World Academy of Science, Engineering and Technology
International Journal of Medical and Health Sciences
Vol:13, No:2, 2019

Prediction Modeling of Alzheimer's Disease and Its Prodromal Stages from Multimodal Data with Missing Values

M. Aghili, S. Tabarestani, C. Freytes, M. Shojaie, M. Cabrerizo, A. Barreto, N. Rishe, R. E. Curiel, D. Loewenstein, R. Duara, M. Adjouadi

***Abstract*—**A major challenge in medical studies, especially those that are longitudinal, is the problem of missing measurements which hinders the effective application of many machine learning algorithms. Furthermore, recent Alzheimer's Disease studies have focused on the delineation of Early Mild Cognitive Impairment (EMCI) and Late Mild Cognitive Impairment (LMCI) from cognitively normal controls (CN) which is essential for developing effective and early treatment methods. To address the aforementioned challenges, this paper explores the potential of using the eXtreme Gradient Boosting (XGBoost) algorithm in handling missing values in multiclass classification. We seek a generalized classification scheme where all prodromal stages of the disease are considered simultaneously in the classification and decision-making processes. Given the large number of subjects (1631) included in this study and in the presence of almost 28% missing values, we investigated the performance of XGBoost on the classification of the four classes of AD, NC, EMCI, and LMCI. Using 10-fold cross validation technique, XGBoost is shown to outperform other state-of-the-art classification algorithms by 3% in terms of accuracy and F-score. Our model achieved an accuracy of 80.52%, a precision of 80.62% and recall of 80.51%, supporting the more natural and promising multiclass classification.

***Keywords*—**eXtreme Gradient Boosting, missing data, Alzheimer disease, early mild cognitive impairment, late mild cognitive impairment, multiclass classification, ADNI, support vector machine, random forest.

I. Introduction

ALZHEIMER'S Disease (AD) is a pervasive neurodegenerative disorder and the most prevalent form of dementia. According to the Alzheimer's Association, AD is the 6th leading cause of death in the United States with over 5.7 million Americans affected by 2018, a number expected to increase up to 14 million [1]. Presently, there is no known remedy for AD.

AD progresses gradually, typically resulting in episodic memory loss and behavioral changes. This combination of ambiguous and varied effects and unsettling, yet subtle progression of the disease, complicates the pursuit of realistic AD diagnosis models. Currently, cognitive tests are the best diagnostic tool, but they are still far from ideal in determining the causality of the disease. In an attempt to improve AD predictive models, researchers are exploring other biomarkers to enhance the diagnosing of AD, including structural Magnetic Resonance Imaging (MRI), functional MRI, Positron Emission Tomography (PET), Cerebrospinal Fluid (CSF) and genetic biomarkers (APOE) [2]-[5].

In recent years, machine learning algorithms have become the subject of interest for different applications [6]-[8]. AD studies have in large part dealt with binary classification with mixed success, especially for the challenging classification of CN vs. EMCI subjects. These reported results are further affected when the more pragmatic multiclass classification is assumed. Therefore, early reliable diagnosis of AD through an amalgamation of functional, structural, metabolic measurements, together with the cognitive tests (preferably not the ones used at baseline), genetic data, and other biomarkers like CSF is crucial to understand the disease and its transition phases and for the prospects of planning early treatment. Accordingly, it is essential to understand and delineate the different stages of AD in a multiclass process and be able to develop concise methods to predict and detect the disease in its earliest manifestations. To address this, many scientists agree that AD is not a binary classification problem (AD from CN) but a multiclass problem where AD progresses through multiple stages [9]-[13].

In clinical trials, which are longitudinal in nature, one of the major impediments is the problem of missing data. Many instances of input data are partially missing even in the most popular and comprehensive datasets, such as the ADNI (Alzheimer's Disease Neuroimaging Initiative) dataset. This problem becomes more complex when the dataset is crafted from multiple modalities. The difficulty in scheduling follow up visits and high drop-out rates, due to health and aging issues, are problematic obstacles. In addition to the aforementioned issues, data corruption, noise and

Maryamossadat Aghili, Christian Freytes, Mehdi Shojaie, Mercedes Cabrerizo, Armando Barreto and Malek Adjouadi are with the Center for Advanced Technology and Education, Department of Electrical and Computer Engineering, Florida International University, Miami 33174, FL, USA.

Solale Tabarestani is with the Center for Advanced Technology and Education, Department of Electrical and Computer Engineering, Florida International University, Miami 33174, FL, USA (corresponding author, e-mail: staba006@fiu.edu).

Naphtali Rishe, and Maryamossadat Aghili are with School of Computing and Information Sciences, Florida International University, Miami 33174, FL, USA.

Rosie E. Curiel, and David Loewenstein are with the Center for Cognitive Neuroscience and Aging, University of Miami Miller School of Medicine, Miami, FL, USA.

Ranjan Duara is with the Wien Center for Alzheimer's Disease & Memory Disorders, Mount Sinai Medical Center, Miami, FL, USA.

David Loewenstein, Rosie E. Curiel, Ranjan Duara, and Malek Adjouadi are with Florida ADRC (Florida Alzheimer's Disease Research Center at Gainesville).

misalignment exacerbate this problem further, making the potentially significant data samples difficult to render or restore [14]. These missing measurements hinder the performance of potentially good models. Often, algorithms would either (a) discard subjects with missing values from further experimentation, or (b) replace the missing values with zero values or the mean of the given attribute. AD diagnosis and prognosis could vastly benefit from an effective mechanism for estimating missing values. However, high dimensional datasets usually have nonlinear and complicated correlations; an issue that greatly interferes with conclusive estimation. Therefore, it is important to develop a model to overcome the missing data challenge and make use of the statistical information of the available data.

In determining avenues for correctly modeling AD in its different stages, the large number of dimensions (which is defined by the number of variables in the data) has made it difficult to obtain an optimal solution even when deploying the most effective learners. This challenge has led to utilizing a technique known by researchers as boosting [4], [11], [15]. In the development of machine learning algorithms, there exists the notion of weak versus strong learners, where a weak learner has less than random chance accuracy and a strong learner results in an accuracy greater than random chance. Boosting methods combine weak learners to create a strong learner in situations where a strong learner is not available. The study of [11] utilized a Gradient Tree Boosting (TreeBoost) technique with M5 decision trees to make predictions of cognitive scores in Alzheimer's patients. This study was able to accurately predict cognitive scores up to 14 months, which is potentially helpful for detecting MCI to AD converters. However, this investigation does not segregate EMCI and LMCI subjects and has focused only on predicting cognitive scores. Natarajan et al. introduces a three-way classifier that bases its decisions on structural MRIs [16]. They accomplished this by segmenting clinically significant areas and using a relational learning algorithm to provide classification, combined with the use of gradient-tree boosting. The experiments in this paper were successful in discriminating AD from MCI and AD from CN subjects, but had a high number of false positives and false negatives in delineating CN from MCI, a more crucial stage for early intervention and treatment planning. The methodology employed here is promising since it achieved progress producing predictions based on neuroimaging modalities, but it falls short in accurately detecting the prodromal stages of AD. Zhang et al. developed an algorithm to effectively predict the cognitive scores of subjects and compared the use of different boosting methods [4].

Following the promising results of the aforementioned articles, this paper applies gradient boosting technique to create a model that can discriminate all the prodromal stages of AD to include CN, EMCI, LMCI and AD. This will allow for determining whether a subject has the potential to convert to AD at an earlier stage, while ample time for treatment is also made possible.

The remainder of this paper is organized as follows. The method and data considered in this study are described in Section II, followed by results and discussion in Section III. Finally, conclusion is provided in Section IV.

II. Methods and Data

A. Gradient Boosting

Bagging and boosting are well-known forms of ensemble techniques which rely on a collection of predictors to make a final prediction by voting or averaging. In bagging, multiple independent learners are built on the data and the predictions are combined by some model averaging techniques such as weighted average, normal average or majority voting. This model takes a bootstrap of data for each sub-model, making sub-models slightly different from each other. Each observation is selected with a replacement to be available as input for other sub-models. In this way, many uncorrelated sub-models are trained to make a final model with minimized error.

On the other hand, in boosting, sub-models are not built independently but sequentially. A subsequent sub-model learns from errors of the previous sub-model. As such, observations do not have equal probability of occurrence in subsequent learners, and one with the higher error is weighted more to appear with higher chance. This causes the observation to be picked based on the error made in the previous sub-model. These sub-models can be any of the decision trees or other regressors and classifiers models [17]. Fig. 1 demonstrates the difference of the boosting and bagging techniques.

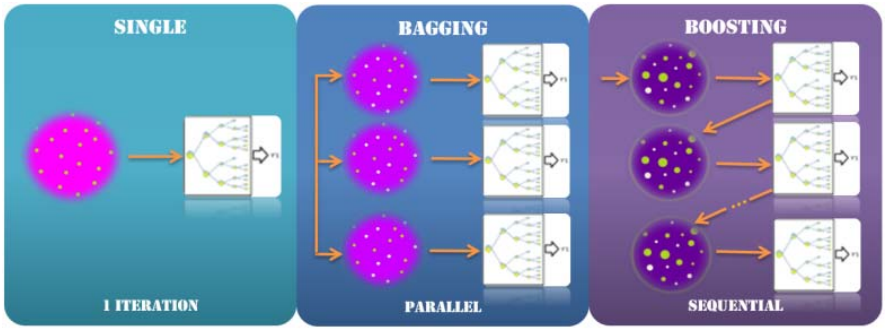

Fig. 1 Illustration of bagging and boosting techniques

Gradient boosting is a form of boosting technique which is applied in both regression and classification problems. It defines a Mean Squared Error (MSE) loss function and tries to minimize it by using gradient descent, and consequently updating the prediction based on a learning rate. It basically updates the prediction in order to minimize the sum of the residuals. Fig. 2 shows the main steps of the gradient boosting.

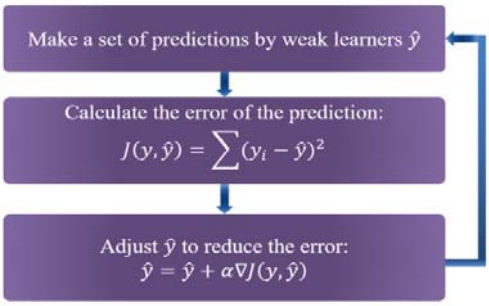

Fig. 2 Main steps of gradient boosting

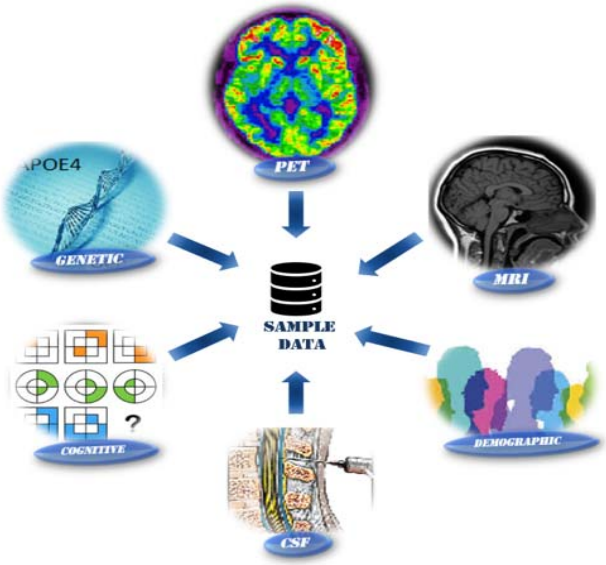

Fig. 3 Sample data point

TABLE I
BIOMARKERS

Source	Features
Cognitive tests	EcogPtMem, EcogPtLang, EcogPtVisspat, EcogPtPlan, EcogPtOrgan, EcogPtDivatt, EcogPtTotal, EcogSPMem, EcogSPLang, EcogSPVisspat, EcogSPPlan, EcogSPLang,EcogSPOrgan, EcogSPDivatt, EcogSPTotal,FAQ, MOCA, RAVLT
MRI	Ventricles, Hippocampus, WholeBrain, Entorhinal, Fusiform, MidTemp, ICV, FLDSTRENG, FSVERSION
PET	FDG, PIB amyloid, AV45 amyloid, CDRSB
Genetic	APOE4
Demographic	AGE, Gender, Education
CSF	Ab1, t-tau, p-tau

B. Data

To verify the described model in this paper, the data provided by the Alzheimer Disease Neuroimaging Initiative (ADNI) is used (adni.loni.usc.edu). This dataset is released in 2003 as a project guided by Principal Investigator Michael W. Weiner, MD. This dataset mainly concentrates on biomarker assessments to estimate the progression of mild cognitive impairment (MCI) and AD [18].

ADNI data are composed of multiple modalities of data which are processed with a standard pipeline resulting in a large matrix of subject features and their test measurements. Subjects are ordered in rows and biomarkers in columns. From all biomarkers, the following indicators were selected; CSF, MRI, PET, genetics, cognitive tests, and demographics data. A sample data point is shown in Fig. 3. with the details of the features in the sample data summarized in Table I. The dataset characteristics are shown in Table II. In order to address the issue caused by pulling the data from various scanners with difference settings, we included the attribute defining the scanner type.

TABLE II
DATASET CHARACTERISTICS

Group	Number of Subject	MMSE	Age	Education years
AD	342	23.21 ± 2.07	75.02	15.21
CN	417	29.07 ± 1.12	74.75	16.31
EMCI	310	28.30 ± 1.55	71.19	15.96
LMCI	562	27.18 ± 1.80	73.99	15.86

III. RESULTS AND DISCUSSION

To illustrate the performance of the proposed approach, we performed experiments on the ADNI database. A total of 1631 subjects including 342 AD patients, 562 LMCI, 310 EMCI, and 417 CN have been selected from the ADNI-Merge dataset. Almost all the samples in the pool of 1631 subjects and 39 features have at least one missing value, which in total results in around 28% of the data missing. Thus, discarding the samples with missing value takes a large part of this data out of consideration, so a solution for addressing the missing values should be implemented prior to any further investigations. Many studies are deploying the sparsity aware constrains to handle different types of sparsity patterns in the data including eXtreme Gradient Boosting (XGBoost) [19], [20].

Therefore, in this paper, XGBoost classifier is proposed which not only addresses all four potential groups (CN, EMCI, LMCI, and AD) simultaneously, but also instinctively handles the missing data. A tenfold cross-validation is used to estimate the best hyper parameters for the XGBoost method. The results are compared with two state-of-the-art classification algorithms: Random forest (RF), which is a well-known bagging algorithm, and support vector machine (SVM). While both XGBoost and RF algorithms are robust to the unscaled datasets, the data must be scaled and normalized before being fed to the SVM.

Particularly, subjects are separated into 10 parts (each part with a relatively equivalent size). Subjects from one part are chosen as the testing data, and the remainder is used as the training data. XGBoost from the XGB library [21], RF and SVM from the Scikit-learn library [22] have been adopted for training the classifiers. From Table III, which represents the

accuracy, F1-score, Precision and Recall of all classification techniques, it can be observed that XGBoost provides considerably better performance in multiclass classification in AD.

Figs. 4-6 illustrate the AUC and ROC curves of the three algorithms on the classification of the four defined groups. Receiver Operating Characteristics (ROC) curve is one of the most important evaluation metrics, which shows the performance of a classification model at all thresholds. ROC is a probability curve, and AUC represents the degree of separability. These curve plots are based on the two parameters of true positive rate (TPR) and false positive rate (FPR) defined as follows:

$$TPR = \frac{TP}{TP+FN}; \quad FPR = \frac{FP}{FP+TN} \tag{1}$$

TABLE III
ALGORITHMS PERFORMANCE COMPARISON

Model / Metrics	Accuracy	F-Score	Precision	Recall
XGBoost	80.52	80.31	80.62	80.51
RF	77.17	76.95	77.40	77.17
SVM	74.58	74.40	75.98	74.58

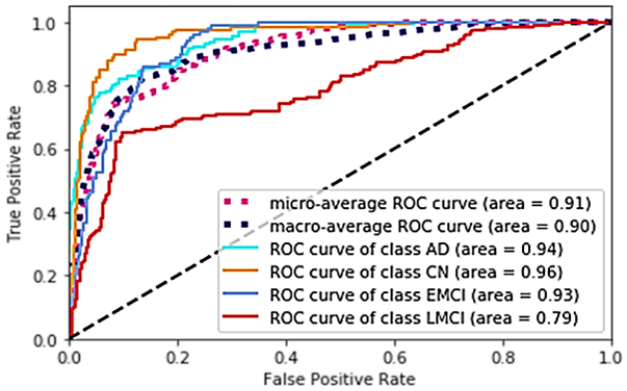

Fig. 4 Receiver Operating Characteristics for SVM

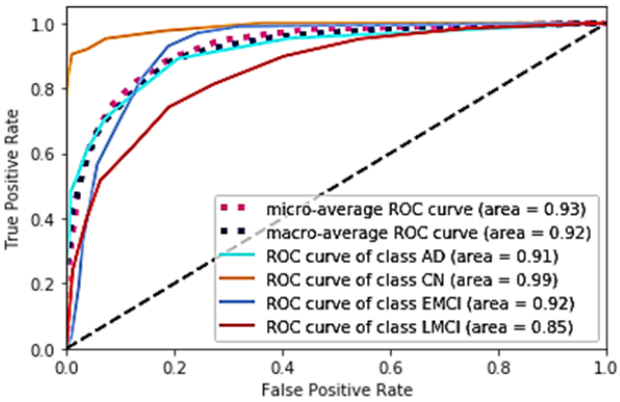

Fig. 5 Receiver Operating Characteristics for RF

Plotted in Fig. 6 is the ROC which shows how much the XGBoost model is capable of distinguishing between classes in comparison to SVM and RF, shown earlier in Figs. 4 and 5, respectively. These graphs show that the most difficult task is the correct classification of the patients in the class of LMCI with the AUC of 87% in XGBoost, 85% in RF and 79% in SVM. The challenge is that the LMCI subjects are usually misclassified with either EMCI or AD.

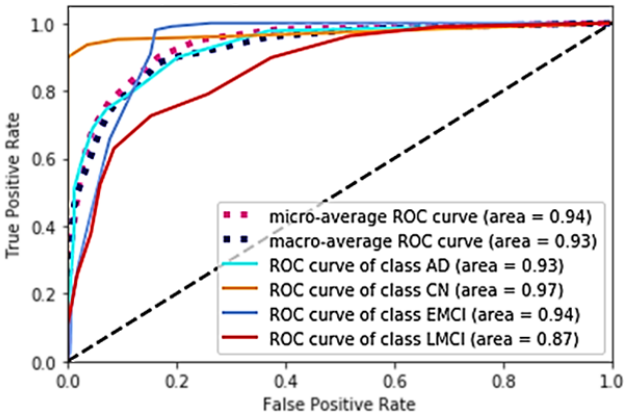

Fig. 6 Receiver Operating Characteristics for XGBoost

IV. CONCLUSION

This paper demonstrates the superiority of XGBoost in the complex scenario of multiclass classification of AD subjects while contending with the missing data challenge. The focus was placed in segregating the MCI group into EMCI and LMCI subjects, which has not been well explored yet. The intention is to use all prodromal stages of the disease in a multiclass process, which is more natural and realistic.

Multimodal studies that incorporate several biomarkers are shown to improve the classification and prediction in Alzheimer's related research. However, capturing data from multiple sources imposes the process of handling the missing data issues. A primary objective of this paper was to seek a high classification performance in delineating the subtler and challenging groups of CN and EMCI. These ensure early detection and enable early intervention and treatment planning. We received the best accuracy possible showing that the XGBoost algorithm offers a clever way to handle missing data.

ACKNOWLEDGMENT

This research is supported by National Science Foundation (NSF) under NSF grants CNS-1532061, CNS-1429345, CNS-1551221, CNS-1338922, 1Florida Alzheimer's Disease Research Center (ADRC) (NIA 1P50AG047266-01A1) and the Ware Foundation. 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 (Departmenwt 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 other sources. Details of acknowledgement are available at http://adni.loni.usc.edu/wpcontent/uploads/how to apply/ADNI Manuscript Citations.pdf.

REFERENCES

[1] J. Karlawish, C. R. Jack, W. A. Rocca, H. M. Snyder, and M. C. Carrillo, "2017 Alzheimer's disease facts and figures," *Alzheimer's Dement.*, 2017.

[2] Cuingnet R, Gerardin E, Tessieras J, Auzias G, Lehéricy S, Habert M-O, Chupin M, Benali H, Colliot O, "Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database," *Neuroimage*, vol. 56, no. 2, pp. 766–781, 2011.

[3] K. Ritter, J. Schumacher, M. Weygandt, R. Buchert, C. Allefeld, and J. D. Haynes, "Multimodal prediction of conversion to Alzheimer's disease based onincomplete biomarkers," *Alzheimer's Dement. Diagnosis, Assess. Dis. Monit.*, vol. 1, no. 2, pp. 206–215, 2015.

[4] H. Zhang, F. Zhu, H. H. Dodge, G. A. Higgins, G. S. Omenn, and Y. Guan, "A similarity-based approach to leverage multi-cohort medical data on the diagnosis and prognosis of Alzheimer's disease," *Gigascience*, vol. 7, no. 7, pp. 1–10, 2018.

[5] A. Farzan, S. Mashohor, A. R. Ramli, and R. Mahmud, "Boosting diagnosis accuracy of Alzheimer's disease using high dimensional recognition of longitudinal brain atrophy patterns," *Behav. Brain Res.*, vol. 290, pp. 124–130, 2015.

[6] S. Tabarestani, M. Eslami, and F. Torkamni-Azar, "Painting style classification in Persian Miniatures," in *Iranian Conference on Machine Vision and Image Processing, MVIP*, 2016.

[7] L. Ghorbanzadeh and A. E. Torshabi, "An Investigation into the Performance of Adaptive Neuro-Fuzzy Inference System for Brain Tumor Delineation Using ExpectationMaximization Cluster Method; a Feasibility Study," Frontiers in Biomedical Technologies, vol. 3, pp. 8-19, 2017.

[8] S. Tabarestani, M. Aghili, M. Shojaie, C. Freytes, "Profile-Specific Regression Model for Progression Prediction of Alzheimer's Disease Using Longitudinal Data," *17th IEEE Int. Conf. Mach. Learn. Appl.*, 2018.

[9] R. C. Petersen and J. C. Morris, "Mild cognitive impairment as a clinical entity and treatment target," *Arch. Neurol.*, vol. 62, no. 7, pp. 1160–1163, 2005.

[10] E. Moradi, A. Pepe, C. Gaser, H. Huttunen, and J. Tohka, "Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects," *Neuroimage*, vol. 104, pp. 398–412, 2015.

[11] W. Izquierdo et al., "Robust prediction of cognitive test scores in Alzheimer's patients," 2017 IEEE Signal Processing in Medicine and Biology Symposium (SPMB), Philadelphia, PA, 2017, pp. 1-7.

[12] H. Il Suk, S. W. Lee, and D. Shen, "Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis," *Neuroimage*, vol. 101, pp. 569–582, 2014.

[13] M. Aghili, S. Tabarestani, M. Adjouadi, and E. Adeli, "Predictive Modeling of Longitudinal Data for Alzheimer's Disease Diagnosis Using RNNs," in *PRedictive Intelligence in MEdicine*, 2018, pp. 112–119.

[14] M. Mafi, S. Tabarestani, M. Cabrerizo, A. Barreto and M. Adjouadi, "Denoising of ultrasound images affected by combined speckle and Gaussian noise," in IET Image Processing, vol. 12, no. 12, pp. 2346-2351, 12 2018.

[15] W. Izquierdo, H. Martin, M. Cabrerizo, A. Barreto, J. Andrian, N. Rishe, S. Gonzalez-Arias, D. Loewenstein, R. Duara, M. Adjouadi, "Robust prediction of cognitive test scores in Alzheimer's patients," 2017 IEEE Signal Processing in Medicine and Biology Symposium (SPMB), Philadelphia, PA, 2017, pp. 1-7.

[16] S. Natarajan, B. Saha, S. Joshi, A. Edwards, T. Khot, E. M. Davenport, K. Kersting, C. T. Whitlow, and J. A. Maldjian. "Relational learning helps in three-way classification of Alzheimer patients from structural magnetic resonance images of the brain," Intl. Journal of Machine Learning and Cybernetics, pages 1–11, 2013.

[17] J. H. Friedman. "Greedy function approximation:a gradient boosting machine," The Annals of Statistics, 29(5):1189–1232, 2001.

[18] M. W. Weiner, D. P. Veitch, P. S. Aisen et al., "The Alzheimer's disease neuroimaging initiative: a review of papers published since its inception", Alzheimer's Dement., vol. 9, no. 5, pp. e111-e194, 2013.

[19] M. Eslami, F. Torkamani-Azar, and E. Mehrshahi, "A Centralized PSD Map Construction by Distributed Compressive Sensing," *IEEE Commun. Lett.*, 2015.

[20] M. Eslami, A. H. Gazestani, and S. A. Ghorashi, "Introduction and Patent Analysis of Signal Processing for Big Data," *Adv. Parallel Comput.*, vol. 33, no. Big Data and HPC: Ecosystem and Convergence, pp. 101-–119, 2018.

[21] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," arXiv preprint arXiv:1603.02754, 2016.

[22] T. Klikauer, "Scikit-learn: Machine Learning in Python," *TripleC*, 2016.