ROBUST PREDICTION OF COGNITIVE TEST SCORES IN ALZHEIMER'S PATIENTS

W. Izquierdo[1], H. Martin[1], M. Cabrerizo[1,5], A. Barreto[1], J. Andrian[1], N. Rishe[1], S. Gonzalez-Arias[5], D. Loewenstein[2,3,6], R. Duara[3,4,5,6], M. Adjouadi[1, 5, 6]

[1]Center for Advanced Technology and Education-CATE, Florida International University, Miami, Florida, USA.
[2]Center on Aging and Dept. of Psychiatry and Behavioral Sc., Miller School of Medicine, UM, Miami, FL, USA.
[3]Wien Center, for Alzheimer's Disease & Memory Disorders, Mount Sinai Medical Center, Miami, FL, USA.
[4]Departments of Neurology, University of Florida College of Medicine, Gainesville, FL, USA
[5]Herbert Wertheim College of Medicine, Florida International University, Miami, FL, USA
[6]Florida Alzheimer's Disease Research Center, Gainesville, Miami Beach, Miami, FL, USA and BR, FL, USA
[*]for the Alzheimer's Disease Neuroimaging Initiative
{wizqu003, hmart027, cabreriz, barretoa, andrianj, rishe, semgonz, adjouadi}@fiu.edu,
dloewenstein@med.miami.edu, ranjan.duara@msmc.com

***Abstract*— Predicting future cognitive status from current and past scores on objective cognitive tests and imaging measures would be useful in diagnosing Alzheimer's disease (AD) and to assess the progression of the disease. We used stochastic gradient boosting of decision trees on over 1,141 individuals whose clinical and imaging studies were available from the Alzheimer's disease Neuroimaging Initiative (ADNI) database. The proposed method outperformed all the algorithms tested in all five cognitive scores (MMSE, CDRS, RAVLT, ADAS11 and ADAS13), outranking all other state-of-the-art algorithms in terms of both Pearson's correlation coefficient and root mean square error. All correlation measures between predicted and actual cognitive scores were higher than 0.9. Given the large number of subjects included in this study, all correlations were statistically significant. For the subset of MCI patients, we compared the proposed method with state of the art algorithms. Here, the proposed method outperformed all the algorithms tested in all five cognitive scores.**

I. INTRODUCTION

Alzheimer's disease (AD) is a progressive neurodegenerative disease that according to the National Institute on Aging affects around 5.3 million people in the United States making it the most common cause of dementia in the elderly.

Unfortunately, definitive diagnosis can only be made by exhaustive examination of brain tissue through autopsy or brain biopsy. Thus, no accurate diagnosis exists, and only a provisional clinical estimation of the disease presence or stage can be provided based on the patient's clinical history, neuropsychological testing and neuroimaging. However, it is known that there exists strong correlation between cognitive test scores such as the Mini-Mental State Examination (MMSE) with the true presence of the disease in its different stages. Therefore, doctors and clinicians usually take advantage of this correlation and use the clinical cognitive scores of tests such as the MMSE, or the Clinical Dementia Rating Scale (CDRS) in order to clinically conclude in which of the categories or stages of AD a patient fits best (mild, moderate, or severe dementia). They complement their decision process with detailed clinical history, physical and mental status, neurological assessment, and other traditional objective tests in order make a final clinical diagnosis [1].

This methodology used by physicians usually limits the merits of cognitive tests in longitudinal studies burdened by the missing data challenge, and for analysts to predict the disease diagnosis based on machine learning. This is because if the machine uses these scores, as attributes for prediction, there will be an unduly high accuracy in the classification task because the algorithm would be simulating the already known heuristic used by the doctors for the initial clinical diagnosis of the disease. A widely used approach to overcome this issue is to base the computer decision process on other kinds of biomarkers including biochemical, genetic, neurophysiological, and neuroimaging biomarkers as provided in taxonomy [2], while ignoring completely the useful information that the cognitive test scores provide.

This study takes advantage of the useful information provided by the Mini Mental State Examination (MMSE) and the Clinical Dementia Rating Scale (CDRS) and their known correlation with the definitive diagnosis of AD. Instead of eliminating these variables from the decision process, our approach is to actually predict these test scores over time. Our aim is to provide added evidence to the prospects of a patient possibly developing AD

*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

through the prediction of future test scores, enhancing the means for a doctor to estimate a future prognosis and assess the evolution of the disease.

II. Pre-processing

A. The data

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). ADNI provides a longitudinal multivariate data collection of clinical, imaging, biochemical, and genetic features for the diagnosis of Alzheimer's disease and its prodromal stages [1].

The creators of ADNI took advantage of the strong correlation between the cognitive test scores such as MMSE and CDRS with the definitive AD diagnosis. Thus, they base their initial classification methodology on these test scores. This is documented in the ADNI general eligibility criteria described at www.adni-info.org. It follows a summary of the mentioned class eligibility as provided by [3]:

"All AD patients met the NINCDS/ADRDA criteria for probable AD, had mild level of dementia, defined as having a Mini-Mental State Examination (MMSE) score between 20 and 26, and had a Clinical Dementia Rating (CDR) score of 1.0.

The inclusion criteria for MCI is as follows:

- MMSE score between 24 and 30.
- Memory complaints and objective memory impairment measured by the Logical Memory II subscale of the Wechsler Memory Scale (education adjusted).
- CDR of 0.5
- Absence of significant levels of impairment in other cognitive domains.
- Preserved activities of daily living.
- Absence of dementia."

Since ADNI currently uses MMSE and CDR test scores to infer the disease stage, there would be no contribution to be made in developing a new classifier that takes into consideration the aforementioned scores. In other words, it will be akin to estimating a true function G that maps the set of input variables $\boldsymbol{x} = \{x_1, x_2, \dots, x_n\}$ onto the output variable $g = G(\boldsymbol{x})$:

$$G\left(\boldsymbol{x_i}(t < t_j)\right) = \begin{cases} CN\ if\ the\ i^{th}\ patient\ is\ labeled\ as\ CN\ at\ time\ t_j \\ EMCI\ if\ the\ i^{th}\ patient\ is\ labeled\ as\ EMCI\ at\ time\ t_j \\ LMCI\ if\ the\ i^{th}\ patient\ is\ labeled\ as\ LMCI\ at\ time\ t_j \\ AD\ if\ the\ i^{th}\ patient\ is\ labeled\ as\ AD\ at\ time\ t_j \end{cases}$$

The output variable $g = G(\boldsymbol{x})$ does not represent the true diagnosis of the disease. Instead, it represents the clinical labels assigned by the physicians involved, and as previously discussed these labels were assigned by taking into account the cognitive test scores. This means we can decompose the original function G as a function of a function, in other words:

$$G\left(\boldsymbol{x_i}(t < t_j)\right) = H\left(\{F^{(n)}(\boldsymbol{x_i})\}\right) \quad (1)$$

Where n defines the specific neuropsychological test considered, that is $n = MMSE, CDR, \dots etc.$

In other words, $\{F^{(n)}(x_i)\}$ could be expressed as:

$$\{F^{(n)}(x_i)\} = \{F^{(MMSE)}(x_i), F^{(CDR)}(x_i), \dots, F^{(N)}(x_i)\} \quad (2)$$

H represents the already known heuristic used by physicians to do the classification of the stage of the disease. The aforementioned ADNI general eligibility criteria is an example of this heuristic H.

Modeling H has no additional value other than automatization. Actually, if we target the problem of modeling H we can only obtain as good a classifier as the person (or persons) who labeled our training data, and evidently this H would differ from one physician to another and from dataset to dataset.

This paper focuses on the task of predicting the values of all or a subset of the functions $\{F^{(n)}(\boldsymbol{x_i}(t < t_j))\}$, where $F^{(n)}$ represents the future cognitive test scores of the n^{th} test (MMSE, CDR, or any other) at time t_j using all available information at any other time $t < t_j$. Such predicted scores will serve as a tool to help physicians gauge future AD stage classification. In this way, we are constructing a model that could be used across different databases leaving the actual classification heuristic H up to the physicians.

The challenge here is to predict the cognitive test scores on the 24th month starting from the baseline examination. In order to validate our results, the one constraint we enforced over the database, is that a record for the 24th month (the intended prediction) must exist for every patient used. Within this constraint, our dataset from ADNI consisted of 1,141 individuals. The age of 1,141 individuals was approximately symmetrically distributed (bell shaped curve) around a 73.97 years mean value, with a minimum of 55, a maximum of 91.4 and a standard deviation of 7.05 years. The sample includes 649 (57%) male and 492 (43%) female subjects. Also, the baseline classification is distributed such that 349 individuals are cognitively normal (CN), 208 are classified as early mild cognitive impaired (EMCI), 411 are classified as late mild cognitive impaired (LMCI) and 173 as AD.

B. Data Handling

In the past, several classification and regression methods have attempted to classify the different stages of the disease, predict either future cognitive test scores, or the MCI-AD conversion directly [3-8]. Most of these approaches only use baseline data and do not take advantage of the longitudinal studies available from ADNI. The study in [8] uses longitudinal data. Their method outperformed the state of the art algorithms. However, the intrinsic design of their model did not allow it to handle missing information, which remains a limiting factor for all longitudinal studies of AD. Therefore, the missing data problem is actually a big challenge for the ADNI database and a hindrance when performing longitudinal studies in AD.

Figure 1 shows the percentage of missing values per attribute over the 151 available measurements collected over the 18-month longitudinal study. It can be observed that some attributes are missing more than 50% of the time, which amplifies the need in handling the missing data challenge. Some of the values are mostly missing, rendering them impractical in longitudinal studies. However, values that are missing less than 80 percent of the time can still contribute to the system with valuable information.

Ignoring fields with missing values is not often the best approach, specifically in databases limited to only few entries. Algorithms like mean imputation, expectation maximization, and combined imputation are more widely used because they can handle missing information in a clever way. In [9], the authors performed a more detailed analysis of these methods and provide an assessment on how the final results could be affected.

Making use of decision trees is another useful way for handling missing data in a very sophisticated manner. Some algorithms treat the missing entries as allowable values for the attributes. As a result, an additional branch is created at each node to separate the subsamples that contain missing values of the attribute in question that will further develop into a sub-tree that will thereafter initiate a decision based on new relevant attributes.

In this study, the challenge was in the attempt to handle 151 attributes, some of which had a very high missing value rate (e.g., above 77% of the volumetric measurements were missing), providing additional merit to the proposed approach.

III. Methods

A. Gradient Boosting

Gradient boosting handles the function approximation problem through additive regression models. This is achieved by using the least squares method to sequentially fit a simple parameterized function to current pseudo-residuals at each iteration [10].

Gradient tree boosting (or TreeBoost) uses trees as the simple parameterized functions called base learners. This method produces competitive, highly robust procedures for both regression and classification processes. This procedure inherits the favorable characteristics of decision trees while mitigating many of the unfavorable ones such as classification accuracy, and stability among others. In 1999, reference [11] provided some of the main advantages of TreeBoost. It is worth to point out some of them to help understand what makes gradient boosting suitable for our approach: Invariance to strictly monotonic transformations is one of the most important advantages of decision trees

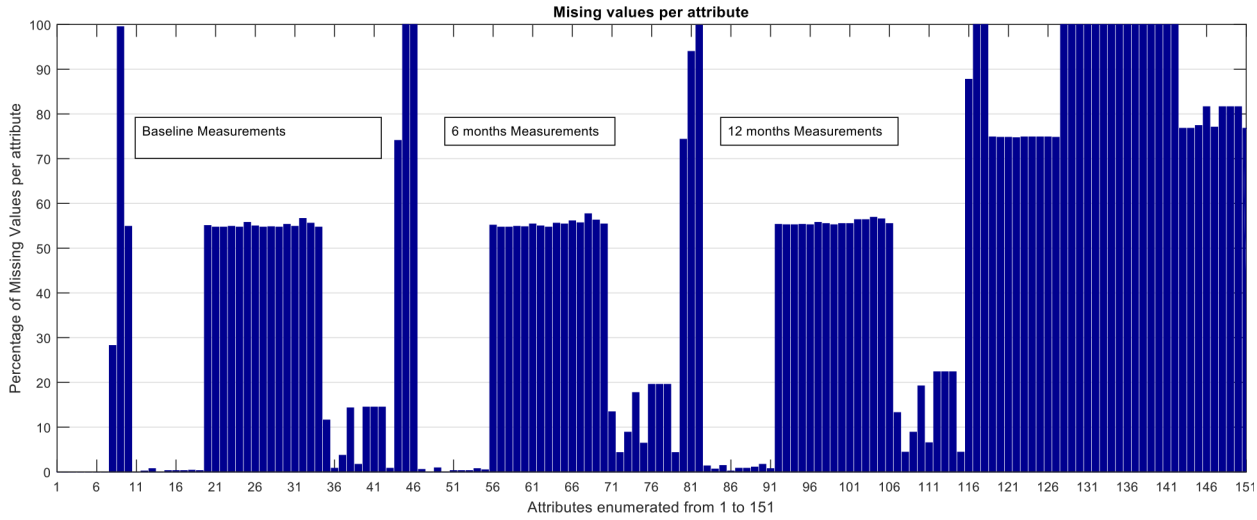

Fig. 1. Percentage of missing values per attribute. Attributes were enumerated from 1 to 151 for simplicity of the figure.

inherited by gradient boosting algorithms. Thus, using $x_j, \log x_j, e^{x_j}$ or $x_j{}^a$ as the j^{th} input variable would all yield the same result; (b) Inherent robustness in the presence of redundant or irrelevant attributes. Since decision trees usually lead to sparse solutions, they get rid of the non-useful information and take into account only those features deemed relevant in the decision making process; (c) Ability to handle missing data; an important property directly inherited from the base learners, or the decision trees; (d) Improved accuracy and stability while it prevents over-fitting by using the average over a large set of small trees.

Interpretability, one of the biggest advantages of decision trees, is not quite certain to be still attained when using gradient boosting. However, recent studies have proven that gradient boosting can be synthetized to a single small decision tree, while preserving its accuracy and stability properties [12].

B. M5' decision trees

Classical regression trees predict constant values at its leaves. In contrast, M5' trees can have multivariate linear models at each leaf. Thus, M5' trees are thus analogous to piecewise linear functions [13].

The key factors that differentiate M5' from classical regression trees are: (a) They build linear models at each leaf; (b) The simplification stage leads to sparse solutions, eliminating all the unused attributes and keeping only those that contribute the most; (c) The pruning stage compares each linear model generated at each non-leaf node with the subtree below it in terms of estimated error and prunes the subtree if it does not have a lower value; (d) The smoothing stage deals with any discontinuity, a deficiency of classical regression trees.

The pruning stage aims to acquire smaller decision trees for interpretability. This is a key property of decision trees that makes them very suitable for medical applications, in contrast to other algorithms like Artificial Neural Networks (ANNs) that suffer from opacity. Although ANNs are generally more accurate than decision trees in some applications, when the physician's experience disagrees with the ANN outcome, the doctor tends to follow his instincts making the machine's output not useful.

When using decision trees, the model's inner workings can be observed to gain valuable insight of the machine's deliberation process. A possible effect of the transparency of decision trees could be helpful in persuading physicians to adhere to the machine's decision.

In this study we make use of M5' trees as the base learners for the stochastic gradient boosting algorithm. In doing so, we have applied a method that could yield insights into the progression of AD, taking steps towards interpretability and visual mapping of the decision making process as provided by the decision trees, as well as achieving superior performance through the integration of stochastic gradient boosting as a machine learning method.

IV. RESULTS AND DISCUSSION

In order to predict future cognitive tests scores of AD and control patients at the 24th month mark, 1,141 individuals whose information was available for the 24th month exam date, were selected from the ADNI database. We handled 151 attributes, some of which had a very high missing value rate (i.e. above 77% of the volumetric measurements were missing).

By using the stochastic gradient boosting algorithm, while setting M=50 as the number of iterations or successive approximations, and setting ν=0.1 as the learning rate, we obtained an impressively high correlation coefficient between the predicted and true tests score values. We repeated the experiment 1,500 times to get a better estimation of the true correlation coefficient. We also performed a 10-fold cross validation of the learning model. We use M5' regression trees as base learners allowing no less than four instances per leaf as one of the measures to prevent overfitting. The Weka 3.8.1 software package was used to build the model and obtain the regression results. MATLAB was used to corroborate these results.

Although we were primarily interested in predicting MMSE and CDR tests scores, we also performed additional experiments to predict three other well-known cognitive tests (Alzheimer's Disease Assessment Score 11 and 13 (ADAS11 and ADAS13) and The Rey Auditory Verbal Learning Test (RAVLT) immediate) as means to corroborate the effectiveness of the proposed method as applied to any cognitive task that can be used for the diagnosis AD.

Fig. 2 shows the obtained correlation measures between the provided and predicted values for these five tests. It provides a visual perception of how accurate the prediction is for any given test. Ideally, we were expecting all of the points to be on the straight line y=x or predicted value=actual value, meaning that the prediction is 100% accurate. However, we are currently not able to reach the ideal case, mainly because the data are noisy due to the imperfections inherit from the cognitive tests. Nonetheless, although these tests are not perfect, they are some of the best tools we currently have. Another source of errors is that we are making predictions in continuous space using discretely sampled data.

Nonetheless, we can visually gauge how these plots come close to the ideal straight-line case. Even without

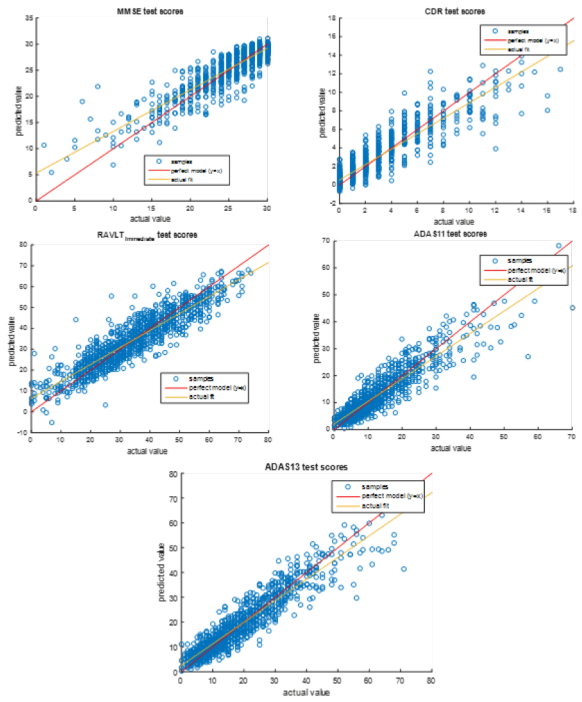

Fig. 2. Observed correlation between the five recorded tests scores with the predicted values.

looking at the numbers, we can infer from Fig. 2 that the correlation between the prediction and the ground truth is very high. We also looked at the residual plots of the five cognitive test scores and observed that the behavior was random and could not possibly generate any bias on the correlation coefficient value.

Table 1 provides a numeric appreciation of the results. It contains the values for the Pearson correlation coefficient (CORR), the mean absolute error (MAE), and root mean squared errors (RMSE) for the five tests performed. It also shows the baseline of this study (what R^2 value would it get if the 18th month test results were used as the predicted value for the 24th month).

We ran two statistical tests to prove the significance of the obtained correlation coefficients. The first test is to reject $H_0: R = 0$ in favor of $H_1: R \neq 0$ and the second one is to reject $H_0: R = R^*$ in favor of $H_1: R > R^*$, where R^* is the correlation coefficient obtained for the baseline of this study.

METHOD	MSE		ADAS	
	CORR	RMSE	CORR	RMSE
CONCAT (bl)	0.635	2.541	0.657	4.771
Ensemble (bl)	0.666	2.637	0.677	4.943
SVM (bl)	0.659	2.457	0.682	4.763
SVM (bl to M18)	0.786	2.035	0.777	7.004
Proposed (bl to M18)	**0.904**	**1.996**	**0.939**	**4.442**

Table 2. Comparison with most prominent algorithms for the cognitive tests MMSE and ADAS

From the results provided in Table 1, all correlation measures were higher than 0.9. Given the large number of subjects (over 1,100 individuals) included in this study, together with the high Pearson correlation coefficient, results show very small p-values ($< 10^{-11}$). Thus, any logical alpha value (0.05, 0.01, or even 0.001) will be good enough to reject the null hypothesis in favor of the alternative hypothesis.

Table 2 compares the results obtained by the study in [24] against our proposed method for various cognitive tests. Note that the mentioned study in [24] was done for only 88 individuals and using sparse attributes. Therefore, discrepancies with our study are expected.

These numeric values suggest two very important facts, the suitability of the stochastic gradient boosting of decision trees method for the regression problem of predicting the cognitive test scores and the additional contributions of using longitudinal data.

Since an important goal for physicians and the scientific community is to predict MCI to AD conversion, it is of special interest to explore how these numbers fit within the MCI group and how they vary in time in an attempt to model the dynamics of such measures in longitudinal studies, especially in light of the missing data challenge.

FITTING DETAILS	MMSE	CDR	RAVLT	ADAS11	ADAS13
Pearson Correlation Coefficient (R)	0.9037	0.9293	0.9113	0.9223	0.9413
R Squared	0.8167	0.8636	0.8305	0.8506	0.8860
Mean absolute error	1.4352	0.7368	4.6488	2.6872	3.1822
RMSE	1.9780	1.1512	6.0166	3.8077	4.3514
p - value (H0: R=0 vs H1: R !=0)	0.0000	0.0000	0.0000	0.0000	0.0000
Baseline Study					
Pearson Correlation Coefficient Baseline (R*)	0.7875	0.8475	0.8256	0.8224	0.8644
R* Squared	0.6202	0.7183	0.6816	0.6763	0.7472
p - value (H0: R=R* vs H1: R > R*)	0.0000	0.0000	0.0000	0.0000	0.0000

Table 1. The reported values of the Pearson's correlation coefficient (CORR), mean absolute error (MAE) and root mean squared

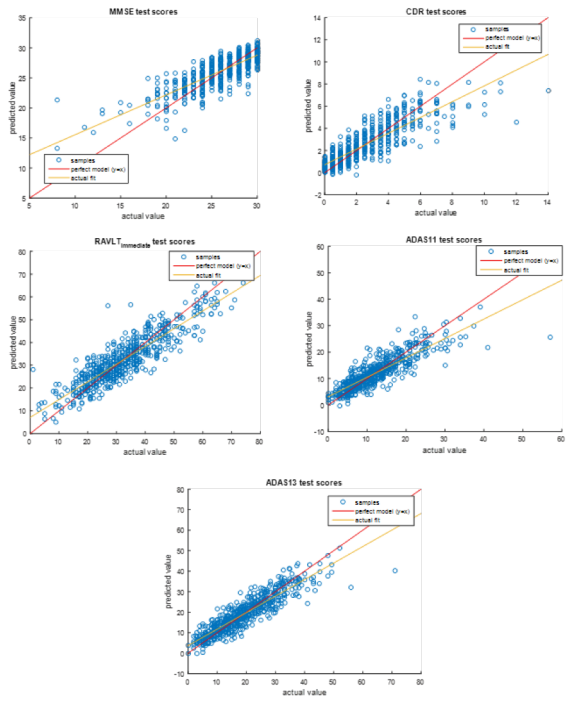

Fig. 3. Observed correlation between the five recorded tests scores with the predicted values for MCI patients only

Fig. 3 shows the observed correlation between the five recorded test scores with their corresponding predicted value for patients diagnosed with MCI at the baseline mark. For this subgroup (MCI patients), the proposed algorithm reports excellent results.

This suggests the suitability of stochastic gradient boosting for handling longitudinal data with high missing value ratios, in particular for predicting the diagnosis of AD and its prodromal stages.

Table 3 shows the numerical results for the Pearson's correlation coefficient, mean absolute error, and root mean squared error focusing on MCI patients only.

Furthermore, as a comparative study, Table 4 contrasts the results of the stochastic gradient boosting algorithm

	MMSE	CDR	RAVLT	ADAS11	ADAS13
CORR	0.8276	0.8504	0.8916	0.8648	0.9072
p-value	0.0000	0.0000	0.0000	0.0000	0.0000
MAE	1.4958	0.7858	4.3872	2.5950	3.1960
RMSE	1.9949	1.1295	5.7560	3.6595	4.3357

Table 3. The reported values of the Pearson's correlation coefficient (CORR), mean absolute error (MAE) and root mean squared error (RMSE), averaged on 10-fold tests.

to those of current state-of-the-art methods. From these results, we conclude that the proposed method proved to be very stable, outperforming the other algorithms that were tested. In contrast, other algorithms like ridge regression, SVM, or bagging, although they had an overall good performance, were not as stable, as they would outperform some in specific measures but underperform in others. However, this corroborates that using longitudinal data improves the outcome, allowing three more algorithms (Ridge regression, SVM, and Bagging) to accurately perform such prediction and classification tasks.

V. Conclusions

In this paper, we used stochastic gradient boosting over 1,141 individuals whose clinical and imaging studies were available from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database, in order to predict future cognitive scores and provide added credence in the diagnosis of AD and its prodromal stages.

Our proposed method was able to use the available longitudinal data, take advantage of the multimodality of this dataset, and handle the problem of missing values. This led us to obtain significantly better prediction results, outranking all other state-of-the-art algorithms in terms of both Pearson's correlation coefficient and root mean squared error (RMSE).

METHOD	MMSE		CDR		RAVLT		ADAS11		ADAS13	
	CORR	RMSE	CORR	RMSE	CORR	RMSE	CORR	RMSE	CORR	RMSE
ML Perceptron	0.5517	3.386	0.6365	1.9759	0.6882	10.709	0.6371	6.6223	0.6972	8.3219
KNN (K=10)	0.6101	2.9136	0.6831	1.6056	0.7220	8.9356	0.7094	5.2413	0.7562	7.0065
CART	0.6558	2.7084	0.7591	1.4028	0.8479	6.7444	0.7836	4.5504	0.8441	5.5354
Bagging	0.7699	2.2754	0.8176	1.2403	0.8878*	5.8594*	0.8375*	3.9826*	0.8757	4.9842
SVM	0.7819	2.2374	0.8387*	1.1732*	0.8547	6.6615	0.8338	4.0486	0.8829*	4.8537*
Ridge regression	*0.8030	*2.1328	0.8190	1.2562	0.8620	6.4786	0.8314	4.0893	0.8740	5.0491
Proposed	**0.8276**	**1.9949**	**0.8504**	**1.1295**	**0.8916**	**5.7560**	**0.8648**	**3.6595**	**0.9072**	**4.3357**

Table 4. Performance comparison of proposed method against state-of-the-art algorithms over MCI patients. All algorithms used the longitudinal data available (bl to M18) for prediction. Marked in "bold" the competition winner and "*" denotes 2nd place.

The enhanced accuracy of our algorithm over the subgroup of Mild Cognitive Impaired (MCI) patients was of special interest as a measure of its potential to help doctors predict MCI to AD conversion.

Providing the proven valuable longitudinal data to state of the art algorithms like Neural Networks, Support Vector Machines, Ensembles and other outstanding techniques still led to the proposed method outperforming said algorithms for every single predicted score of each cognitive test. Thus, our results give evidence of the suitability of the stochastic gradient boosting of decision trees for the regression problem, specifically for predicting the cognitive test scores of the ADNI database's patients.

By using M5' regression trees as base functions, this algorithm yielded sparse solutions and relatively small trees for enhanced interpretability of the results. At this point, the emphasis is placed on performing regression instead of classification. However, the specific aim of this study was to provide the necessary information to the doctors for them to make a more accurate diagnosis in the progression of AD and hence take more informed decisions in the planning and treatment of the disease.

ACKNOWLEDGEMENTS

We acknowledge the critical support provided by the National Science Foundation under grants: CNS-1532061, CNS-1551221, CNS-1642193, CNS-0959985, HRD- 0833093, IIP 1338922, and CNS-1429345. The generous support of the Ware Foundation is greatly appreciated. Harold Martin is supported through the NSF Graduate Fellowship Program (GRFP) DGE-1610348.

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. 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.

We appreciate the support provided by the Alzheimer's Disease Research Center (ADRC), the Florida Department of Health, Ed and Ethel Moore Alzheimer's Disease Research Program, and Wien Center for Alzheimer's Disease & Memory Disorders, Mount Sinai Medical Center, Miami, FL

REFERENCES

[1] Weiner MW, Veitch DP, Aisen PS, Beckett LA, Cairns NJ, 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; Alzheimer's Disease Neuroimaging Initiative., *The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception*. Alzheimers Dement. 2013.

[2] Mirzaei, G., Adeli, A., and Adeli, H., "Imaging and Machine Learning Techniques for Diagnosis of Alzheimer Disease," Reviews in the Neurosciences, 27:8, 2016, pp. 857-870.

[3] Lebedev AV, Westman E, Van Westen GJ, Kramberger MG, Lundervold A, Aarsland D, Soininen H, Kłoszewska I, Mecocci P, Tsolaki M, Vellas B, Lovestone S, Simmons A; Alzheimer's Disease Neuroimaging Initiative and the AddNeuroMed consortium., *Random Forest ensembles for detection and prediction of Alzheimer's disease with a good between-cohort robustness.*, NeuroImage, Clinical 6, 2014, pp. 115-125

[4] Stonnington CM, Chu C, Klöppel S, Jack CR Jr, Ashburner J, Frackowiak RS; Alzheimer Disease Neuroimaging Initiative., *Predicting clinical scores from magnetic resonance scans in Alzheimer's disease*, Neuroimage, 50, 2010, pp.1519-1535.

[5] Duchesne S, Caroli A, Geroldi C, Collins DL, Frisoni GB., *Relating one-year cognitive change in mild cognitive impairment to baseline MRI features*, Neuroimage, 47,2009, pp.1363-1370.

[6] Wang Y, Fan Y, Bhatt P, Davatzikos C, High-dimensional pattern regression using machine learning: from medical images to continuous clinical variables. Neuroimage, 50, 2010, pp.1519-1535.

[7] Zhang D, Shen D, Predicting future clinical changes of MCI patients using longitudinal and multimodal biomarkers., PLoS ONE, 7(3), 2012

[8] Korolev IO, Symonds LL, Bozoki AC; Alzheimer's Disease Neuroimaging Initiative, Predicting Progression from Mild Cognitive Impairment to Alzheimer's Dementia Using Clinical, MRI, and Plasma Biomarkers via Probabilistic Pattern Classification, PLoS ONE, 2016

[9] Ritter K, Schumacher J, Weygandt M, Buchert R, Allefeld C, Haynes JD, *Multimodal prediction of conversion to Alzheimer's disease based on incomplete biomarkers*, Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 1, 2015, pp.206-215.

[10] Friedman JH, Stochastic Gradient Boosting, Computational Statistics & Data Analysis - Nonlinear methods and data mining, 38(4), 2002, pp.367-378.

[11] Friedman JH, Greedy Function Approximation: A Gradient Boosting Machine, Stanford University, Statistics 1999.

[12] Valdes G, Luna JM, Eaton E, Simone CB, Ungar LH, Solberg TD, MediBoost: a Patient Stratification Tool for Interpretable Decision Making in the Era of Precision Medicine. Nature: Article number 37854, 2016

[13] Wang Y, Witten I, *Inducing Model Trees for Continuous Classes*, in Proc. of Poster Papers, European Conf. on Machine Learning. Prague, 1997