# Model Based on Deep Feature Extraction for Diagnosis of Alzheimer's Disease

Iago R. R. Silva*, Gabriela S. L. Silva†, Rodrigo G. de Souza‡,
Wellington P. dos Santos*, Roberta A. de A. Fagundes*
*Department of Computer Engineering, University of Pernambuco, Brazil
†Department of Biomedical Engineering, Federal University of Pernambuco, Brazil
‡Center for Informatics, Federal University of Pernambuco, Brazil
irrs@ecomp.poli.br, gabriela.lucas@ufpe.br, rgs@cin.ufpe.br, wps@ecomp.poli.br, roberta.fagundes@upe.br

*Abstract*—**Alzheimer's disease (AD) is a neurodegenerative disease that results in loss of cognitive ability of the patient. Computational intelligence, more specifically Deep Learning, has been a powerful method for AD diagnosis. In this work we propose a model for AD diagnosis based on deep feature extraction for the classification using magnetic resonance imaging. This model aims to classify AD vs. HC (Healthy Controls). The database used in this project is the Minimal Interval Resonance Imaging in Alzheimer's Disease (MIRIAD), for validation of the proposed method. We select thirty slices from the upper region of the brain, above the eyes, for the apprenticeship in this work. The Convolutional Neural Network (CNN) architecture is designed in three convolutional layers to extract the best features of the selected region. After that, we put the selected attributes in a vector for learning and detection of patterns by another technique of computational intelligence. Finally, the data are partitioned with the 10-folds cross-validation method and trained with the Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbor (K-NN) algorithms with different parameters for evaluation. The results of accuracy are 0.8832, 0.9607 and 0.8745, for the algorithms mentioned above, respectively. According to a comparative analysis performed with other works of the literature, we can prove the efficiency and reliability of the model for the diagnosis of Alzheimer's disease.**

*Keywords*—**Alzheimer's Disease, Diagnosis, Deep Learning, CNN, SVM.**

## I. INTRODUCTION

Alzheimer's disease (AD) is a neurodegenerative disease. AD presents as a progressive loss of behavioral and intellectual characteristics. The decline of memory, language, and perception are some of the problems [1]. AD is related to aging and being the most common form of dementia. The investigation of its causes and repercussions are of interest to researchers. Researchers search to understand the pathology of AD because it's not entirely understood.

Neurochemical changes are some of the defining traits of the disease. The presence of neurofibrillary tangles and the accumulation of amyloid plaques characterize the AD. Also, there are anatomical neuro alterations that are closely associated with AD. The atrophy of the cortical regions and the hippocampus region, in contrast to a ventricular are alterations [1].

Magnetic Resonance Imaging (MRI) is essential for the diagnosis of AD. The specialists in neurology make the diagnosis through the image or signals [2] examination analysis. Currently, specialists have used clinical applications that assist them in making decisions. Applications of this type use Computational Intelligence. These applications can determine with high accuracy if the patient is healthy or has the disease. Currently, several Computational Intelligence techniques are used for the classification of AD. Among these techniques, we can highlight Deep Learning. It's is an approach that has extensive use in last years in AD diagnosis [3] [4] [5].

Currently, the majority of pattern recognition applications use Deep Learning models. When there is an image classification problem with a large quantity of data, the Convolutional Neural Networks (CNN) is better to use. This model can generate the best features to classification task [6]. The issue of Alzheimer's disease classification provides a lot of image data [7]. Also, it's necessary to get a high accuracy performance, offering the use of CNN's.

In this work, we propose a model for binary AD diagnosis based on deep feature extraction for the classification using MRI. The aim is to correctly classify patients that contain AD and who does not have the disease with high precision. The model consists in the use of features generated by CNN and another technique to classify these features. Firstly, the CNN is pre-trained to get the best features that can be used to classification. Then, with these data, the supervised machine learning algorithms are used to classification task. The use of other algorithms aims to maximize the accuracy of the model.

The rest of this work is organized as follow. Section II presents the background on binary AD classification. Section III shows the methods such as dataset used, preprocessing, network architecture, and generally used parameters in learning algorithms. Section IV provides the obtained results of the proposed approach. Finally, section V provides the conclusions and future works.

## II. LITERATURE REVIEW

### A. AD Binary Classification

The classification of AD is the target of researchers in the literature. In general, the methods for this purpose consist of three main steps, such as: the acquisition of images, extraction of features, and the pattern recognition. In this section, we present some works present in the state of the art that use the approach we mentioned.

The work in [8] proposes an approach for the extraction of characteristics in MRI images. The method is called Hybrid Forward Sequential Selection (HFS), to identify regions of the brain that are correlated with the brain. Also, the statistical technique Principal Component Analysis (PCA) is used to reduce the features. The work gets an accuracy of 0.912 and sensitivity and specificity equals to 0.886 and 0.923, respectively.

Another work [9] uses data obtained through the MRI voxels in the feature extraction step. The voxels used are named Volumes of Interest (VOIs) and are selected from the Gray Matter (GM) segmentation. The generated features are put into a genetic algorithm for rankings of relevance and correlation. Finally, ranked data is trained with the SVM classifier. The work gets an accuracy of 0.9301 and sensitivity and specificity equals to 0.8913 and 0.9680, respectively.

The work in [10] proposes an approach with the extraction of features of multi-modal data. The data are extracted from MRI and FDG-PET images. Also, they use Cerebrospinal Fluid Data (CSF) and genetic data from patients. All this information is combined for training with the Random Forest algorithm. The work gets an accuracy of 0.862 and sensitivity and specificity equals to 0.851 and 0.861, respectively.

### B. AD Binary Classification using a Deep Learning Approach

Approaches using Deep Learning are more familiar to be used in recent years. The same phenomenon happens with AD classification. The use of depths is highlighted in this context due to its high capacity for generalization. The use of deep networks in sorting has some major strands having the use of a CNN, deep MLP, or autoencoders [11]. While the MLP stands out in the classification context, CNN and autoencoders stand out in the context of feature extraction. These techniques are not limited to application to MRI images, they are also used in PET-SCAN images [12] and Electroencephalography signals [13], then consequently achieve good results in their methodologies.

The MLP approach has the same guidelines as explained in the previous section. In this network, the input neurons are usually the feature arrays generated by the extraction of features. Then there is a complex structure that defines the network structure, containing many layers with many neurons [14]. The number of executions takes place through epochs, while the network learns and updates the weights of all neurons. In the paper [3] the authors use deep MLP

for classification. The features were generated through a spectral matching. The results for accuracy are equal to 0.84, and for sensitivity and specificity are equal to 0.73 and 0.89, respectively.

Another type of network with a deep approach is the autoencoders. Autoencoders work by displaying their input in their output. The purpose of this type of network is to perform the learning of the encodings of the features [15]. The work [16] presents an approach of detection of the AD using autoencoders in the extraction of information of the data. In the classification step, the softmax regression is used, which is an activation function for multiclass classification [17]. The model presented an accuracy, sensitivity, and specificity of 0.8776, 0.8857 and 0.8722, respectively. The same authors applied the same neural network structure in another approach at work [18]. The difference is the use of MRI images along with PET. The metrics mentioned above had an increase to 0.8792, 0.9366 and 0.8722, respectively.

Finally, another type of deep network used for AD classification is CNN. A CNN works with the learning of convolutional layer filters to extract features (more details in section III-B). The work [4] uses an approach to AD classification using CNN. First, all images in the database corresponding to the upper brain are selected. These images are then used as input to the network. The learning is performed, and classification is done through the sigmoid activation function. The work reached an accuracy of 0.8485, while sensitivity and specificity equal 0.8593 and 0.8314, respectively. A similar approach was proposed in [5], where a CNN was also used. The difference is the data input and the robustness of the network. CNN data entry was previously selected landmarks in the database, while the network structure had more layers. In turn, the data classification was performed through the softmax activation function. About the metrics mentioned above, the model reached 0.9275, 0.9348 and 0.9130, respectively.

## III. METHODS

In this section, we discuss the methods used in this work, such as dataset used, data pre-processing, CNN network proposed, how we get the features from CNN network, data preparation for classifiers and classification. The pipeline of the these methods is illustrated in Fig. 1. Then, the next subsections explain each of the stages declared in the pipeline.

### A. Data Acquisition and Pre-processing

We utilized all data from MIRIAD (Minimal Interval Resonance Imaging in Alzheimer's Disease) dataset [19] from UCL (University College London). This dataset contains Magnetic Resonance Imaging (MRI) of 69 persons divided into 2 groups: the first with 23 healthy pacients (healthy controls, or HC), and the second with 46 patients with Alzheimer's Disease (AD). This dataset provides two classes for binary classification. In this work, we perform the
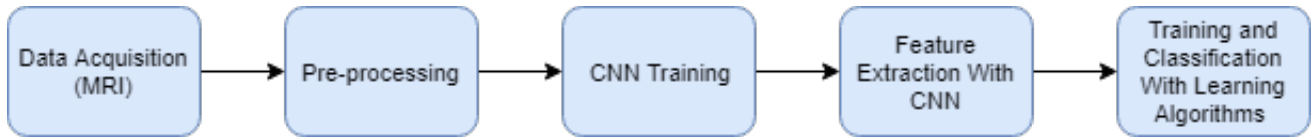
paper N-19554.pdf

Fig. 1. The methods used in this work.

investigation of the application of the proposed model in this context.

The dataset was created by the longitudinal approach, where the patients make the image acquisition in 7 visits over 52 weeks. In total, 39 patients completed the visits, and among these, 22 patients were sent to make another exam. The total quantity of completed exams in this dataset is 708. Table I shows the demographic data of the dataset.

TABLE I
DEMOGRAPHIC DATA OF THE MIRIAD DATASET.

| Class | Male/Female | Age (Mean±std) |
|-------|-------------|----------------|
| NC | 12/11 | 69.7±7.2 |
| AD | 19/27 | 69.4±7.1 |

Each examination represents a file containing images corresponding to the brain of each patient, representing a three-dimensional plane. First, we did the conversion of these files into two-dimensional images, represented by slices. The choice of slices is an essential process for the recognition of patterns in medical images, because if the slices are selected incorrectly, it may impair the learning of the algorithms. Therefore for each examination we selected 30 slices in the axial plane, located above the eyes. Fig. 2 Illustrates this procedure.
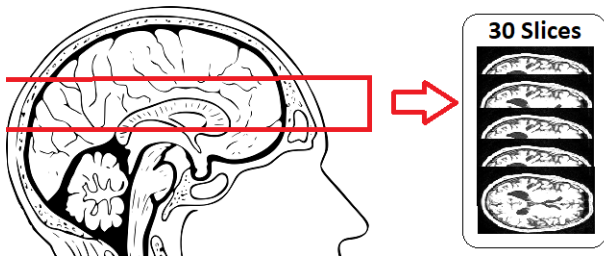


Fig. 2. Procedure for acquisition of MR slices.

After the acquisition of the slices, we normalized the images to values ranging between 0 and 1 to support the sigmoid decoders present at CNN, finishing the pre-processing step.

### B. CNN Training and Feature Extraction

With the use of a CNN, it is not necessary to perform image segmentation. Thus, we did not previously segment images into our approach. Other methods of image classification need this step to cut possible noise and binarization of the image [20]. CNN works by extracting the best features present in an image, following a definition through the

class. In our model, the slices obtained are directly placed as input to CNN. Then the process of extracting attributes is performed on all the slices. Fig. 3 shows the CNN architecture used in this work.
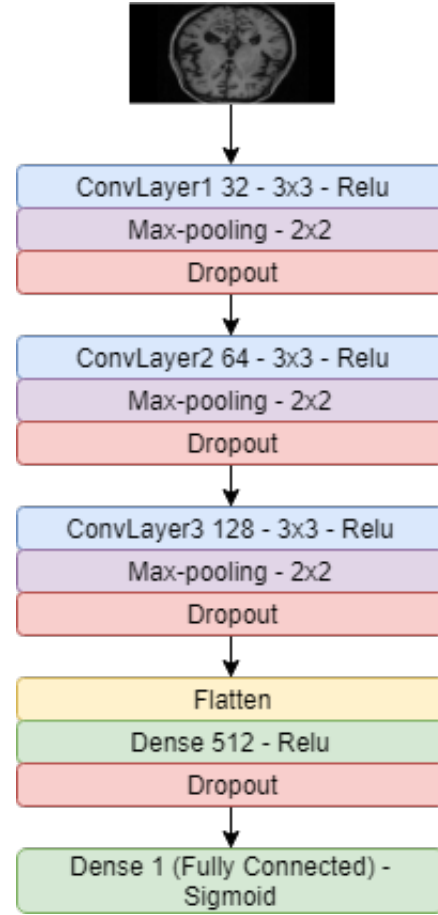


Fig. 3. CNN Architecture used in this work.

We define three convolutional layers that are responsible for the application of 3x3 filters. These layers are responsible for extracting the features present in an image [21]. At each layer, we increase the number of filters by multiplying by 2. The number of filters used in the first layer was 32, the second layer was 64, and the third layer was 128. At each layer drop, the number of features used becomes lower, causing loss of information. Because of this, only three convolutional layers were defined. With the network training process, the filters are automatically adjusted for activation in the most relevant features. The activation function we use in the layers for this work is the

Rectified Linear Units (ReLU). The Equation (1) shows the ReLU function.

$$f(x) = max(0, x) \tag{1}$$

There are layers of pooling after the convolution layers [6]. A single value replaces the values belonging to a region in the features map. This is done to decrease the number of attributes to drop to the next convolution layer. In the pooling layers, we use the $max()$ function, with 2x2 window, to get the highest values of each region. After each layer of pooling, we define layers of the dropout type. The dropout layers are responsible for ignoring a certain amount of neurons in a network [22]. This process is performed in the training phase of the network. The goal of adding these layers is to reduce the risk of overfitting on CNN. This forces the deep network to learn the most robust features for better network learning. The parameter used in the three dropout layers was 0.20, this is the rate of neurons to be ignored in the network.

In the fully-connected layer, we use only one neuron. Then we use the sigmoid activation function for simple classification. It's a binary classification problem, and then this activation function is better for it. Equation (2) shows the sigmoid activation.

$$f(x) = \frac{1}{1 + e^{-x}} \tag{2}$$

The fully-connected layer input data is generated through the last convolution layer of this project, with 128 filters. For that, the data is reshaped and placed in a single array with 512 elements, through the flatten. Thus, the model is pre-trained in $n$ epochs to learn the best features. Upon completion of CNN learning, the data from each image generated through the flattening are placed in a multidimensional array. These data are used for the learning step with another computational intelligence algorithm.

*C. Training and Classification With Another Learning Algorithms*

We apply this step with the aim of maximizing the accuracy and other metrics of the model. For this, three classification algorithms used in AD classification were defined: Random Forest [23], Support Vector Machine (SVM) [9], and K-Nearest Neighbors (K-NN) [24]. The Fig. 4 illustrates this step.

## IV. EXPERIMENTS AND RESULTS

The first subsection present the environment configuration parameters of the experiments. Then the next subsections present the results obtained.

*A. Experimental Settings*

We validate our method by applying the Alzheimer's versus Healthy Controls (AD vs. HC) classification task. We partitioned the MIRIAD database to perform CNN pre-training, was according to the holdout method. The size of
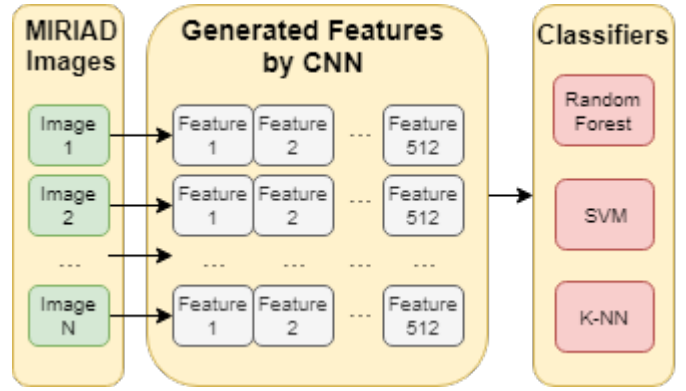


Fig. 4. The process of classification of extracted features. First, the features of the images are collected by CNN. Then the features are put into classifiers.

the training set used is 80%, while the other 20% is to test the model. Finally, the network is pre-trained, it's executed in 50 epochs to generate the feature array.

The feature array is used for the new learning. We use the before mentioned algorithms for this purpose. Table II shows the parameters used in each algorithm for the machine learning process.

TABLE II
THE PRINCIPAL PARAMETERS UTILIZED IN THE LEARNING ALGORITHMS.

| Algorithm | Parameter | Values Used |
|---|---|---|
| Random Forest | Iterations | 50, 100, and 150 |
| SVM Linear Kernel | - | - |
| SVM RBF Kernel | gamma (g) | 0.04, 0.07 and 0.10 |
| K-NN | K | 1, 4, and 7 |

The data validation method of this array is cross-validation with K = 10 (folds), where ten executions are performed to evaluate the model.

We adopt metrics to evaluate the prediction results. These depend on the amount of correctness of the positive (TP) and negative (TN) classes, as well as the errors of the positive (FP) and negative (FN) classes. The metrics were chosen to evaluate our model were: accuracy (ACC), sensitivity (SEN) and specificity (SPE), and the area under the ROC curve (AUC). The Equations (3), (4), and (5) illustrate the formulas that represent these metrics:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

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

$$SPE = \frac{TN}{TN + FP} \tag{5}$$

*B. Classification Results*

After applying the proposed method with the experimental configurations before defined, we get very close results in several metrics. Table III shows the results obtained.

paper N-19554.pdf

TABLE III
RESULTS OBTAINED WITH THE EXPERIMENTS OF AD VS HC CLASSIFICATION.

| Algorithm | ACC | AUC | SEN | SPE |
|---|---|---|---|---|
| Random Forest, 50 iterations | 0.8700 | 0.9570 | 0.8547 | 0.9400 |
| Random Forest, 100 iterations | 0.8800 | 0.9650 | 0.8585 | 0.9438 |
| Random Forest, 150 iterations | 0.8832 | **0.9680** | 0.8657 | **0.9567** |
| SVM, linear kernel | 0.9508 | 0.9460 | 0.9636 | 0.9264 |
| SVM, RBF kernel, g=0.04 | 0.9552 | 0.9552 | 0.9448 | 0.9500 |
| SVM, RBF kernel, g=0.07 | 0.9583 | 0.9583 | 0.9495 | 0.9539 |
| SVM, RBF kernel, g=0.10 | **0.9607** | 0.9569 | 0.9607 | 0.9531 |
| K-NN, K=1 | 0.8512 | 0.8750 | 0.9650 | 0.7142 |
| K-NN, K=4 | 0.8745 | 0.9440 | 0.9623 | 0.7672 |
| K-NN, K=7 | 0.8118 | 0.9520 | **0.9836** | 0.6599 |

When we analyze the results, we can observe some contrasts. The first one is with the application of the algorithm Random Forest. About classification, the algorithm achieved a high ACC rate of 0.8832, as well as the AUC rate of 0.9680, using 150 iterations. We note that there is a slight discrepancy between the two metrics because the hit rates of the two classes are unbalanced. We can reach this conclusion because of the values obtained for SEN to be equal to 0.8585 and SPE equal to 0.9438. That is, there is a higher hit rate than true negatives (patients who do not have AD). This fact makes Random Forest have the highest AUC rate. We can also observe that the number of iterations causes little influence on the metrics analyzed. The difference in values occurs only in the third decimal place.

The second contrast is with the application of the SVM algorithm. We have observed that the RBF kernel has obtained superior results in some metrics compared to the Linear kernel. About the linear kernel, the algorithm obtained a great ACC value among the algorithms under analysis, with the result equal to 0.9508, while the RBF kernel with g=0.10 obtained the highest ACC equal to 0.9607. Still, the other two algorithms obtained better results in other metrics. The AUC rate equals 0.9569 with RBF kernel is almost 0.01 points lower than the AUC rate obtained with Random Forest. However, despite being also exceeded in SPE with a value equal to 0.8531, it obtained a SEN equal to 0.9607, with its best parameter choice. It shows that although the AUC is smaller about Random Forest, it got a high capacity of generalization. There is a greater balance in the classification with both parameters and kernels, even though the classes are unbalanced. Through the high rate of correctness in both classes, we can come to this conclusion.

Finally, the K-NN algorithm has the best SEN rate, which is equal to 0.9836 using K=7. We used three different K values to evaluate our model, and in both we observed that the SEN metric always high. This indicates that the algorithm had a high success rate in the positive classes. This fact positively influenced the AUC metric, obtaining 0.9520. The result obtained in the SPE metric is much lower when compared to those obtained with Random Forest and with SVM, using any parameters. This fact has a direct impact on the ACC, also getting the lowest of the three

algorithms, with a result corresponding to 0.8118, using K=4.

We used the RBF kernel in SVM, and it get the best classification results. This provided not only a high ACC result but also a high hit ratio in both classes. While in the other two algorithms there has always been a disproportion about the hit rates of the classes. With this, we can affirm that the CNN generated a hyperplane that can separates correctly the features. The configuration of the gap chosen provided good separation between the two classes. This is relevant for medical diagnosis, as false diagnoses can affect patients' quality of life. In spite of obtaining a considerable high accuracy, one can not rely on K-NN and Random Forest in our method, due to the low rate of hits of the true or false negatives. Thus, we can also say that SVM with RBF kernel is the most suitable for application in our approach.

## V. COMPARISON WITH PUBLISHED WORKS

State of the art for classification and diagnosis of AD for the binary class using either deep learning or non-deep learning is around 0.84. For comparison, we selected works of this type and using similar experimental configuration parameters. Table IV presents a comparative table with approaches, configuration parameters and results obtained in each work cited. Our approach is shown with high precision to solve this classification problem, reaching an ACC equal to 0.9607. This indicates that it is competitive about previously published works.

Analyzing methods that do not use deep learning, we can attest some observations. Regardless of the feature extractor used, accuracy was always higher using the SVM. This fact also occurs in our work, with this algorithm having superior results in the ACC metric. In this context of the use of SVM, the work [9] obtained the best results for SPE, with the value equal to 0.9680. It demonstrates that the SVM is one of the algorithms that can best to apply to solve the problem of AD classification. Despite the mathematical robustness of these feature extractors, none of these works outperformed the ACC of our model. Thus, we can verify a superiority of the methods using CNN to extract characteristics about other methods presented.

In the analysis of the methods that use deep learning, we can also attest some observations. About the attribute extraction criterion for this research problem, CNN provides better results when compared to autoencoders observing all metrics analyzed. But, this does not detract from these approaches, which presented ACC greater than 0.87. Also, we can attest that the use of a hybrid model (CNN + SVM) gives a better result when compared to that obtained using only CNN. This is the case of works [4] and [5], which presented an approach with sigmoid and softmax in the classification layer and obtained ACC results smaller than ours. With this, we can attest that the use of a more robust classification algorithm will tend to get better results of accuracy, from a hybrid approach.

paper N-19554.pdf

TABLE IV
COMPARISON OF ALZHEIMER'S DISEASE BINARY CLASSIFICATION (AD VS HC) WITH PUBLISHED WORKS.

| Ref. | Dataset | Feature Extraction | Classification Algorithm | Data Partitioning | ACC | AUC | SEN | SPE |
|---|---|---|---|---|---|---|---|---|
| Liu et al. (2014) [16] | ADNI | Stacked Autoencoders in Grey Matter | Softmax | Cross Validation (10 folds) | 0.8776 | - | 0.8857 | 0.8722 |
| Liu et al. (2015) [18] | ADNI | Stacked Autoencoders (MRI+PET) | Softmax | - | 0.8792 | - | 0.9366 | 0.8861 |
| Shakeri et al. (2016) [3] | ADNI | Spectral Matching | Deep MLP | - | 0.8400 | - | 0.7300 | 0.8900 |
| Han and Zhao (2016) [8] | ADNI | HFS | SVM | Cross Validation (10 folds) | 0.9120 | - | 0.8860 | 0.9230 |
| Beheshti et al. (2017) [9] | ADNI | Voxel Values | SVM | Cross Validation (10 folds) | 0.9301 | 0.9351 | 0.8913 | **0.9680** |
| Tong et al. (2017) [10] | ADNI | Graphs | SVM | Holdout (75% train \| 25% test) | 0.8620 | 0.9300 | 0.8510 | 0.8610 |
| Silva et al. (2018) [4] | MIRIAD | CNN | Sigmoid CNN | Holdout (80% train - 20% test) | 0.8485 | - | 0.8593 | 0.8314 |
| Liu et al. (2018) [5] | MIRIAD | CNN (landmarked points) | Softmax CNN | ADNI-1 as train and ADNI-2 as test | 0.9275 | **0.9716** | 0.9348 | 0.9130 |
| Our Work | MIRIAD | CNN | SVM | Cross Validation (10 folds) | **0.9607** | 0.9569 | **0.9607** | 0.9531 |

Our approach provide a balance in the learning of both classes (AD and HC). For this reason, it obtained superior ACC about the analyzed works, being a deep approach or not. Our approach presented lower values of AUC concerning [5], and SPE concerning [9]. However, our SEN surpassed the others, and we obtained a very close SPE. This fact reflects in the balanced amount of correctness of the two classes, even though the base is unbalanced.

The main contribution of our work is the fact that CNN generates optimal and precise features for the research problem of this work. Also, these features cause a booster in accuracy if use in a hybrid approach with another more robust classifier. We can prove it by the high hit rates obtained in all algorithms that have been tested. Moreover, the application of the SVM algorithm was fundamental for the balance in the quantity of correctness of both classes. The SVM algorithm has robustness and ease of learning about AD classification. This also provides a booster in accuracy. The proposed architecture is not new in the literature [25], however its application in MRI for diagnosis of AD has not yet been applied. The contribution is clearly explicit through the results obtained in all metrics analysed.

## VI. CONCLUSION

In this work, we propose a method for AD diagnosis. The method consists of a approach between feature extraction through CNN and classification with another algorithm. The acquisition of images of this method includes the use of MRI slices from the upper region to the eye. Experiments were performed with the Random Forest, SVM and K-NN

algorithms, which obtained results of the accuracy of 0.88, 0.9508 and 0.8512, respectively.

The results prove the effectiveness of the model, which obtained an excellent generalization capacity, although the database is unbalanced. The model can generate optimal and precise features for the research problem of this work. Also, these features cause a booster in accuracy if used in a hybrid approach with another more robust classifier. Comparing the results of other works of state of the art, we can attest that our proposal is competitive with the others. It is superior in two of the four metrics evaluated. We conclude that our model is efficient for the diagnosis of Alzheimer's disease limited to diagnosis of AD using two classes (AD vs HC) for binary classification. As future work, we intend to apply and evaluate the proposed model in databases of three classes, including the Mild Cognitive Impairment (MCI) stage.

paper N-19554.pdf

## REFERENCES

[1] E. R. Kandel, J. H. Schwartz, T. M. Jessell, D. of Biochemistry, M. B. T. Jessell, S. Siegelbaum, and A. Hudspeth, *Principles of neural science.* McGraw-hill New York, 2000, vol. 4.

[2] N. Mammone, L. Bonanno, S. D. Salvo, S. Marino, P. Bramanti, A. Bramanti, and F. C. Morabito, "Permutation disalignment index as an indirect, eeg-based, measure of brain connectivity in mci and ad patients," *International journal of neural systems*, vol. 27, no. 05, p. 1750020, 2017.

[3] M. Shakeri, H. Lombaert, S. Tripathi, S. Kadoury, and ADNI, "Deep spectral-based shape features for alzheimer's disease classification," in *International Workshop on Spectral and Shape Analysis in Medical Imaging.* Springer, 2016, pp. 15–24.

[4] I. R. R. Silva, R. G. de Souza, G. S. Silva, C. S. de Oliveira, L. H. Cavalcanti, R. S. Bezerra, A. d. A. Roberta, and W. P. dos Santos, "Utilização de redes convolucionais para classificação e diagnóstico da doença de alzheimer," in *II Simpósio de Inovação em Engenharia Biomédica*, 2018, pp. 73–76.

[5] M. Liu, J. Zhang, E. Adeli, and D. Shen, "Landmark-based deep multi-instance learning for brain disease diagnosis," *Medical image analysis*, vol. 43, pp. 157–168, 2018.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[7] C. R. Jack Jr, M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P. J. Britson, J. L. Whitwell, C. Ward *et al.*, "The alzheimer's disease neuroimaging initiative (adni): Mri methods," *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 27, no. 4, pp. 685–691, 2008.

[8] Y. Han and X.-M. Zhao, "A hybrid sequential feature selection approach for the diagnosis of alzheimer's disease," in *Neural Networks (IJCNN), 2016 International Joint Conference on.* IEEE, 2016, pp. 1216–1220.

[9] I. Beheshti, H. Demirel, H. Matsuda, and ADNI, "Classification of alzheimer's disease and prediction of mild cognitive impairment-to-alzheimer's conversion from structural magnetic resource imaging using feature ranking and a genetic algorithm," *Computers in biology and medicine*, vol. 83, pp. 109–119, 2017.

[10] T. Tong, K. Gray, Q. Gao, L. Chen, D. Rueckert, and ADNI, "Multi-modal classification of alzheimer's disease using nonlinear graph fusion," *Pattern recognition*, vol. 63, pp. 171–181, 2017.

[11] L. Deng, D. Yu *et al.*, "Deep learning: methods and applications," *Foundations and Trends® in Signal Processing*, vol. 7, no. 3–4, pp. 197–387, 2014.

[12] Y. Ding, J. H. Sohn, M. G. Kawczynski, H. Trivedi, R. Harnish, N. W. Jenkins, D. Lituiev, T. P. Copeland, M. S. Aboian, C. Mari Aparici *et al.*, "A deep learning model to predict a diagnosis of alzheimer disease by using 18f-fdg pet of the brain," *Radiology*, vol. 290, no. 2, pp. 456–464, 2018.

[13] F. C. Morabito, M. Campolo, N. Mammone, M. Versaci, S. Franceschetti, F. Tagliavini, V. Sofia, D. Fatuzzo, A. Gambardella, A. Labate *et al.*, "Deep learning representation from electroencephalography of early-stage creutzfeldt-jakob disease and features for differentiation from rapidly progressive dementia," *International journal of neural systems*, vol. 27, no. 02, p. 1650039, 2017.

[14] G. F. Montufar, R. Pascanu, K. Cho, and Y. Bengio, "On the number of linear regions of deep neural networks," in *Advances in neural information processing systems*, 2014, pp. 2924–2932.

[15] L. Deng, M. L. Seltzer, D. Yu, A. Acero, A.-r. Mohamed, and G. Hinton, "Binary coding of speech spectrograms using a deep auto-encoder," in *Eleventh Annual Conference of the International Speech Communication Association*, 2010.

[16] S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, and D. Feng, "Early diagnosis of alzheimer's disease with deep learning," in *Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on.* IEEE, 2014, pp. 1015–1018.

[17] D. Heckerman and C. Meek, "Models and selection criteria for regression and classification," in *Proceedings of the Thirteenth conference on Uncertainty in artificial intelligence.* Morgan Kaufmann Publishers Inc., 1997, pp. 223–228.

[18] S. Liu, S. Liu, W. Cai, H. Che, S. Pujol, R. Kikinis, D. Feng, M. J. Fulham *et al.*, "Multimodal neuroimaging feature learning for multiclass diagnosis of alzheimer's disease," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 4, pp. 1132–1140, 2015.

[19] I. B. Malone, D. Cash, G. R. Ridgway, D. G. MacManus, S. Ourselin, N. C. Fox, and J. M. Schott, "Miriad - public release of a multiple time point alzheimer's mr imaging dataset," *NeuroImage*, vol. 70, pp. 33–36, 2013.

[20] I. R. R. Silva, R. A. A. Fagundes, and T. S. M. C. de Farias, "Techniques for automatic liver segmentation in medical images of abdomen," *IEEE Latin America Transactions*, vol. 16, no. 6, pp. 1801–1808, June 2018.

[21] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[22] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[23] K. Oppedal, T. Eftestøl, K. Engan, M. K. Beyer, and D. Aarsland, "Classifying dementia using local binary patterns from different regions in magnetic resonance images," *Journal of Biomedical Imaging*, vol. 2015, p. 5, 2015.

[24] S. H. Nozadi, S. Kadoury, and ADNI, "Classification of alzheimer's and mci patients from semantically parcelled pet images: A comparison between av45 and fdg-pet," *International journal of biomedical imaging*, vol. 2018, 2018.

[25] X.-X. Niu and C. Y. Suen, "A novel hybrid cnn–svm classifier for recognizing handwritten digits," *Pattern Recognition*, vol. 45, no. 4, pp. 1318–1325, 2012.

paper N-19554.pdf