Extracting Explainable Assessments of Alzheimer's disease via Machine Learning on brain MRI imaging data

K.G. Achilleos[1], S. Leandrou[2,3], N. Prentzas[1], P.A. Kyriacou[3], A.C. Kakas[1], C.S. Pattichis[1,4]
for the Alzheimer's Disease Neuroimaging Initiative*

[1] University of Cyprus, Department of Computer Science, Nicosia, Cyprus
[2] European University Cyprus/Department of Health Sciences, Nicosia, Cyprus
[3]City University London, School of Mathematical Sciences, Computer Science and Engineering
[4]Research Centre on Interactive Media Smart Systems and Emerging Technologies (RISE), Nicosia, Cyprus

Abstract— **A plethora of machine learning and deep learning methods are used for the assessment of Alzheimer's Disease (AD) from brain structural changes as seen in Magnetic Resonance Imaging (MRI) with highly satisfactory results. However, these models are black-box and lack an explicit declarative knowledge representation and thus there is a difficulty in generating the underlying explanatory imaging structures. The objective of this study was to investigate the usefulness of rule extraction in the assessment of AD using decision trees (DT) and random forests (RF) algorithms and integrating the extracted rules within an argumentation-based reasoning framework in order to make the results easy to interpret and explain. The DT and RF algorithms were applied on brain MRI images acquired from normal controls (NC) and AD subjects. The KNIME analytics platform was used to compute the DT and the R project was used for the RF. The argumentation model implemented in the Gorgias framework achieved an average accuracy of 91%, exhibiting improved results compared to the models of DT and RF. The overall performance of all models in this study is in agreement with other studies. In addition, the explanations given by our approach for the various possible predictions provide a more useful and complete assessment of the state of the patient/case at hand. This study demonstrated the usefulness of rule extraction in the assessment of AD based on MRI features and the positive results of the use of the argumentation based symbolic reasoning for composing and interpreting the ML results.**

Keywords— **Alzheimer's disease, decision trees, random forests, quantitative MRI, Explainable AI, Argumentation.**

I. Introduction

Alzheimer's disease (AD) is a progressive neurodegenerative disorder which affects the elderly. AD is one of the most frequent causes of dementia and in a 2019 estimate the number of people living with dementia globally was over 50 million [1]. AD causes impairment of memory and other mental functions i.e. language. At present, there is no cure and medications available do not treat the underlying causes of the disease or slow its progression [2].

The problem of AD clinical diagnosis is up to this day considered difficult [3]. Clinical diagnosis relies on cognitive tests that most of the time are affected by external factors or patient's psychology, therefore, they are not objective. These tests are Mini Mental State Exam (MMSE) scores [4] and Clinical Dementia Rating (CDR) scores [5]. Magnetic Resonance Imaging (MRI) is also used to detect brain structural changes derived from neurodegeneration and its importance in the assessment of AD was underlined by its inclusion in the new diagnostic criteria [6].

The hippocampus is one of the first structures in the brain affected by AD and the anatomical neural changes can be measured in the terms of hippocampal volume and shape using MRI. The hippocampus represents the most valid and most used region of interest (ROI) in the classification and prediction of AD [7]–[9].

The challenge for modern neuroimaging is to provide early diagnosis of AD since currently diagnosis is established after there is progressive and irreversible neurodegeneration. Even though hippocampal atrophy is the most established structural AD imaging biomarker to date, texture analysis has the advantage of detecting earlier, microscopic alterations. Consequently, in this study the effectiveness of hippocampal texture is combined with hippocampal volume in the classification of normal control (NC) from AD subjects using decision trees (DT) and random forests (RF).

Explainability of a prediction has become a crucial provision for Artificial Intelligence (AI) [10]. Classification models and especially medical diagnosis systems need to be transparent and explainable [11] to earn the trust of end users. "If the users do not trust a model or a prediction they will not use it" [12]. For the problem of diagnosing AD from MRI images there is previous work using different AI models like neural networks, RF and DT. In general, these approaches are not primarily concerned with Explainable Learning: their methods are not driven or influenced by the need to provide explanations supporting their diagnostic predictions.

In this paper, argumentation is used as the underlying framework for learning in order to study the problem of diagnosing Alzheimer's disease from MRI images. Existing Machine Learning (ML) methods are utilized for learning and extracting rules from imaging data. In the argumentation framework rules are viewed as arguments supporting a diagnosis rather than strict rules imposing a diagnosis. With a further learning process for learning preferences amongst these argument rules, a learned argumentation theory is reached where diagnostic predictions are at least as accurate as these achieved by other ML methods but also explainable.

*Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

2471-7819/20/$31.00 ©2020 IEEE
DOI 10.1109/BIBE50027.2020.00175

Explanations for a diagnostic prediction would be composed by both arguments or reasons supporting the prediction as well as reasons for the relative strength of the prediction in comparison to other opposing predictions.

Furthermore, in difficult cases the learned argumentation theory could be in a prediction dilemma offering explanations for several possible and conflicting predictions. The learned theory is thus viewed more as offering an assessment of the case at hand and the aim is for the explanations in such dilemma cases to help towards a final decision by another external process, e.g. by a human expert, that would be less inconclusive and more accurate.

Argumentation theories can provide a diagnostic model that is easily comprehensible by experts so that it can be useful to them in taking a final decision. Moreover, such argumentation-based models would be easily modified and refined by the experts in a knowledge acquisition methodology that combines ML with knowledge elucidation from expert sources of information.

II. Methodology

A. Material

Data were acquired from the Alzheimer's Disease Neuroimaging Initiative (ADNI) (http://adni.loni.usc.edu/). The ADNI was launched in 2003 by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, the Food and Drug Administration, private pharmaceutical companies and non-profit organizations as a public-private partnership. The goal of the ADNI study is to determine biological biomarkers of AD through neuroimaging, genetics, neuropsychological tests and other measures in order to develop new treatments and monitor their effectiveness and lessen the time of clinical trials.

Enrolled subjects were all between 55 and 90 years of age and each subject was able to perform all test procedures described in the protocol and had a study partner able to provide an independent evaluation of functioning. In total, 237 subjects were included in this study of two distinct groups: NC=153 (73 males and 80 females) and AD=84 (40 males and 44 females). Subjects with outlier values were not removed, as DT and RF are resilient to noisy data [13]. However, all subjects with missing values were dropped. As a result, the final dataset comprised of NC=144 and AD=69. Due to the unbalanced dataset and relatively small number of records, we used a 30% equal distribution test set (NC=22, AD=22) and an unbalanced training set (NC=122, AD=47).

B. MRI data

Brain MRI images downloaded were part of the ADNI-1 Complete 2- and 3-year dataset. All the subjects had a standardized protocol (see http://adni.loni.usc.edu/methods/mri-analysis/adni-standardized-data/) on 1.5-T MRI units from Siemens Medical Solutions and General Electric Healthcare. MR protocols included high-resolution (typically 1.25×1.25×1.25 mm^3 voxels) T1-weighted volumetric 3D sagittal magnetization prepared rapid gradient-echo (MPRAGE) scans. MRI data acquisition techniques were standardized across different sites according to ADNI protocol.

C. Feature extraction and selection

ROI segmentation was performed using the Freesurfer v6.0 software (Massachusetts General Hospital, Boston, MA), which is documented and freely available for download online (http://surfer.nmr.mgh.harvard.edu/). The Freesurfer pipeline has been discussed in previous publications [14], [15]. The Hippocampus Volume was computed using the Freesurfer segmentations.

The Haralick texture features [16] were calculated using a KNIME Analytics platform workflow [17]. The following brain MRI texture features were extracted from the hippocampus: Angular Second Moment, Contrast, Correlation, Sum Average, Entropy, Sum Entropy, Cluster Shade, Variance, Sum Variance, Cluster Shade and Cluster Prominence.

RF models with 5000 trees were built to measure the mean Decrease Gini index in order to identify the most important features. Gini Index [18] is an impurity splitting method. It was observed that feature selection based on the Gini index increases the overall performance of the RF models [19]. The 10 most promising features are shown in Figure 1.

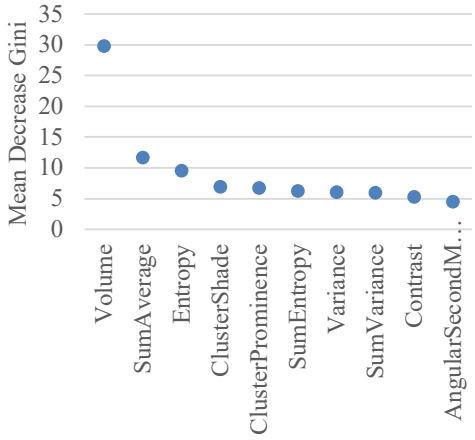

Figure 1: 10 most promising hippocampal features based on mean decrease of the gini index

In a final step aiming to produce more comprehensible rules, we discretized the features in five categories (1-Low and 5-High). Equal frequency binning [20] was selected for the discretization (see Table 1 for Hippocampus Volume)

Table 1: Feature discretization to categories 1-5

Feature	Category	Values (volume in mm^3)	Record count
Hippocampus Volume (hipVolume)	1-Very Low	<2890	43
	2-Low	>=2890 and <3170	42
	3-Nominal	>=3170 and <3530	43
	4-High	>=3530 and <3800	42
	5-Very High	>=3800	43

As noted in the introduction Hippocampus Volume is the most important feature in the classification of NC vs AD. This is visualized in the graphical representation of the categories formed by discretization (see Figure 2). 53% of NC subjects are in the higher categories and 74% AD are in the lower categories. In the middle category (3) there are 22% NC and 17% AD.

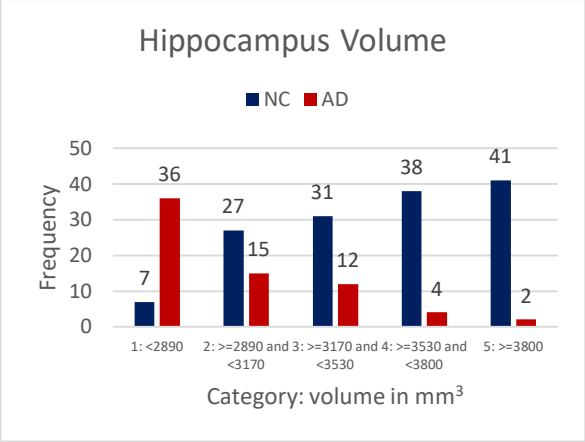

Figure 2: NC/AD subjects Hippocampus Volume categories versus frequency

D. Decision trees and Random Forest models

The discretized data were used by DT and RF ML algorithms to produce a collection of rules.

The KNIME analytics platform was utilized for the configuration and testing of the DT. Ten-fold cross validation DT were trained and tested on the random equal sized datasets created. The DT configuration was set to 10 minimum records per node and reduced error pruning [21]. The KNIME node, "Decision Tree to Ruleset" [22], provided by KNIME core, extracted the rules in text form with the number of rules ranging from 2 to 8 for each tree.

On the other hand, the R project [23] was used to train RF with 10-fold cross validation. One RF model was selected for rule extraction based on best accuracy on the test set. Again, testing was carried out with equal sized testing sets of 22 records per class. The "inTrees" [24] library in R was used to extract 45 rules from the selected RF.

For each rule collected the following classification metrics were used (where P=no of positive cases and N=no of negative cases(N)):

$$Precision = \frac{true\ positive(TP)}{true\ positive(TP) + false\ positive(FP)}$$

$$Sensitivity = \frac{true\ positive(TP)}{no\ of\ positive\ cases(P)}$$

$$Specificity = \frac{true\ negative(TN)}{no\ of\ negative\ cases(N)}$$

$$Accuracy = \frac{true\ positive(TP) + true\ negative(TN)}{P + N}$$

E. Argumentation component

Based on the output of the DT and RF models described above, we apply an argumentation-based learning component to arrive at our final learned model. Gorgias-B authoring tool was used. It is built on top of Gorgias [25], a preference based argumentation framework, where arguments are constructed using a basic argument scheme of Modus Ponens to link a set of premises with the claim [25]. A Gorgias argumentation theory represents the knowledge that describes the application in terms of object level arguments and priority argument rules.

As first step for the design of the model, two options for subject P, healthy(P) and diagnosedAlz(P) were defined. The definitions of the options predicates which are basically the features, was the second step. One example is hipVolume(P,V), where P is the subject and V is value of the feature. The next step was authoring of the arguments (rules). Each argument consisted of a set of options predicates, the variables restrictions and a claim, healthy or diagnosedAlz. For example *rule(r2(P), diagnosedAlz(P), []):-hipVolume(P, V), 1=V* is a rule claiming that if hipVolume=1 then diagnosedAlz(P) holds.

In the final step, the rules priority was set. Arguments for options were defined giving preference to a rule over another, either under conditions or unconditionally. For example, rule(p1(P), prefer(r4(P), r1(P)), []) is an unconditional preference rule stating always prefer r4 over r1.

The procedure of determining the rules to be included in the argumentation theory and their respective priority was carried out manually. The proposed model has a total of eleven arguments.

III. Results

To evaluate the argumentation model, the 10 random test sets produced for the DT with equal class distribution, 22(NC) and 22(AD) were used. The results are shown in Table 2. The model achieved accuracy in the range of 86% to 95%, with an average of 91%, as opposed to the 10 DT and 10 RF models that achieved a max of 86%.

Table 2: Classification Results for models NC versus AD

Classifier	Accuracy	Sensitivity	Specificity
	average on 10 runs		
Decision Trees	77%	66%	88%
Random Forests	74%	56%	91%
Argumentation Rules	91%	87%	95%

A selection of five rules is presented in Table 3, along with classification metrics for each rule. A demonstration test case in the Gorgias framework is given in Table 4.

Table 3: A selection of arguments defined in Gorgias framework

Priority →	Sample Rules from the argumentation model *(in parenthesis: rule no in the model)*	Precision	Sensitivity	Specificity
	(r1) If hipVolume >= 2890mm^3 ***return*** subject in NC	84%	96%	51%
	(r2) If hipVolume < 2890mm^3 ***return*** subject in AD	83%	51%	96%
	(r3) If hipVolume<3170mm^3 and hipClusterPRominence<802000 ***return*** subject in AD	87%	28%	98%
	(r5) If hipEntropy >= 3.67 and hipVAriance >= 198 and hipContrast < 194 ***return*** subject in NC	100%	14%	100%
	(r8) If hipSumVariance < 543 and hipSumEntropy < 2.92 ***return*** subject in NC	100%	17%	100%

Table 4: A demonstration testcase in the Gorgias framework

Radiology Input	patient (p445, 2439.1, 38.13, 3.76, 10067.75, 769877.6, 200.64, 623.58, 2.91, 0.1, 178.98). patient (p445, hipVolume, SumAverage, hipEntropy, hipCLusterShade, hipClusterPRominence, hipVAriance, hipSumVariance, hipSumEntropy, hipAngularSecondMoment, hipContrast).

↓

Gorgias Input	patient(p445,1,2,4,2,1,3,3,4,3,3).
Argument 1	diagnosedAlz(p445)?
Gorgias Output	Found solution: When [hipVolume(p445, V), 1=V] choose diagnosedAlz(p445) When [hipVolume(p445, V), hipClusterPRominence(p445, PR), 1=PR, 3>V] choose diagnosedAlz(p445) When [hipEntropy(p445, E), hipVolume(p445, V), 2<VA, hipContrast(p445, C), hipClusterPRominence(p445, PR), 1=PR, 5>C, 3>V, hipVAriance(p445, VA), 3<E] prefer diagnosedAlz(p445) over healthy(p445)
Explanation	*p445 satisfies two AD rules (r2, r3), one NC rule (r5) and one preference rule p10 (line 4-5 in output).* List of relative arguments defined in Gorgias *rule(r2(P), diagnosedAlz(P), []): -hipVolume(P, V), 1=V.* *rule(r3(P), diagnosedAlz(P), []): -hipVolume(P, V), hipClusterPRominence(P, PR), 3>V, 1=PR.* *rule(r5(P), healthy(P), []): -hipEntropy(P, E), hipVAriance(P, VA), hipContrast(P, C), 3<E, 2<VA, 5>C.* *rule(p10(P), prefer(r3(P), r5(P)), []).*
Argument 2	healthy(p445)?
Output	No solution for this goal.
Explanation	*Although p445 satisfies rule r5, rule r3 has precedence over rule r5 so the argument is not triggered.*

↓

Radiology output	Diagnosis: Alzheimer's (ALZ) Criteria in favour AD: 1. Hippocampal Volume is very Low. 2. Cluster Prominence is very Low in combination with Low Volume. Criteria in favour NC: 1. Entropy is High, Variance is Nominal and Contrast is not Very High The criteria supporting AD are in general stronger than those in favour for NC and therefore the predicted diagnosis is AD.

IV. Discussion

In this research the accuracy of the NC-AD Classification model proposed was significantly higher in comparison to the DT and RF models constructed. Previous work from other researchers on ADNI data and RF classification has analogous results [26]. The accuracy achieved is favorably comparable to results achieved by black box models like neural networks as seen in the work of B.-K. Choi et al [27] and A. Payan and G. Montana [28] among others.

Only MRI data were used as input to the model design process whereas other models also opt to include demographics like age, gene information (Apolipoprotein e4 gene (AP-Oe4)). These added features do not tend to increase RF model's accuracies as discussed in [26]. MMSE test scores, on the other hand vary within the population by age and education and should not be used to make formal diagnoses [29].

The proposed model consists of simple rules structured in a form intelligible by the user. As predictive models frequently have to address the tradeoff between performance accuracy and interpretability, Gorgias framework may be a good compromise combining the two qualities.

Given the above, the rules proposed offer an alternative model where the expert can evaluate the prediction and subsequently accept or reject it.

However, in this work the features offered by the argumentation component are not explored to the maximum. The rules and priorities could be further adjusted to present dilemmas to the user and not true or false predictions. Hence, an assessment of the disease and not a prediction. Additionally, an explanation of the prediction/assessment can be automatically extracted in natural language by using an explanation schema and suitable naming of the rules.

Another hindrance to the development process was the need for access to additional datasets for further validation and refinement of the methods. However, DT and RF may be applied on smaller datasets contrary to the requirement of large datasets by other AI/ ML methods. The limited training set of this study, does reduces the probability of good generalization performance on new datasets, however this is a challenge for all researchers in the neuroimaging field.

V. Concluding remarks

This paper has presented a novel argumentation model for the classification of NC, AD subjects. While research efforts in this domain are ongoing and have produced significant results, our approach addresses the challenge of an explainable and yet significantly accurate model harnessing the capabilities of the argumentation theory. There is a growing interest for explainable models addressing human trust issues. The aim of our research to demonstrate that argumentation can be successfully applied was demonstrated.

Future research plans include the application of the argumentation process to the more difficult problem of NC vs Mild Cognitive Impairment (MCI) classification.

Future work will also further investigate the application of symbolic reasoning on more structures. Recent work suggests that including the entorhinal cortex features can provide better results [30] in the classification of NC from MCI subjects.

Conflict of interest

"The authors declare that they have no conflict of interest".

Funding

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 739578 and the Government of the Republic of Cyprus through the Directorate General for European Programmes, Coordination and Development.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 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.

Acknowledgments

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 is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

REFERENCES

[1] J. Bhatt *et al*., "World Alzheimer Report 2019 Attitudes to dementia," 2019, [Online]. Available: https://www.alz.co.uk/research/WorldAlzheimerReport2019.pdf.

[2] "Alzheimer's Association | Alzheimer's Disease & Dementia Help." https://alz.org/ (accessed Jun. 23, 2020).

[3] Q. Feng and Z. Ding, "MRI Radiomics Classification and Prediction in Alzheimer's Disease and Mild Cognitive Impairment: A Review," *Curr. Alzheimer Res.*, vol. 17, no. 3, pp. 297–309, Mar. 2020, doi: 10.2174/1567205017666200303105016.

[4] A. Kaiser, R. Gusner-Pfeiffer, H. Griessenberger, and B. Iglseder, "Mini-Mental-State," *Zeitschrift für Gerontopsychologie - psychiatrie*, vol. 22, no. 1, 2009, doi: 10.1024/1011-6877.22.1.11.

[5] J. C. Morris, "The clinical dementia rating (cdr): Current version and scoring rules," *Neurology*, vol. 43, no. 11, pp. 2412–2414, 1993, doi: 10.1212/wnl.43.11.2412-a.

[6] B. Dubois *et al*., "Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria," *Lancet Neurology*, vol. 6, no. 8. 2007, doi: 10.1016/S1474-4422(07)70178-3.

[7] L. Ferrarini, G. B. Frisoni, M. Pievani, J. H. C. Reiber, R. Ganzola, and J. Milles, "Morphological hippocampal markers for automated detection of alzheimer's disease and mild cognitive impairment converters in magnetic resonance images," *J. Alzheimer's Dis.*, vol. 17, no. 3, 2009, doi: 10.3233/JAD-2009-1082.

[8] C. R. Jack, "Alzheimer disease: New concepts on its neurobiology and the clinical role imaging will play," *Radiology*, vol. 263, no. 2. pp. 344–361, 2012, doi: 10.1148/radiol.12110433.

[9] C. R. Jack *et al.*, "Magnetic resonance imaging in Alzheimer's Disease Neuroimaging Initiative 2," *Alzheimer's Dement.*, vol. 11, no. 7, pp. 740–756, Jul. 2015, doi: 10.1016/j.jalz.2015.05.002.

[10] A. Holzinger, C. Biemann, C. S. Pattichis, and D. B. Kell, "What do we need to build explainable AI systems for the medical domain?," no. Ml, pp. 1–28, 2017, [Online]. Available: http://arxiv.org/abs/1712.09923.

[11] A. Singh, S. Sengupta, and V. Lakshminarayanan, "Explainable deep learning models in medical image analysis," pp. 1–18, 2020, [Online]. Available: http://arxiv.org/abs/2005.13799.

[12] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should i trust you?' Explaining the predictions of any classifier," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, vol. 13-17-August-2016, doi: 10.1145/2939672.2939778.

[13] J. R. Quinlan, "Induction of Decision Trees," 1986.

[14] B. Fischl *et al*., "Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain," *Neuron*, vol. 33, no. 3, 2002, doi: 10.1016/S0896-6273(02)00569-X.

[15] B. Fischl *et al.*, "Automatically Parcellating the Human Cerebral Cortex," *Cereb. Cortex*, vol. 14, no. 1, 2004, doi: 10.1093/cercor/bhg087.

[16] R. M. Haralick, I. Dinstein, and K. Shanmugam, "Textural Features for Image Classification," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-3, no. 6, 1973, doi: 10.1109/TSMC.1973.4309314.

[17] M. R. Berthold *et al*., "Knime : The Konstanz Information Miner Knime : The Konstanz Information Miner," *Comput. (Long. Beach. Calif).*, vol. 11, no. 1, pp. 319–326, 2008, doi: 10.1145/1656274.1656280.

[18] L. Breiman, J. Friedman, C. Stone, and R. Olshen, "Classification and regression trees," 1984.

[19] A. Sarica, A. Cerasa, and A. Quattrone, "Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: A systematic review," *Front. Aging Neurosci.*, vol. 9, no. OCT, pp. 1–12, 2017, doi: 10.3389/fnagi.2017.00329.

[20] J. Dougherty, R. Kohavi, and M. Sahami, "Supervised and Unsupervised Discretization of Continuous Features," in *Machine Learning Proceedings 1995*, 1995, pp. 194–202.

[21] Quinlan J.R, "Simplifying Decision Trees," *International Journal of Man-Machine Studies*, vol. 27, no. 3. pp. 221–234, 1987.

[22] "KNIME Hub." https://hub.knime.com/ (accessed Jun. 23, 2020).

[23] R Core Team, "R Core Team (2014). R: A language and environment for statistical computing.," *R Found. Stat. Comput. Vienna, Austria. URL http//www.R-project.org/.*, 2014.

[24] H. Deng, "Interpreting tree ensembles with inTrees," *Int. J. Data Sci. Anal.*, vol. 7, no. 4, 2019, doi: 10.1007/s41060-018-0144-8.

[25] A. C. Kakas, P. Moraitis, and N. I. Spanoudakis, "GORGIAS: Applying argumentation," *Argument Comput.*, vol. 10, no. 1, 2019, doi: 10.3233/AAC-181006.

[26] A. V. Lebedev *et al*., "Random Forest ensembles for detection and prediction of Alzheimer's disease with a good between-cohort robustness," *NeuroImage Clin.*, vol. 6, pp. 115–125, 2014, doi: 10.1016/j.nicl.2014.08.023.

[27] B.-K. Choi *et al*., "Convolutional Neural Network-based MR Image Analysis for Alzheimer's Disease Classification," *Curr. Med. Imaging Former. Curr. Med. Imaging Rev.*, vol. 16, no. 1, pp. 27–35, Nov. 2019, doi: 10.2174/1573405615666191021123854.

[28] A. Payan and G. Montana, "Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks," 2015. Accessed: Jun. 21, 2020. [Online]. Available: http://adni.loni.usc.edu/wp-content/uploads/how.

[29] R. M. Crum, J. C. Anthony, S. S. Bassett, and M. F. Folstein, "Population-Based Norms for the Mini-Mental State Examination by Age and Educational Level," *JAMA J. Am. Med. Assoc.*, vol. 269, no. 18, pp. 2386–2391, May 1993, doi: 10.1001/jama.1993.03500180078038.

[30] S. Leandrou, S. Petroudi, P. A. Kyriacou, C. C. Reyes-Aldasoro, and C. S. Pattichis, "Quantitative MRI Brain Studies in Mild Cognitive Impairment and Alzheimer's Disease: A Methodological Review," *IEEE Reviews in Biomedical Engineering*, vol. 11. 2018, doi: 10.1109/RBME.2018.2796598.