2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)

Radiomics textural features extracted from subcortical structures of grey matter probability for Alzheimers disease detection.

César A. Ortiz Toro*, Nuria Gutiérrez Sánchez†, Consuelo Gonzalo-Martín‡, Roberto Garrido García§, Alejandro Rodriguez González¶ and Ernestina Menasalvas Ruiz‖
Centro de Tecnología Biomédica
Universidad Politécnica de Madrid
Pozuelo de Alarcón, Spain
Email:*ca.ortiz@upm.es, †nuria.gutierrezs@alumnos.upm.es, ‡consuelo.gonzalo@upm.es,
§roberto.garrido.garcia@alumnos.upm.es, ¶alejandro.rg@upm.es, ‖ernestina.menasalvas@upm.es

***Abstract*—Alzheimer's disease (AD) is characterized by a progressive deterioration of cognitive and behavioral functions as a result of the atrophy of specific regions of the brain. It is estimated that by 2050 there will be 131.5 million people affected. Thus, there is an urgent need to find biological markers for its early detection and monitoring. In this work, it is present an analysis of textural radiomics features extracted from a gray matter probability volume, in a set of individual subcortical regions, from a number of different atlases, to identify subject with AD in a MRI. Also, significant subcortical regions for AD detection have been identified using a ReliefF relevance test. Experimental results using the ADNI1 database have proven the potential of some of the tested radiomic features as possible biomarkers for AD/CN differentiation.**

***Index Terms*—Alzheimer's disease, Radiomics, Support vector machines, Magnetic resonance imaging**

I. Introduction

Alzheimer's disease (AD) is a disease characterized by an accumulation of beta-amyloid plaques and neurofibrillary tangles that leads to a progressive deterioration of cognitive and behavioral functions [1]. According to Alzheimer's Disease International (ADI) in its 2016 World Alzheimer's Report, only 50% of people with dementia are being diagnosed, a figure that drops to 10% in less developed countries.

To this day, the diagnosis of Alzheimer's disease remains essentially clinical, meaning that it cannot be detected until the first symptoms appear, or even later, when the neuropathological damage is already significant. It is estimated that the evolution of this disease normally takes between 20 and 30 years from its pre-clinical stage, until recognizable and conclusive symptoms are presented for the diagnosis of Alzheimer's disease [2]. At an intermediate stage in this evolution, the first cognitive symptoms appear, in which subjects show only a slight deterioration in memory, but do not meet the criteria for dementia [3]. This stage is called Mild Cognitive Impairment (MCI) and, although not all patients with MCI develop AD, studies show that between 10% and 15% of MCI cases progress to AD per year [4]. Therefore, detection and monitoring of Alzheimer's disease from possible biomarkers has become a necessity.

Among the set of possible biomarkers for Alzheimer's disease, those based on the analysis of the different modalities of medical imaging of the brain have demonstrated a considerable potential. Medical imaging offers the ability to visualize degenerative histological and methabolical changes, which occur long before the neurodegenerative disorder is clinically detectable [5]. The commonly used imaging modalities in dementia diagnosis include the magnetic resonance imaging (MRI), positron emission tomography (PET), and single-photon emission computed tomography (SPECT). As a result, the use of medical imaging for early diagnosis of AD has grown significantly in the last years, especially the use of MRI, given its non-invasive nature, wide availability and relative absence of discomfort for the patient.

There is a large amount of published research on possible biomarkers for MRI neuroimaging-based computer-aided detection of AD [6]. In general, most of these features are based on variations in density of the whole brain [7] or use subtle changes on thickness extracted at the vertex level on the cortical surface [8]. In the last decade, radiomic approaches [9] to the analysis of medical images have become widespread. The concept of radiomics involves the conversion of digital medical images into high-dimensional minable data, using a large panel of phenotypic features, such as those based on shape, intensity or texture. This characterization process can unveil information in the biomedical image associated with possible underlying pathophysiological anomalies. Although radiomic techniques are primarily used in oncological studies, several types of radiomic textural features have demonstrated their relevance in the detection and quantification of Alzheimer's disease, as can be seen in [10] and [11].

Taking into account prior knowledge of the magnitude and spatial pattern of the evolution of AD, it is possible to focus studies on areas of the brain that have been shown to contain discriminatory information related to AD (i.e. [12]). The main

2372-9198/19/$31.00 ©2019 IEEE
DOI 10.1109/CBMS.2019.00084

approach in these methods implies the use of biomarkers extracted from the hippocampus [13] as volume, shape or textural features. Work on other cerebral areas has been done in [14], in which a diffeomorphometry study has been carried out in a number of regions, including the right and left hippocampus, thalamus, and lateral ventricles, in order to perform a linear discriminant analysis for AD prediction.

The generation of biomarkers based on specific areas of the brain relies on the partitioning of the original MRI volumes in a set of cortical regions using an existing atlas template. However, despite the existence of multiple atlases of the brain, both anatomical and functional, there is no accepted standard for the partitioning of the cortex and subcortical structures, or for the assignment of labels to the resulting regions [15]. There is a considerable lack of agreement among the available parceling schemes [16], which represents a considerable problem when it comes to selecting regions of interest in the study of discriminatory features extracted from specific cerebral regions.

This article deals with the analysis of textural radiomic features on a set of individual subcortical regions, selected for their relevance in distinguishing subjects with AD from healthy subjects. Given the existing discrepancies between atlases, a comparison will be made between the results obtained from different brain parcelations. So far, radiomic studies of Alzheimer's disease have focused on extracting features from the MRI volume intensity values. However, the inherent variability of brain tissues [17] and the various textural patterns resulting from the model and configuration of the MRI hardware can significantly affect textural analyses based on the intensities of the MRI volume [18].

As can be seen in [19] the result of a brain tissue segmentation will tend to show poorly classified areas in regions of the brain where the neuronal atrophy associated with AD tends to manifest. This especially affects areas of grey matter, where such atrophy is particularly apparent [20]. Much in the same way as can be seen in [21], we will take advantage of this fact. Therefore in this work, a volume that represents the probability that, in a MRI, a voxel corresponds to grey matter is used as a base for features extraction.

II. Materials and Methods

In this work, we present an analysis of textural radiomics features in a set of individual subcortical regions, from a number of different atlases, selected for their relevance in classifying subjects with Alzheimer's disease from healthy subjects. At the same time, a significance test has been carried out to identify which areas in relation to which atlases and features show significant differences between AD and CN subjects. Unlike related studies, a volume representing the probability that each voxel of an MRI volume corresponds to grey matter will be used as a base for feature extraction.

The analysis process presented is divided into three stages. In the first stage, the MRIs are processed in order to carry out bias correction and spatial normalization of the data set, using the Clinica software platform (version 0.1.0). Clinica is a software platform developed by the ARAMIS lab[1], for research studies in clinical neurosciences, specialized in multimodal data (neuroimaging, clinical and cognitive evaluations, genetics, etc.). Also, at this stage, a segmentation of the MRIs representing the probability of being white matter, grey matter and cerebrospinal fluid is obtained. In the second stage, from the volume of grey matter, a set of textural features are extracted from the areas of interest. Finally, the resulting vectors for each feature and atlas are used to train classification models based on SVM. These vectors are tested to check the relevance of each of the areas of interest selected in each atlas by using a ReliefF test [22].

The subjects included in this study were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI).

A. ADNI dataset

Launched in 2003 as a public-private partnership and led by Principal Investigator Michael W. Weiner, MD, the Alzheimer's Disease Neuroimaging Initiative[2] is a project aimed at testing whether different modalities of medical imaging such as MRIs and PETs, image based biomarkers, and clinical and neuropsychological evaluation data can be combined to assess the progression of MCI and early AD.

The population of this study during the its three main phases (ADNI1, ADNI GO and ADNI2) consists of 1650 subjects, with a total of 3193 magnetic resonance images, of which 350 are control subjects (CN) 900 are MCI patients and the remaining 400 subjects are diagnosed with AD.

B. MRI volume pre-processing

In the first stage, the dataset is processed in order to carry out bias correction and spatial normalization, and to generate a white matter/grey matter/cerebrospinal fluid segmentation.

To work with a standardized preprocessing workflow, compatible with multiple neuroimaging databases, volume preprocessing and overall data set management is performed using Clinica software platform. Clinica is developed with Python and is designed, using Nipype, as a modular architecture. Before beginning the pre-processing process, Clinica performs the conversion of the dataset to the BIDS format [23].

The Clinica software pre-processing pipeline involves the SPM [24] segmentation procedure that performs all preprocessing processes simultaneously, in a procedure known as "Unified Segmentation" [25]. The pipeline then calculates a group template by applying DARTEL (Diffeomorphic Anatomical Registration Through Exponentiated Lie. Algebra [26]) DARTEL is based on the idea of producing a bidirectional "flow field" as the nucleus for the "deformation" of the image in the image registration process. An example of a pre-processed, segmented volume slice can be seen in 1.

Finally, the volumes are normalized using the space defined by the Montreal Neurological Institute (MNI) template. Clinica also offers a modular way to perform a classification based on automatic learning by combining different inputs, algorithms

[1] www.clinica.run/www.aramislab.fr
[2] adni.loni.usc.edu

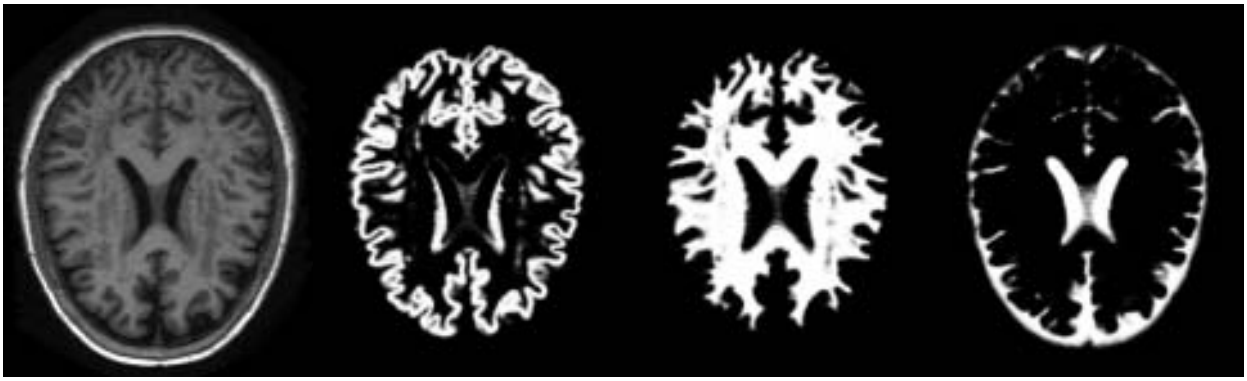


Fig. 1: Example of tissue segmentation, from a volume part of the pre-processed ADNI dataset. Original MRI slice (left), grey matter probabilities (center left), white matter probabilities (center right), cerebrospinal fluid (right)

and validation strategies. These modules are based on scikit-learn [27].

The Clinica pre-processing pipeline has been modified to generate the normalized atlases for each subject, in the form of a volumetric image labelled with the different regions of the parcelation.

C. Feature extraction

Taking into account the particularities of the volume of grey matter probability, composed mainly of homogeneous zones where tissue degeneration is presented as degradation in intensity, 10 textural features have been selected. As general textural features, from the grey level co-occurrence matrix (GLCM), autocorrelation, cluster tendency, correlation, Sum Average, Sum Entropy and Sum of Squares or Variance are retrieved, from each region of interest. As specific textural features for the evaluation of the homogeneity if the distribution of intensities, using the grey-level size zone matrix (GLSZM), the Small Area Low Grey Level Emphasis (SALGLE), the Small Area High Grey Level Emphasis (SAHGLE), the Large Area Low Grey Level Emphasis (LALGLE) and the Large Area High Grey Level Emphasis (LAHGLE) are retrieved, from each region of interest. An in-depth definition of these features can be found in [28] and [29].

These features are extracted using the PyRadiomics [30] platform. PyRadiomics provides a flexible analysis platform with a simple and convenient back-end interface that allows automation in data processing, feature definition and batch management. PyRadiomics is also implemented in Python.

From the results presented in [5] and [12], 9 subcortical regions (for right and left hemispheres) have been chosen as regions of interest: the hippocampus, the parahippocampal gyri, the amygdala, the middle temporal gyri, the superior temporal gyri, the lateral orbital gyri, the medial orbital gyri, the cingulate gyri and the precuneus.

Four anatomical atlases and one functional atlas will be tested. Although these atlases cover the entire cortex and the main subcortical structures, it is possible that some parcelations may not be present in some atlases, or may appear over-parceled. As a result, the size of the vectors of extracted features changes according to the atlas.

- AAL2 [31]: Anatomical atlas manually created from volumes of a single subject. AAL2 is composed of 120 anatomical regions. In this atlas, it has not been possible to identify the equivalents to the amygdala, the lateral orbital Gyrus or the medial orbital Gyrus, so the feature vectors for this atlas represent 12 partitions.
- Hammers [32]: Anatomical atlas created manually from MRI volumes of 30 healthy subjects. This atlas is composed of 69 regions. The feature vectors produced with this atlas represent a total of 18 partitions.
- LPBA40 [33]: Anatomical atlas created manually from volumes of 40 subjects previously transformed into MNI space. LPBA40 consists of 56 regions. It has not been possible to identify in this atlas the equivalents to the amygdala or the medial orbital gyrus, so the feature vectors for this atlas represent 14 partitions.
- Neuromorphometrics [3]: Anatomical atlas created manually from MRI volumes of 30 healthy subjects, previously transformed into MNI space. 140 regions have been parcelated in this atlas. The feature vectors generated with this atlas represent a total of 18 partitions.
- AICHA [34]: Functional atlas created using functional magnetic resonance images from 281 subjects. AICHA represents 345 regions. Many of the regions of interest in this atlas appear over-partitioned, resulting in feature vectors representing 78 partitions.

D. Classification and significance analysis

The regions of interest selected for this study have shown their importance in the evolution of Alzheimer's disease. However, taking into account the differences in shape and position of regions of interest between atlases, and the different

[3]Neuromorphometrics Inc, Building a Model of the Living Human Brain. http://www.neuromorphometrics.com/

stages of the evolution of AD, it is not possible to guarantee the relevance of a specific region for a specific feature on a specific atlas. In this work, the relevance of the association between feature and region of interest is tested using the ReliefF algorithm [22]. The ReliefF algorithm generates a vector of weights for each feature in order to find the possible contribution of a feature to a classification, by finding neighbors from the same class (near-hit) and from the other class (near-miss) using the L1 norm. ReliefF algorithm do not assume conditional independency of the attributes, takes into account contextual information and is able to work with problems with strong dependencies between attributes [35].

To perform the categorization of MRI volumes, in order to separate Alzheimer's disease patients from normal cognitive patients (CN), a Support Vector Machine (SVM) [36] classifier has been selected to generate the classification model. The use of SVMs generally yields reliable models, robust against biases or problems of variance in the data [37]. A linear kernel with a cost parameter of 1 is used.

The models are constructed and examined using ten-fold cross-validation. As performance measures, accuracy, negative prediction value, positive prediction value, sensitivity, specificity, balanced accuracy and F-score are reported. A detailed description of this measures can be found in [38]. Models generation and testing is performed using WEKA; relevance heatmaps and significance tests are carried on using MATLAB.

III. Results

The experiments were carried out using the healthy subjects and subjects diagnosed with Alzheimer's disease of the ADNI1 phase of the ADNI dataset. After converting these ADNI1 data to the BIDS standard, using the Clinica Software's conversion pipeline, the population is reduced to 72 subjects, 44 CN and 28 AD. This reduction may be due to multiple causes, from corrupted images to duplicate or unrelated metadata in the dataset.

Fig. 2 shows the heatmaps of subcortical brain region relevance for each atlas and radiomic textural feature, where each square represents the weight assigned to this specific feature/subcortical region combination by ReliefF. In order to maintain consistency between the different heatmaps, the regions absent in some atlases (AAL2 and LPBA40) are still shown (colored in grey) and, in the atlases with over-partitioned regions (AICHA), it is displayed the mean of the weights assigned by refiefF to each region. As can be seen, among the regions, the hippocampus and the amygdala (where it is present in an anatomical atlas) are the most dominant regions for most radiomic features. This result is consistent with previous studies [10]. To a lesser extent, the temporal, middle and parahippocampal gyri are also noteworthy, especially in the case of the features of correlation, sum of averages, and LAHGLE, although there is no full agreement between the different atlases. This suggests that there are substantial differences in how these regions have been defined.

The results of the classification experiments are summarized in Table I. As a baseline, provided by Clinica, we show the results obtained with a model generated from the mean intensity of each region of the complete parcelation of the brain, for each atlas. Among the set of tested features, the features LAHGLE (for atlases AAL2, Hammers, LPBA40 and Neuromorphometrics), autocorrelation and sum averages (for atlases AAL2, Hammers and Neuromorphometrics) show significant improvements for most of the proposed evaluation metrics (McNemar test $p < 0.05$), compared to the baseline classification model. Although these features also show improvements in relation to the baseline model in the rest of the atlases, it is not possible to claim significance for these results using the McNemar test. On the other hand, features as SALGLE, SAHGLE, LALGLE and correlation result in significantly worse classifiers models than those obtained with the baseline features (McNemar test $p < 0.05$) for many of the tested atlases. The rest of the features (sum entropy, sum squares and cluster tendency) generate models with similar or worse results than the baseline model, although again is not possible to claim significance.

Tested textural features show consistent behavior between atlases; features that produce good results in one atlas produce good results in others, and vice versa (with the exception of SHAGLE). As can be seen from the analysis of the performance measures, the classification errors have a bias towards false negatives, for almost all features and atlases. In general, the existence of relevant regions associated with a feature indicates a good performance of the model generated with it, although it is not a guarantee, as can be seen, for example, for the feature "correlation".

IV. Conclusions

The results presented in this work show the potential of some of the tested radiomic features (i.e LAHGLE) as possible biomarkers in the detection of AD. As in previous studies, it was found that the hippocampus and the amygdala are the most dominant regions for MRI AD/CN differentiation. Finally this works suggest that, taking into account only the overall performance measures for the different features, both Neuromorphometrics and Hammers have a slight advantage over the rest as the best suited atlases for the Alzheimer's disease detection, although performance differences are, in general, small.

The results obtained for some of the tested features suggest the extension of the study to other cases, such as discrimination between normal and mild cognitive impairment, as well as the extension of the group of study subjects using other databases. Also, the possibility of narrowing the set of used subcortical regions is considered, as well as to improve the discrimination capacity of the models using multi-features vectors.

V. ACKNOWLEDGMENT

This paper is supported by European Unions Horizon 2020 research and innovation programme under grant agreement No. 727658, project IASIS (Integration and analysis of heterogeneous big data for precision medicine and suggested treatments for different types of patients)

REFERENCES

[1] A. Serrano-Pozo, M. P. Frosch, E. Masliah, and B. T. Hyman, "Neuropathological alterations in alzheimer disease," *Cold Spring Harbor perspectives in medicine*, vol. 1, no. 1, p. a006189, 2011.

[2] M. Weiner, D. Veitch, P. Aisen, and C. N. Beckett lA, "green rc," *Harvey D, Jack Cr, Jagust W, liu e, Morris JC, Petersen rC, Saykin AJ, Schmidt Me, Shaw l, Siuciak JA, Soares H, Toga AW, Trojanowski JQ*, 2012.

[3] B. Dubois and M. L. Albert, "Amnestic mci or prodromal alzheimer's disease?" *The Lancet Neurology*, vol. 3, no. 4, pp. 246–248, 2004.

[4] R. C. Petersen, R. O. Roberts, D. S. Knopman, B. F. Boeve, Y. E. Geda, R. J. Ivnik, G. E. Smith, and C. R. Jack, "Mild cognitive impairment: ten years later," *Archives of neurology*, vol. 66, no. 12, pp. 1447–1455, 2009.

[5] K. A. Johnson, N. C. Fox, R. A. Sperling, and W. E. Klunk, "Brain imaging in alzheimer disease," *Cold Spring Harbor perspectives in medicine*, p. a006213, 2012.

[6] S. Rathore, M. Habes, M. A. Iftikhar, A. Shacklett, and C. Davatzikos, "A review on neuroimaging-based classification studies and associated feature extraction methods for alzheimer's disease and its prodromal stages," *NeuroImage*, vol. 155, pp. 530–548, 2017.

[7] S. Klöppel, C. M. Stonnington, C. Chu, B. Draganski, R. I. Scahill, J. D. Rohrer, N. C. Fox, C. R. Jack Jr, J. Ashburner, and R. S. Frackowiak, "Automatic classification of mr scans in alzheimer's disease," *Brain*, vol. 131, no. 3, pp. 681–689, 2008.

[8] S. Li, X. Yuan, F. Pu, D. Li, Y. Fan, L. Wu, W. Chao, N. Chen, Y. He, and Y. Han, "Abnormal changes of multidimensional surface features using multivariate pattern classification in amnestic mild cognitive impairment patients," *Journal of Neuroscience*, vol. 34, no. 32, pp. 10 541–10 553, 2014.

[9] V. Kumar, Y. Gu, S. Basu, A. Berglund, S. A. Eschrich, M. B. Schabath, K. Forster, H. J. Aerts, A. Dekker, D. Fenstermacher *et al.*, "Radiomics: the process and the challenges," *Magnetic resonance imaging*, vol. 30, no. 9, pp. 1234–1248, 2012.

[10] A. Chaddad, C. Desrosiers, and T. Niazi, "Deep radiomic analysis of mri related to alzheimers disease," *IEEE Access*, vol. 6, pp. 58 213–58 221, 2018.

[11] F. Feng, P. Wang, K. Zhao, B. Zhou, H. Yao, Q. Meng, L. Wang, Z. Zhang, Y. Ding, L. Wang *et al.*, "Radiomic features of hippocampal subregions in alzheimers disease and amnestic mild cognitive impairment," *Frontiers in aging neuroscience*, vol. 10, 2018.

[12] L. K. McEvoy, C. Fennema-Notestine, J. C. Roddey, D. J. Hagler Jr, D. Holland, D. S. Karow, C. J. Pung, J. B. Brewer, and A. M. Dale, "Alzheimer disease: quantitative structural neuroimaging for detection and prediction of clinical and structural changes in mild cognitive impairment," *Radiology*, vol. 251, no. 1, pp. 195–205, 2009.

[13] L. Sørensen, C. Igel, N. Liv Hansen, M. Osler, M. Lauritzen, E. Rostrup, M. Nielsen, A. D. N. Initiative, the Australian Imaging Biomarkers, and L. F. S. of Ageing, "Early detection of alzheimer's disease using m ri hippocampal texture," *Human brain mapping*, vol. 37, no. 3, pp. 1148–1161, 2016.

[14] X. Tang, D. Holland, A. M. Dale, L. Younes, M. I. Miller, A. D. N. Initiative *et al.*, "Baseline shape diffeomorphometry patterns of subcortical and ventricular structures in predicting conversion of mild cognitive impairment to alzheimer's disease," *Journal of Alzheimer's Disease*, vol. 44, no. 2, pp. 599–611, 2015.

[15] A. C. Evans, A. L. Janke, D. L. Collins, and S. Baillet, "Brain templates and atlases," *Neuroimage*, vol. 62, no. 2, pp. 911–922, 2012.

[16] J. W. Bohland, H. Bokil, C. B. Allen, and P. P. Mitra, "The brain atlas concordance problem: quantitative comparison of anatomical parcellations," *PloS one*, vol. 4, no. 9, p. e7200, 2009.

[17] F. Kruggel, J. S. Paul, and H.-J. Gertz, "Texture-based segmentation of diffuse lesions of the brains white matter," *NeuroImage*, vol. 39, no. 3, pp. 987–996, 2008.

[18] L. R. Schad, "Problems in texture analysis with magnetic resonance imaging," *Dialogues in clinical neuroscience*, vol. 6, no. 2, p. 235, 2004.

[19] D. V. Callaert, A. Ribbens, F. Maes, S. P. Swinnen, and N. Wenderoth, "Assessing age-related gray matter decline with voxel-based morphometry depends significantly on segmentation and normalization procedures," *Frontiers in aging neuroscience*, vol. 6, p. 124, 2014.

[20] P. M. Thompson, K. M. Hayashi, G. De Zubicaray, A. L. Janke, S. E. Rose, J. Semple, D. Herman, M. S. Hong, S. S. Dittmer, D. M. Doddrell *et al.*, "Dynamics of gray matter loss in alzheimer's disease," *Journal of neuroscience*, vol. 23, no. 3, pp. 994–1005, 2003.

[21] C. Toro, C. Gonzalo-Martín, A. García-Pedrero, and E. Menasalvas Ruiz, "Supervoxels-based histon as a new alzheimers disease imaging biomarker," *Sensors*, vol. 18, no. 6, p. 1752, 2018.

[22] I. Kononenko, E. Šimec, and M. Robnik-Šikonja, "Overcoming the myopia of inductive learning algorithms with relieff," *Applied Intelligence*, vol. 7, no. 1, pp. 39–55, 1997.

[23] K. J. Gorgolewski, T. Auer, V. D. Calhoun, R. C. Craddock, S. Das, E. P. Duff, G. Flandin, S. S. Ghosh, T. Glatard, Y. O. Halchenko *et al.*, "The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments," *Scientific Data*, vol. 3, p. 160044, 2016.

[24] W. D. Penny, K. J. Friston, J. T. Ashburner, S. J. Kiebel, and T. E. Nichols, *Statistical parametric mapping: the analysis of functional brain images*. Elsevier, 2011.

[25] J. Ashburner and K. J. Friston, "Unified segmentation," *Neuroimage*, vol. 26, no. 3, pp. 839–851, 2005.

[26] M. Goto, O. Abe, S. Aoki, N. Hayashi, T. Miyati, H. Takao, T. Iwatsubo, F. Yamashita, H. Matsuda, H. Mori *et al.*, "Diffeomorphic anatomical registration through exponentiated lie algebra provides reduced effect of scanner for cortex volumetry with atlas-based method in healthy subjects," *Neuroradiology*, vol. 55, no. 7, pp. 869–875, 2013.

[27] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[28] R. M. Haralick, K. Shanmugam *et al.*, "Textural features for image classification," *IEEE Transactions on systems, man, and cybernetics*, no. 6, pp. 610–621, 1973.

[29] G. Thibault, B. Fertil, C. Navarro, S. Pereira, P. Cau, N. Lévy, J. Sequeira, and J.-L. Mari, "Texture indexes and gray level size zone matrix. application to cell nuclei classification," in *10th International Conference on Pattern Recognition and Information Processing, PRIP 2009*, 2009, pp. 140–145.

[30] J. J. van Griethuysen, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, R. G. Beets-Tan, J.-C. Fillion-Robin, S. Pieper, and H. J. Aerts, "Computational radiomics system to decode the radiographic phenotype," *Cancer research*, vol. 77, no. 21, pp. e104–e107, 2017.

[31] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, "Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain," *Neuroimage*, vol. 15, no. 1, pp. 273–289, 2002.

[32] A. Hammers, R. Allom, M. J. Koepp, S. L. Free, R. Myers, L. Lemieux, T. N. Mitchell, D. J. Brooks, and J. S. Duncan, "Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe," *Human brain mapping*, vol. 19, no. 4, pp. 224–247, 2003.

[33] D. W. Shattuck, M. Mirza, V. Adisetiyo, C. Hojatkashani, G. Salamon, K. L. Narr, R. A. Poldrack, R. M. Bilder, and A. W. Toga, "Construction of a 3d probabilistic atlas of human cortical structures," *Neuroimage*, vol. 39, no. 3, pp. 1064–1080, 2008.

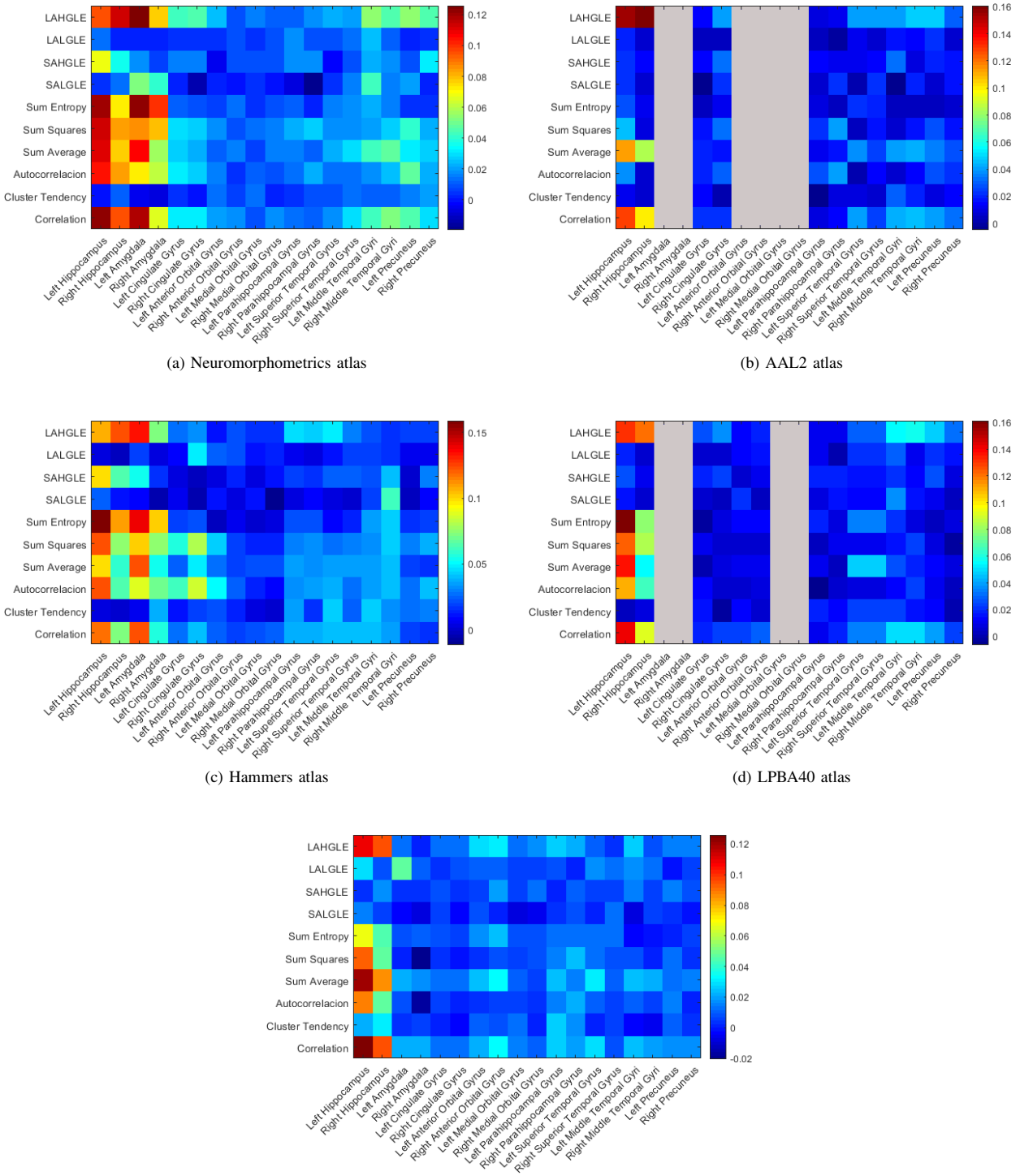
[34] M. Joliot, G. Jobard, M. Naveau, N. Delcroix, L. Petit, L. Zago, F. Crivello, E. Mellet, B. Mazoyer, and N. Tzourio-Mazoyer, "Aicha: an atlas of intrinsic connectivity of homotopic areas," *Journal of neuroscience methods*, vol. 254, pp. 46–59, 2015.

[35] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of relieff and rrelieff," *Machine learning*, vol. 53, no. 1-2, pp. 23–69, 2003.

[36] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural processing letters*, vol. 9, no. 3, pp. 293–300, 1999.

[37] D. Meyer, F. Leisch, and K. Hornik, "The support vector machine under test," *Neurocomputing*, vol. 55, no. 1-2, pp. 169–186, 2003.

[38] D. M. Powers, "Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation," 2011.



(a) Neuromorphometrics atlas

(b) AAL2 atlas

(c) Hammers atlas

(d) LPBA40 atlas

(e) AICHA atlas

Fig. 2: Heatmaps showing the relevance of the subcortical brain regions in relation to the radiomics textural features, for each atlas and radiomic textural feature. Each square represents the weight assigned to this specific feature/subcortical region combination by ReliefF. Regions absent are shown coloured in grey.

TABLE I: Classification results showing a comparison between the evaluation metrics proposed for the different atlas and radiomics textural features. The best results for each atlas are highlighted in bold.

Feature	Accuracy	Balance accuracy	NPV	PPV	Sensitivity	Specificity	F-measure
AAL2							
Mean (Clinica)	0.7909	0.7803	0.8165	0.7616	0.7222	0.8385	0.7418
Autocorrelation	0.8611	0.8579	0.9091	0.7857	0.8462	0.8696	0.8148
Cluster Tendency	0.7361	0.7230	0.8182	0.6071	0.6800	0.7660	0.6415
Correlation	0.6806	0.7222	0.9545	0.2500	0.7778	0.6667	0.3784
Sum Average	0.8611	0.8646	0.9318	0.7500	0.8750	0.8542	0.8077
Sum Entropy	0.7083	0.7321	0.9318	0.3517	0.7692	0.6949	0.4878
Sum Squares	0.7639	0.7519	0.8182	0.6786	0.7037	0.8000	0.6909
SALGLE	0.6944	0.6852	0.8636	0.4286	0.6667	0.7037	0.5217
SAHGLE	0.6806	0.6648	0.8409	0.4286	0.6316	0.6981	0.5106
LALGLE	0.7083	0.7857	**0.9773**	0.2857	0.8889	0.6825	0.4324
LAHGLE	**0.93060**	**0.9296**	0.9545	**0.8929**	**0.9259**	**0.9333**	**0.9091**
AICHA							
Mean (Clinica)	0.8385	0.7927	0.8054	**0.8415**	0.6778	**0.9077**	0.7508
Autocorrelation	0.8333	0.8247	0.8636	0.7857	0.7857	0.8636	0.7857
Cluster Tendency	0.7778	0.7708	0.8636	0.6429	0.7500	0.7917	0.6923
Correlation	0.7361	0.7222	0.7955	0.6429	0.6667	0.7778	0.6545
Sum Average	0.8472	0.8407	0.8864	0.7857	0.8148	0.8667	0.8000
Sum Entropy	0.7778	0.7708	0.8636	0.6429	0.7500	0.7917	0.6923
Sum Squares	0.8056	0.8091	0.9091	0.6429	0.8182	0.8000	0.7200
SALGLE	0.6389	0.6171	0.7273	0.5000	0.5385	0.6957	0.5185
SAHGLE	0.7639	0.7518	0.7955	0.7143	0.6897	0.8140	0.7018
LALGLE	0.6528	0.6305	0.8409	0.3571	0.5882	0.6727	0.4444
LAHGLE	**0.8611**	**0.8539**	**0.8864**	0.8214	**0.8214**	0.8864	**0.8214**
Hammers							
Mean (Clinica)	0.7909	0.7940	0.8652	0.7193	0.8111	0.7769	0.7625
Autocorrelation	0.9028	**0.9000**	**0.9320**	0.8571	**0.8889**	0.9111	0.8727
Cluster Tendency	0.8472	0.8455	0.9091	0.7500	0.8400	0.8511	0.7925
Correlation	0.7639	0.7536	0.8409	0.6429	0.7200	0.7872	0.6792
Sum Average	0.8889	0.8880	0.9318	0.8214	0.8846	0.8913	0.8519
Sum Entropy	0.8611	0.8539	0.8864	0.8214	0.8214	0.8864	0.8214
Sum Squares	0.8611	0.8646	0.9318	0.7500	0.8750	0.8542	0.8077
SALGLE	0.7361	0.7363	0.8864	0.5000	0.7368	0.7358	0.5957
SAHGLE	0.8194	0.8111	0.8636	0.7500	0.7778	0.8444	0.7636
LALGLE	0.7083	0.7321	0.9318	0.3571	0.7692	0.6949	0.4878
LAHGLE	**0.9028**	0.8962	0.9091	**0.8929**	0.8621	**0.9323**	**0.8772**
LPBA40							
Mean (Clinica)	0.8545	0.8513	0.8856	0.8463	0.8333	0.8692	0.8398
Autocorrelation	0.9028	0.9068	0.9545	0.8214	0.9200	0.8936	0.8679
Cluster Tendency	0.8056	0.8021	0.8864	0.6786	0.7917	0.8125	0.7308
Correlation	0.7500	0.7723	0.9318	0.4643	0.8125	0.7321	0.5909
sum Average	0.8750	0.8851	0.9545	0.7500	0.9130	0.8571	0.8235
sum Entropy	0.8472	0.8407	0.8864	0.7857	0.8148	0.8667	0.8000
sum Squares	0.8056	0.8021	0.8864	0.6786	0.7917	0.8125	0.7308
SALGLE	0.6944	0.7167	0.9318	0.3214	0.7500	0.6833	0.4500
SAHGLE	0.7222	0.7222	0.8864	0.4643	0.7222	0.7222	0.5652
LALGLE	0.7222	0.7667	0.9545	0.3571	0.8333	0.7000	0.5000
LAHGLE	**0.9444**	**0.9416**	**0.9555**	**0.9286**	**0.9286**	**0.9545**	**0.9286**
Neuromorphometrics							
Mean (Clinica)	0.8136	0.8064	0.8510	0.8016	0.7667	0.8462	0.7838
Autocorrelation	0.8889	0.8880	**0.9330**	0.8214	0.8846	0.8913	0.8519
Cluster Tendency	0.7917	0.7815	0.8409	0.7143	0.7407	0.8222	0.7273
Correlation	0.7361	0.7596	0.9318	0.4286	0.8000	0.7193	0.5581
Sum Average	0.8611	0.8646	0.9318	0.7500	0.8750	0.8542	0.8077
Sum Entropy	0.8611	0.8579	0.9091	0.7857	0.8462	0.8696	0.8148
Sum Squares	0.8194	0.8095	0.8409	0.7857	0.7586	0.8605	0.7719
SALGLE	0.7361	0.7230	0.8182	0.6071	0.6800	0.7660	0.6415
SAHGLE	0.8333	0.8333	0.9091	0.7143	0.8333	0.8333	0.7692
LALGLE	0.6667	0.6576	0.8864	0.3214	0.6429	0.6724	0.4286
LAHGLE	**0.9028**	**0.9000**	0.9318	**0.8571**	**0.8889**	**0.9111**	**0.8727**