity, and F1-score, inform the network's performance.

Accuracy represents the proportion of correct predictions compared to the total predictions and serves as a general performance metric for the model. Sensitivity, also known as recall, highlights the model's ability to correctly identify positive cases, which is particularly relevant in Alzheimer's detection, where early detection is crucial. Specificity measures the model's ability to correctly identify negative cases, which is crucial in minimizing false alarms. The F1-score, in turn, is a metric that combines precision and recall, aiming for a balance between the model's ability to classify both positive and negative cases correctly. Together, these metrics provide a comprehensive and balanced evaluation of the model's performance in Alzheimer's diagnosis.

Experimental setup

The analysis of MRI and the implementation of CNNs were conducted on an Acer Nitro 5 notebook (AN-515-

44-R629). This equipment is equipped with an AMD Ryzen 7-4800H Octa-Core processor (with 16 threads), 16 GB of DDR4 RAM operating at 3,200 MHz, and a dedicated NVIDIA GeForce GTX 1650 graphics card with 4 GB of memory. The operating system used was Windows 10, version 2H22. For the implementation of CNNs, Jupyter Notebook (version 6.5.4) was used, with codes written in Python (version 3.11.5).

Analysis and statistics

Statistical analysis plays a pivotal role in assessing the performance of CNNs in our study. We employed descriptive statistics, encompassing means and standard deviations, to scrutinize key metrics evaluating CNN performance: accuracy, sensitivity, specificity, and the F1-score. Through a meticulous statistical examination of these metrics, we gained a profound understanding of CNN's performance in our classification task, enabling a critical assessment of our models' efficacy against established criteria. This robust statistical approach was indispensable for result interpretation and informed decision-making within the scope of our research.

Results

We conducted a comparative analysis of three categories of medical image slices (axial, coronal, and sagittal) and four batch size configurations (8, 16, 32, and 64 batches). The performance metric employed was accuracy, evaluated across different epochs (20, 50, 80, and 100 epochs). *Figure 2* presents radar-type graphs illustrating the accuracy achieved for each slice concerning the various epochs and batches configurations.

At 20 epochs, the coronal slice yielded an accuracy 0.99 across all batch sizes. The axial slice also achieved an accuracy of 0.99, but only for batches 16 and 64, with a slight decrease to 0.98 for batches 8 and 32. The sagittal slice demonstrated an accuracy of 0.99 for batches 8 and 16, with a minor decrease to 0.98 and 0.97 for batch sizes 32 and 64, respectively.

Upon increasing to 50 epochs, the accuracy of the coronal slice improved to 1.00 for batches 8 and 16 while remaining at 0.99 for batches 32 and 64. The axial and sagittal slices showed equal accuracies, reaching 0.99 for batches 8 and 16 and decreasing slightly to 0.98 for batches 32 and 64.

With a further increase to 80 epochs, the coronal slice

maintained a high accuracy 1.00 for batch size 16 and 0.99 for the other batch sizes. The axial slice achieved an accuracy of 0.99 for batches 16 and 32 and 0.98 for the other batch sizes. The sagittal slice consistently hit the 0.99 mark for all batch sizes excluding batch 64, where it achieved an accuracy of 0.98.

Lastly, at 100 epochs, the coronal slice achieved an impressive accuracy of 1.00 for batches up to a size of 32, with a slight decrease to 0.99 for batch 64. The axial slice maintained a high accuracy of 0.99 for batch sizes 8 and 16 but was reduced to 0.98 for larger sizes. The sagittal slice demonstrated consistency by achieving an accuracy 0.99 across all batch sizes.

This comprehensive analysis allowed for a thorough evaluation of the performance of the different slices across various batch sizes and epochs.

The analysis of *Figure 3* provides substantial insights into the performance of CNNs in medical image classification. The layered metrics present a holistic perspective on performance fluctuations concerning varying batch sizes and slice orientations. The bar graphs in this figure meticulously examine metrics such as "Precision-NC", "Precision-MCI", "Recall-NC", "Recall-MCI", "F1-score-NC", and "F1-score-MCI". Each metric is organized into grouped arrangements to facilitate a more straightforward interpretation. Moving from left to right, these groups correspond to different configurations of the number of training epochs.

Furthermore, within each metric grouping, we present three distinct sets, each representing a different orientation of magnetic resonance image slices—axial, coronal, and sagittal. For each slice, we separate each batch size (8, 16, 32, and 64 batches), which are identified by different colors in the legend, this detailed visual layout enables a thorough comparison and analysis of performance metrics across various batch sizes and image orientations. By incorporating these details, the figure enhances clarity. It provides a more nuanced understanding of the study's outcomes in the context of AD diagnosis through medical image classification using CNNs.

All evaluated parameters exhibit values above 90%, demonstrating a consistently robust model performance. Overall, the coronal slice showed the highest precision, followed by the axial and sagittal slices. However, it is noteworthy that in specific configurations, the sagittal slice demonstrates superior performance to the axial slice, especially in metrics such as recall and the F1-score for the MCI case. This variation underscores the importance

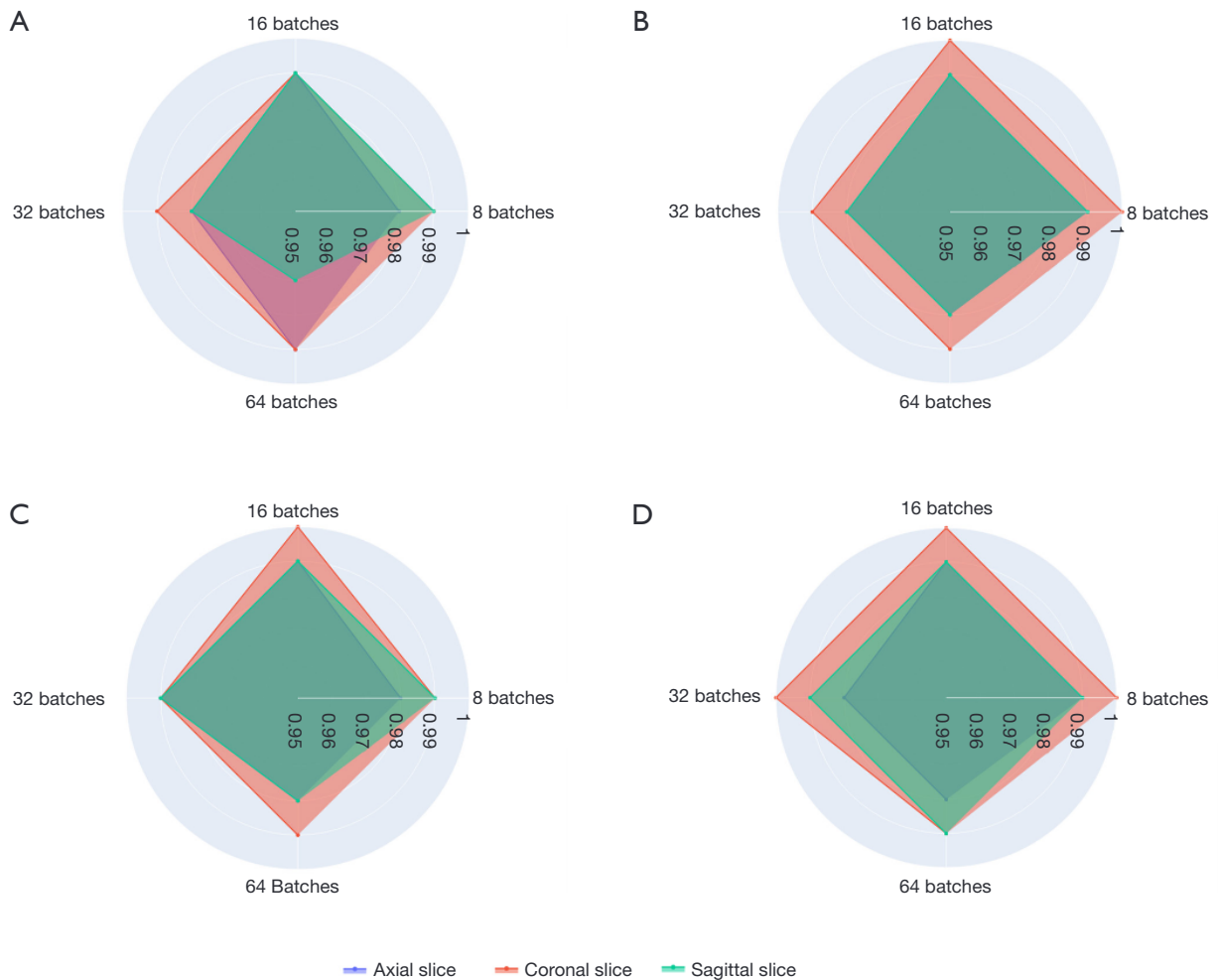


Figure 2 Radar graphs display the accuracy outcomes of network classification for various epoch configurations (20, 50, 80, and 100, corresponding to images A, B, C, and D) and batch sizes (8, 16, 32, and 64). It is evident that, for most model configurations, the coronal slice (depicted in red) consistently achieved the highest accuracy values.

of considering different image slices when optimizing the model for specific cases in classifying neurodegenerative diseases.

Discussion

This research aimed to explore the impact of the selection of anatomical slice orientation in medical images, in conjunction with filtering by education level and MMSE scores, on the accuracy of AD classification using CNNs. The CNN architecture employed consisted of three 2D convolutional layers (Conv2D) interspersed with MaxPooling2D layers. A flattening layer was used to convert the image into a one-dimensional vector,

followed by three dense layers (comprising 1,024, 512, and 2 neurons, respectively).

Patients who had completed a minimum of 16 years of education were categorized into two groups based on their MMSE scores: those with scores ranging from 21 to 24 were classified as having MCI, while those between 25 and 30 were classified as healthy. The differentiation of patients with MCI from healthy individuals using MRI poses a challenge due to the limited visibility of brain degeneration in the early stages of AD. Moreover, our CNN demonstrated exceptional performance in classifying these two groups using T1-weighted magnetic resonance images, achieving an accuracy exceeding 99% in most batch and epoch combinations. However, the F1-score

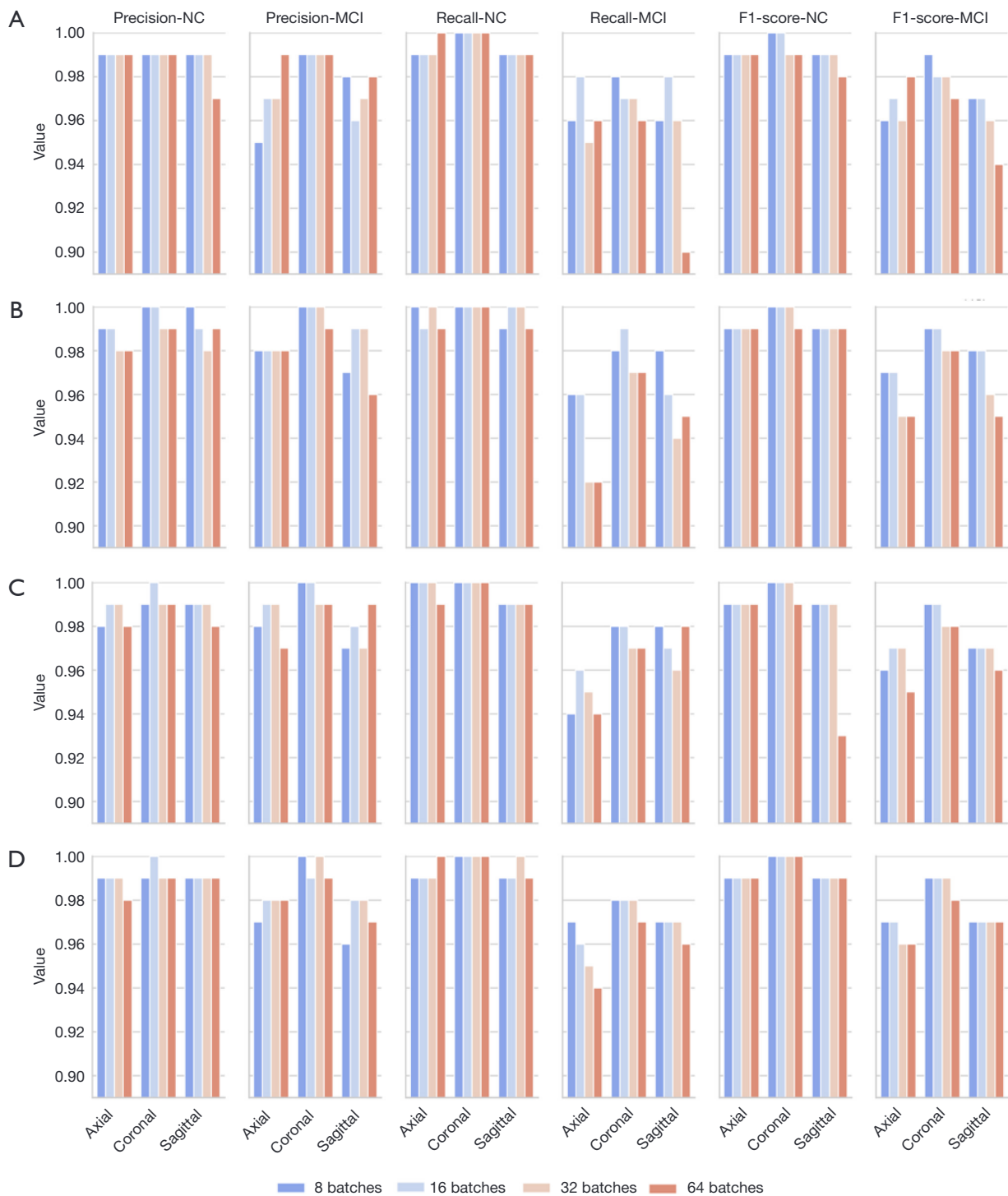


Figure 3 Performance evaluation of CNNs in various configurations. This figure shows 24 stacked bar charts encompassing the metrics “Precision-NC”, “Precision-MCI”, “Recall-NC”, “Recall-MCI”, “F1-score-NC” and “F1-score-MCI”. These metrics are evaluated concerning four different batch size configurations (8, 16, 32, and 64 batches) for four different epoch configurations (A, B, C, and D refer to 20, 50, 80, and 100 epochs, respectively). It can be seen that the coronal slice achieves better results than the sagittal and axial slices, which perform similarly, with a slight advantage for the sagittal slice. CNNs, convolutional neural networks; NC, normal control; MCI, mild cognitive impairment.

results exhibited significant variation among combinations, attributable to the differences in the number of images in the dataset between MMSE scores of 21–24 and 25–30.

Numerous studies (17-19,24,31,32,47-49) employ medical images to train CNNs using sagittal, coronal, or axial slices, depending on the available database. Nevertheless, the choice of slice can significantly influence the network's learning effectiveness, as evidenced by the results shown in *Figure 2*. The findings indicate that coronal slice orientation may be more suitable in medical image classification scenarios, given the consistently high accuracy rates achieved, followed by axial and sagittal slices. The difference in accuracy between coronal and axial slices is small, at just 0.01 percentage points. However, the difference in accuracy between the sagittal and axial slices is greater, at 0.02 percentage points. On the other hand, as more parameters such as Recall, F1-score, and others are analyzed, it can be seen that classification using the sagittal slice has a higher average performance than that based on the axial slice, and this difference is not directly related to the number of images used to train the neural network. In MRI scans, the number of images for each plane (coronal, sagittal, and axial) varies substantially, which makes it unfeasible to maintain a uniform proportion of images for each slice when training the model. In addition, we chose not to use the data-augmentation technique to align the number of images to avoid any potential bias that could be introduced by artificially manipulating the training images.

The variation in accuracy between slices can be attributed to several factors, including the viewing angle and the amount of information in the image. The coronal slice provides a frontal view of the body, the axial slice offers a transverse view, and the sagittal slice provides a longitudinal view. As a result, the coronal slice may offer more pertinent information for image classification.

The batch size configuration might also explain the discrepancy in accuracy between slices. Larger batches necessitate a more complex and robust deep-learning model. The sagittal slice, which achieved the lowest accuracy for batch size 64, might pose more of a challenge to classify with a more complex model.

The results presented in *Figures 2,3* demonstrate that the coronal slice is the most appropriate for classifying AD MRI images. This result could be clarified because the coronal slice provides a frontal view of the brain, which could be advantageous in identifying typical features of AD, such as cortical atrophy and senile plaques.

Strengths and limitations

The CNN developed in our work has relatively few layers (a total of 12) and was able to successfully classify healthy patients versus patients with MCI based on MMSE scores through T1-weighted MRI, achieving accuracy above 90% for all analyzed metrics regardless of the cutting orientation. All data were selected from the ADNI 1 database and followed a standardization proposed by the study itself, thus eliminating any unwanted bias for the classification performed by the network. Additionally, no data augmentation technique was used to increase the dataset, thus eliminating any bias the network could acquire during the learning of patterns related to the reuse of the same altered images. Another important point addressed by our study was how hyperparameters affect classification; therefore, we varied the number of training epochs (20, 50, 80, and 100) and batch size (8, 16, 32, and 64) to determine the best combination between image cutting orientation, the optimized number of training epochs that provide the best result, and batch size.

Although our article has demonstrated significant results, it is crucial to recognize that the use of a single database may be considered a potential limitation, as the predominance of data from a single geographical region (the USA) may result in a lack of representative diversity. The absence of validation on external datasets beyond ADNI 1, such as the Open Access Series of Imaging Studies (OASIS) database (50), although it did not compromise the robustness of our results, highlights an area for improvement. We recognize the importance of validating the robustness of our algorithm through comparisons across different datasets and are committed to incorporating this analysis into our future research.

Conclusions

This research emphasizes the significance of considering slice orientation and batch size when classifying medical images using CNNs. The results of this study suggest that the coronal slice is the most accurate for classifying medical images, followed by the axial and sagittal slices. Notably, in specific scenarios, the sagittal slice outperforms the axial slice, particularly in metrics like recall and the F1-score for classifying patients with MCI. This nuanced understanding emphasizes the significance of tailoring image slice considerations for optimizing the model, providing valuable insights for refining neurodegenerative

disease classification approaches.

Acknowledgments

Funding: We gratefully acknowledge the financial support provided by the Brazilian funding agencies CAPES (process number: 88887.912474/2023-00), which greatly contributed to the successful completion of this research. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U19 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study was coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for NeuroImaging at the University of Southern California.

Footnote

Data Sharing Statement: Available at <https://jmai.amegroups.com/article/view/10.21037/jmai-24-51/dss>

Peer Review File: Available at <https://jmai.amegroups.com/article/view/10.21037/jmai-24-51/prf>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://jmai.amegroups.com/article/view/10.21037/jmai-24-51/coif>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This article does not include any studies with human participants conducted by any of the authors. Institutional review board (IRB) approval and informed consent are waived.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Lutz W, Sanderson W, Scherbov S. The coming acceleration of global population ageing. *Nature* 2008;451:716-9.
2. Erkinen MG, Kim MO, Geschwind MD. Clinical Neurology and Epidemiology of the Major Neurodegenerative Diseases. *Cold Spring Harb Perspect Biol* 2018;10:a033118.
3. Katzman R. Alzheimer's disease. *N Engl J Med* 1986;314:964-73.
4. American Psychiatric Association. *DSM-5: Manual diagnóstico e estatístico de transtornos mentais*. Porto Alegre: Artmed Editora; 2015.
5. Zhang XX, Tian Y, Wang ZT, et al. The Epidemiology of Alzheimer's Disease Modifiable Risk Factors and Prevention. *J Prev Alzheimers Dis* 2021;8:313-21.
6. Bature F, Guinn BA, Pang D, et al. Signs and symptoms preceding the diagnosis of Alzheimer's disease: a systematic scoping review of literature from 1937 to 2016. *BMJ Open* 2017;7:e015746.
7. Alzheimer's Disease International. *Dementia facts & figures*. (2019). Retrieved November 29, 2023. Available online: <https://www.alzint.org/about/dementia-facts-figures/>

8. Cooper C, Li R, Lyketsos C, et al. Treatment for mild cognitive impairment: systematic review. *Br J Psychiatry* 2013;203:255-64.
9. Passeri E, Elkhoury K, Morsink M, et al. Alzheimer's Disease: Treatment Strategies and Their Limitations. *Int J Mol Sci* 2022;23:13954.
10. Yu TW, Lane HY, Lin CH. Novel therapeutic approaches for Alzheimer's disease: An updated review. *International journal of molecular sciences* 2022;22:8208.
11. Guzman-Martinez L, Calfio C, Farias GA, et al. New frontiers in the prevention, diagnosis, and treatment of Alzheimer's disease. *Journal of Alzheimer's Disease* 2021;82:S51-S63.
12. Burke AD, Goldfarb D. Facilitating Treatment Initiation in Early-Stage Alzheimer Disease. *J Clin Psychiatry* 2022;83:LI21019DH2C.
13. Nitrini R, Caramelli P, Bottino CMD, et al. Diagnóstico de doença de Alzheimer no Brasil: critérios diagnósticos e exames complementares. *Recomendações do Departamento Científico de Neurologia Cognitiva e do Envelhecimento da Academia Brasileira de Neurologia. Arquivos de Neuro-Psiquiatria* 2005;63:713-9.
14. Schilling LP, Balthazar MLE, Radanovic M, et al. Diagnóstico da doença de Alzheimer: recomendações do Departamento Científico de Neurologia Cognitiva e do Envelhecimento da Academia Brasileira de Neurologia. *Dementia & Neuropsychologia* 2022;16:25-39.
15. Burns A, Luthert P, Levy R, et al. Accuracy of clinical diagnosis of Alzheimer's disease. *BMJ* 1990;301:1026.
16. Chen Z, Mo X, Chen R, et al. A Reparametrized CNN Model to Distinguish Alzheimer's Disease Applying Multiple Morphological Metrics and Deep Semantic Features From Structural MRI. *Front Aging Neurosci* 2022;14:856391.
17. Basaia S, Agosta F, Wagner L, et al. Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. *Neuroimage Clin* 2019;21:101645.
18. Liu M, Cheng D, Yan W, et al. Classification of Alzheimer's Disease by Combination of Convolutional and Recurrent Neural Networks Using FDG-PET Images. *Front Neuroinform* 2018;12:35.
19. Oh K, Chung YC, Kim KW, et al. Classification and Visualization of Alzheimer's Disease using Volumetric Convolutional Neural Network and Transfer Learning. *Sci Rep* 2019;9:18150. Erratum in: *Sci Rep* 2020;10:5663.
20. Alzubaidi L, Zhang J, Humaidi AJ, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J Big Data* 2021;8:53.
21. Indolia S, Goswami AK, Mishra SP, et al. Conceptual understanding of convolutional neural network-a deep learning approach. *Procedia Computer Science* 2018;132:679-88.
22. Hidaka A, Kurita T. Consecutive dimensionality reduction by canonical correlation analysis for visualization of convolutional neural networks. In: *Proceedings of the ISCIE International Symposium on Stochastic Systems Theory and its Applications*; 2017:160-7.
23. Yamashita R, Nishio M, Do RKG, et al. Convolutional neural networks: an overview and application in radiology. *Insights Imaging* 2018;9:611-29.
24. Carcagnì P, Leo M, Del Coco M, et al. Convolution Neural Networks and Self-Attention Learners for Alzheimer Dementia Diagnosis from Brain MRI. *Sensors (Basel)* 2023;23:1694.
25. ADNI | Alzheimer's Disease Neuroimaging Initiative. (2017). Retrieved October 03, 2023. Available online: <https://adni.loni.usc.edu/>
26. Mueller SG, Weiner MW, Thal LJ, et al. The Alzheimer's disease neuroimaging initiative. *Neuroimage Clin N Am* 2005;15:869-77, xi-xii.
27. Vaz M, Silvestre S. Alzheimer's disease: Recent treatment strategies. *Eur J Pharmacol* 2020;887:173554.
28. Godýń J, Jończyk J, Panek D, et al. Therapeutic strategies for Alzheimer's disease in clinical trials. *Pharmacol Rep* 2016;68:127-38.
29. Atri A. Current and Future Treatments in Alzheimer's Disease. *Semin Neurol* 2019;39:227-40.
30. Congdon EE, Sigurdsson EM. Tau-targeting therapies for Alzheimer disease. *Nat Rev Neurol* 2018;14:399-415.
31. Vieira S, Pinaya WH, Mechelli A. Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications. *Neurosci Biobehav Rev* 2017;74:58-75.
32. Rathore S, Habes M, Iftikhar MA, et al. A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages. *Neuroimage* 2017;155:530-48.
33. ALSaeed D, Omar SF. Brain MRI analysis for Alzheimer's disease diagnosis using CNN-based feature extraction and machine learning. *Sensors* 2022;22:2911.
34. Silveira M, Marques J (2010). Boosting Alzheimer disease diagnosis using PET images. *20th International Conference on Pattern Recognition* 2010;2556-9.
35. Gray KR, Wolz R, Heckemann RA, et al. Multi-region analysis of longitudinal FDG-PET for the classification of

- Alzheimer's disease. *Neuroimage* 2012;60:221-9.
36. Ong HL, Subramaniam M, Abdin E, et al. Performance of Mini-Mental State Examination (MMSE) in long-stay patients with schizophrenia or schizoaffective disorders in a psychiatric institute. *Psychiatry Res* 2016;241:256-62.
 37. Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 1998;86:2278-2324.
 38. Albawi S, Mohammed TA, Al-Zawi S. Understanding of a convolutional neural network. In: 2017 International Conference on Engineering and Technology (ICET); 2017:1-6.
 39. Yamashita R, Nishio M, Do RKG, et al. Convolutional neural networks: an overview and application in radiology. *Insights Imaging* 2018;9:611-29.
 40. Kim P, Kim P. Convolutional neural network. *MATLAB deep learning: with machine learning, neural networks and artificial intelligence*. 2017;121-147.
 41. Kalchbrenner N, Grefenstette E, Blunsom P. A convolutional neural network for modelling sentences. *arXiv preprint* 2014. arXiv:1404.2188.
 42. Akilan T, Wu QJ, Safaei A, et al. A late fusion approach for harnessing multi-CNN model high-level features. In: 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC); 2017:566-71.
 43. Bayar B, Stamm MC. A deep learning approach to universal image manipulation detection using a new convolutional layer. In: *Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security*; 2016:5-10.
 44. Utsch KG. Uso de Redes Neurais Convolucionais para Classificação de Imagens Digitais de Lesões de Pele. Trabalho de Conclusão de Curso (Graduação em Engenharia Elétrica)-Departamento de Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo. Espírito Santo, 2018.
 45. Team K. (2023). Keras documentation: The Sequential class. Retrieved October 10, 2023. Available online: <https://keras.io/api/models/sequential/>
 46. Clevert DA, Unterthiner T, Hochreiter S. Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint* 2015. arXiv:1511.07289.
 47. Bron EE, Klein S, Papma JM, et al. Cross-cohort generalizability of deep and conventional machine learning for MRI-based diagnosis and prediction of Alzheimer's disease. *Neuroimage Clin* 2021;31:102712.
 48. Qiao H, Chen L, Ye Z, et al. Early Alzheimer's disease diagnosis with the contrastive loss using paired structural MRIs. *Computer methods and programs in biomedicine* 2021;208:106282.
 49. Chandra A, Dervenoulas G, Politis M, et al. Magnetic resonance imaging in Alzheimer's disease and mild cognitive impairment. *J Neurol* 2019;266:1293-302.
 50. OASIS Brains - Open Access Series of Imaging Studies. (2019). Retrieved March 13, 2024. Available online: <https://www.oasis-brains.org/>

doi: 10.21037/jmai-24-51

Cite this article as: Ramalho BAC, Bortolato LR, Gomes ND, Wichert-Ana L, Padovan-Neto FE, da Silva MAA, de Lacerda KJCC. The impact of the orientation of MRI slices on the accuracy of Alzheimer's disease classification using convolutional neural networks (CNNs). *J Med Artif Intell* 2024;7:35.

Appendix 1

Python code for obtaining anatomical slices from a DICOM image

```

import os
import pydicom
import numpy as np
import matplotlib.pyplot as plt
import warnings

dicom_dir = 'Add the path of the files in dicom format'

dicom_files = [os.path.join(dicom_dir, filename) for filename in os.listdir(dicom_dir) if filename.endswith('.dcm')]

def load_dicom_data(dicom_files):
    slices = [pydicom.read_file(dicom_file) for dicom_file in dicom_files]
    slices.sort(key=lambda x: int(x.InstanceNumber), reverse = False)

    if len(slices) >= 2:
        slice_thickness = np.abs(slices[0].ImagePositionPatient[2] -slices[1].ImagePositionPatient[2])
    else:
        print('There is only one slice in the list.')

    image_shape = list(slices[0].pixel_array.shape)
    image_shape.append(len(slices))
    volume_data = np.zeros(image_shape, dtype=slices[0].pixel_array.dtype)

    for i, slice in enumerate(slices):
        volume_data[:, :, i] = slice.pixel_array

    return volume_data, slice_thickness

volume_data, slice_thickness = load_dicom_data(dicom_files)
last = int(input('Enter the value of the last image saved in the folder:'))

for i in range(40,140): # This range depends on the slice orientation and the image.
    image_slice = volume_data[i, :, :] #volume_data[axial, coronal, sagittal]

    plt.figure(figsize=(6, 6))
    plt.imshow(image_slice, cmap='gray')
    plt.axis('off')
    warnings.filterwarnings("ignore")
    last += 1
    plt.savefig(f'{last}.png', bbox_inches='tight', pad_inches=0.0)

```

Appendix 2

Python code for creating datasets with shuffled images

```
from pathlib import Path
import splitfolders

input_folder = 'images'
splitfolders.ratio(input_folder, output = 'dataset',
                  seed = 42, ratio = (.7,.15,.15),
                  group_prefix = None)
```