Connecting gut microbial metabolites to amyloid and tau biomarkers for Alzheimer's disease prediction.

Alejandro I. Trejo-Castro
School of Engineering and Sciences
Tecnológico de Monterrey
Monterrey, México
a00818219@tec.mx

***Abstract*—Alzheimer's disease (AD) is the most common dementia and the main neurodegenerative disease. Multiple theories have tried to explain Alzheimer's disease, however they have not resulted in a treatment or preventive strategy. Lately topics such as metabolomics and microbiota are a new approach to find biomarkers and better understand the disease. This study intends to apply data science and machine learning techniques to connect metabolic data with standardized CSF biomarkers for Alzheimer's disease prediction. Data has been gathered from the Alzheimer's Disease Neuroimaging Initiative in a partnership with the Alzheimer's Disease Metabolomics Consortium. The results shows that the combination of gut microbial metabolites with the well-known amyloid and tau biomarkers could enhance the diagnosis and prediction of AD.**

***Index Terms*—Alzheimer's Disease, ADNI, biomarkers, data science, machine learning.**

I. Introduction

AD is the most common among neurodegenerative diseases, accounting for 60 to 80% of all dementia cases reported in 2010. So far, 35.6 million people worldwide have been diagnosed with dementia and yet, it is expected to see an increment of up to 225% on the number of patients during a 40-year time frame (2010-2050) [1]. The prognosis for these patients is not encouraging. AD has been known for over 110 years and a cure has not yet been found, still numerous clinical trials have not been completely successful [2]. With a life expectancy after diagnosis of 3 to 9 years [3], AD is the sixth leading cause of death in the United States. Moreover, there were 5.8 million people living with this condition in 2010, and with life expectancy increasing, it is expected to reach 14 million by the middle of the century [4]. The AD hallmark pathology is so far explained as accumulations of amyloid-β (Aβ) protein fragments outside neurons and hyperphosphorylated tau tangles within neurons [5].

However, recently the gut microbiome has attracted attention to the study of AD. The human gut microbiota is the diverse collection of microorganisms (e.g., bacteria, archaea, and fungi) that reside in the gastrointestinal tract and strongly influence human physiology, including roles in nutrition, digestion, inflammation and immunity [6]. It appears that the theory of the microbiota-gut-brain relationship has a critical role in the development and behavior of the brain. We know now that humans have 10-times more bacteria than the number of cells in our body. This means that around 100 billion bacteria of up to 1000 different bacterial species coexist in the human gastrointestinal tract. Even more interestingly is the fact that the total genome of the gut microbiota is 150 times larger than the human genome composed by 26,600 transcripts encoding proteins that reach a genetic complexity of 4,026,600 transcriptomes [7] [8] Hence, it is clear that humans are an entire symbiotic ecosystem, where not only human genetics plays a role. In fact, there are studies that report that the microbiota may be the etiopathogenesis of amyloidosis in the brain of subject with AD, since bacterial infection induce A peptide oligometry and the gut microbiota can produce its own peptides [9]

Throughout the history of AD there have been several hypotheses that could explain the progression of AD. Among them are cholinergic hypothesis, amyloid hypothesis, tau propagation hypothesis, mitochondrial cascade hypothesis, calcium homeostasis hypothesis, neurovascular hypothesis, inflammatory hypothesis, metal ion hypothesis, and lymphatic system hypothesis [10]. But we are still not there when it comes to treatment or prevention. Moreover, recent failures in clinical trials suggest that it is time to consider alternative novel strategies, one that has had a lot of attention is the microbiota-gut-brain axis and its implication in the development of AD.

In addition, mild cognitive impairment (MCI) represents a transitional state between normal cognition and dementia, since it indicates cognitive deficits including memory and non-memory impairments and an increased risk for the development of dementia. It will be essential to our study because we will distinguish between subjects with MCI who will progress to AD, and those who will not, in order to find biomarkers for an early diagnosis.

The objective of this study was to determine whether gut microbial metabolites features in addition to the amyloid and tau biomarkers could predict MCI to AD progression.

II. Materials and Methods

A. Data

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public–private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. For up-to-date information, see www.adni-info.org.

In addition to the patient's information and demographics, two main data sets were downloaded from ADNI and Alzheimer's Disease Metabolomics Consortium (ADMC). The fist data set is called "ADMC U Hawaii UPLC-MS/MS Gut Metabolites Serum Longitudinal [ADNI1,GO,2]" and it consists of 104 metabolites including bile acids in human serum samples. The version downloaded has as date 2021-12-13 from the University of Hawaii Cancer Center. The second data set named "UPENN CSF Biomarker Master Dataset" includes 3 well-known AD biomarkers: the forty-two amino acid-long Aβ peptide (Aβ $_{1-42}$), total tau (t-tau) and hyperphosphorylated tau at threonine 181 (p-tau$_{181}$) from the University of Pennsylvania Perelman School of Medicine [11].

B. Subject Inclusion

The experiment included only baseline information from subjects with a baseline MCI diagnosis, with more than 65 years old in order to deal with late-onset AD, with available sex and years of education information. They also had to have measurements in each of the gut microbial metabolites and biomarkers,if it was not the case, the patients were ruled out from the study. It was done in this way because the patient either had all the data or had none, there were no cases where they were missing a couple of data, so it was better to remove them.

From these, the 33 subjects who had their first AD diagnosis 2 years after their baseline visit were regarded as progressers (MCIp), while the 74 subjects who never had an AD diagnosis and participated in the study for at least 5 years were labeled as stable (MCIs). Patients who did not meet either the MCIp or the MCIs criteria were excluded from the study.

TABLE I details the demographic characteristics of our population. There was no significant difference in age and years of education between groups when tested using the Wilcoxon rank-sum test nor a significant difference in male/female proportion under a chi-squared test. MCI and AD diagnoses were determined as defined by ADNI guidelines [12].

C. Statistical Analysis

It was performed a univariate analysis. Here hypothesis testing was made where features were compared according to the clasess using the Wilcoxon rank-sum test. A feature was determined significantly different between groups if a p-value lower than 0.05 was found.

TABLE I
Demography of the population

Group of study	MCIs	MCIp	p-value
Subjects (males)	74(42)	33(24)	0.176
Age	73.9±5.6	75.9±5.6	0.079
Education	16.2±3.0	16.0±3.1	0.764

Moreover, a point biserial correlation between the target and every feature in the data set was done, this is equivalent to pearson correlation. This type of correlation coefficient is used when is compared a continuous feature with a binary or dichotomous feature.

For the machine learning modelling, 8 techniques were implemented through sklearn in Python: logistic regression, svm linear, gaussian NB, decision tree classifier, random forest (n estimators=100, criterion='entropy'), bagging, gboosting (n estimators=100, learning rate=1, max depth=1), and XGB (objective= binary logistic, n estimators=2). Each model was performed under a 8-fold cross validation, the metrics that will be use to analyze the results are ROC AUC score and accuracy. It is preferable to give more importance to ROC AUC score because of the imbalance dataset. Two different experiments will be carried out. The first one with the 3 CSF well known biomarkers and the other one with the combination of the 3 biomarkers with the gut metabolites.

D. Results

Twelve significantly different attributes were found between the two classes (see TABLE II). The 3 attributes with the lowest p-value correspond to the 3 well-known CSF biomarkers for AD. This was followed by 9 of the 104 gut microbial metabolites, which corresponds to 8.6% of the data.

Our goal was to compare predictive models between the union of CSF biomarkers and gut microbial metabolites. The selected features for the machine learning algorithms were these 12 attributes. In Figure 1 we can appreciate the distribution between classes for the most significant feature. Here 0 corresponds to MCIs and 1 to MCIp.

In TABLE III, it is presented the top features in correlation with the target through the point biserial correlation. This results told us that is needed a combination of features since a stand alone feature can not predict good the classes.

The mean and standard deviation of the metrics for the different machine learning techniques and experiments trough a 8 fold-cross validation are shown in TABLE IV. Here B means biomarkers for amyloid and tau and B+G stands for biomarkers with gut microbial metabolites. The highest ROC AUC mean was obtained with gboosting, and also is the best model for accuracy in the B+G combination. There was an improvement in the different techniques wil the combination of information. In Figure 2 and Figure 3 are presented the boxplots of the different machine learning techniques for ROC AUC score for B and B+G experiment, respectively. Figure 4

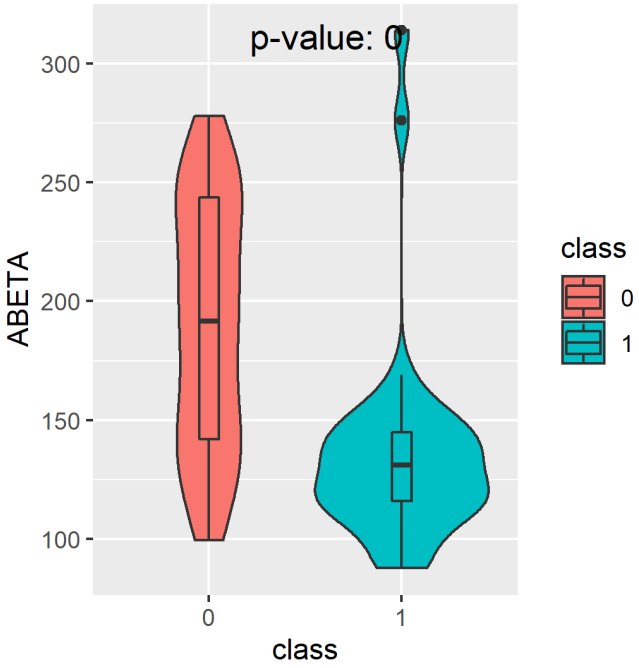

Fig. 1. Distribution of Aβ $_{1-42}$ between classes

TABLE II
SELECTED FEATURES THROUGH UNIVARIATE ANALYSIS

Feature	Type	p-value
Aβ $_{1-42}$	CSF Biomarker	$8.90x10^{-07}$
t-tau	CSF Biomarker	$1.67x10^{-06}$
p-tau$_{181}$	CSF Biomarker	$1.68x10^{-05}$
Butyric acid	Gut Metabolite	0.003
Caprylic acid	Gut Metabolite	0.005
Glyceric acid	Gut Metabolite	0.020
L-Aspartic acid	Gut Metabolite	0.025
3-Hydroxybutyric acid	Gut Metabolite	0.029
L-Tyrosine	Gut Metabolite	0.041
L-Malic acid	Gut Metabolite	0.041
Arachidonic acid	Gut Metabolite	0.045
p-Cresol Sulfate	Gut Metabolite	0.048

and 5 presents a similar information but with the accuracy score.

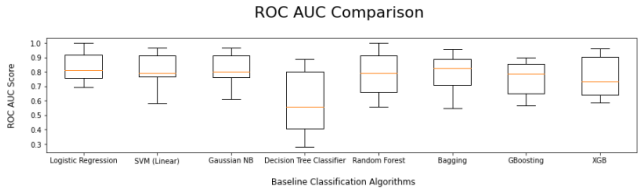

Fig. 2. Boxplots of ROC AUC scores of the different ML techniques in B experiment

E. Discussion

The study showed that gut microbiome metabolites could be used in conjunction with information from amyloid and tau biomarkers to enhance the diagnosis and prediction of AD.

TABLE III
TOP CORRELATED FEATURES WITH THE TARGET

Feature	Correlation	p-value
Aβ $_{1-42}$	-0.444	$1.61x10^{-06}$
t-tau	0.421	$6.34x10^{-06}$
p-tau$_{181}$	0.349	$2.24x10^{-04}$
3-Hydroxybutiric acid	0.294	0.002
Glyceric acid	0.278	0.004
L-Aspartic acid	0.258	0.007
Allolithocholic acid	0.228	0.018
L-Tyrosine	-0.207	0.033
Caprylic acid	-0.207	0.033
p-Cresol Sulfate	0.195	0.044

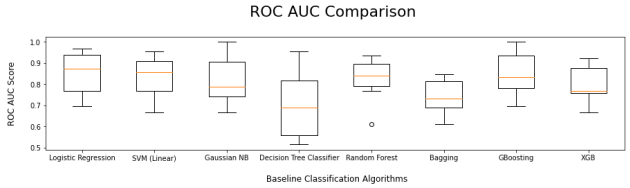

Fig. 3. Boxplots of ROC AUC scores of the different ML techniques in B+G experiment

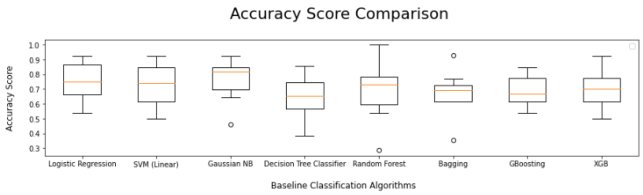

Fig. 4. Boxplots of accuracy scores of the different ML techniques in B experiment

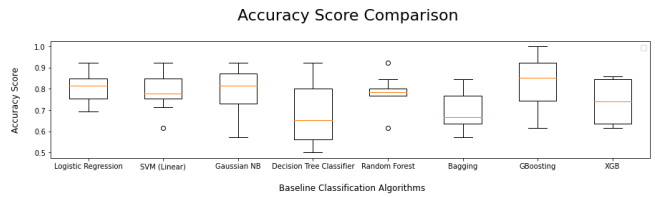

Fig. 5. Boxplots of accuracy scores of the different ML techniques in B+G experiment

This was demonstrated in the TABLE IV where we could see that there was an improvement in the different techniques with the combination of information. More precisely, all the algorithms in the B+G experiment have a better result in accuracy than in the B experiment. Similarly in the ROC AUC scores, but with the exception of Gaussian NB and Bagging where by little is better than B+G configuration.

Clinical trials focused on attending the amyloid biomarkers of the disease have not yielded definite treatment. however, the idea of analyzing gut microbiome metabolites for the study of AD give the opportunity to new treatments, as fecal microbiota transplantation, prebiotics and probiotics.

Moreover we find 9 candidate gut metabolites to study, use in diagnosis and predict AD, such as butyric acid, caprylic

TABLE IV
RESULTS OF THE MULTIVARIATE ANALYSIS

Algorithm	ROC AUC		Accuracy	
	B	B+G	B	B+G
Logistic Regression	83.34 ± 10.65	**85.13 ± 9.97**	74.93 ± 13.77	**80.43 ± 7.30**
SVM (Linear)	81.42 ± 11.72	**83.30 ± 9.71**	72.18 ± 14.38	**78.50 ± 8.82**
Gaussian NB	**81.75 ± 11.22**	81.56 ± 10.82	75.82 ± 13.89	**78.64 ± 12.33**
Decision Tree Classifier	59.04 ± 21.36	**70.50 ± 15.99**	64.42 ± 16.16	**68.41 ± 14.42**
Random Forest	78.40 ± 14.65	**82.36 ± 9.56**	68.41 ± 19.76	**78.50 ± 8.07**
Bagging	**78.25 ± 14.44**	73.76 ± 7.73	67.31 ± 15.14	**69.37 ± 8.73**
GBoosting	75.41 ± 11.98	**85.28 ± 10.31**	68.20 ± 10.59	**82.07 ± 13.26**
XGB	76.07 ± 13.77	**80.12 ± 8.47**	70.19 ± 12.07	**73.83 ± 9.89**

acid, glyceric acid, and l-aspartic acid.

These results motivate to study other sources of information with gut microbial metabolites, such as neuropsychological assessments and genetics from both human and microbiome. Also, the study can be enhanced with the inclusion of other systemb biology techniques, such as network analysis.

Finally these results with other from the literature indicate that gut microbial metabolites are a key part in the study of Alzheimer's and neurodegenerative diseases. In the following link you can consult some codes related to this project, however, the data set it can not be uploaded because it is needed permission from ADNI and ADMC (https://github.com/AlejandroTrejo97).

ACKNOWLEDGMENT

I want to thank to the Consejo Nacional de Ciencia y Tecnología (CONACYT) and Tecnológico de Monterrey, who finance my graduate studies and it is because of them that this work is possible.

REFERENCES

[1] Sosa-Ortiz AL, Acosta-Castillo I, Prince MJ. Epidemiology of Dementias and Alzheimer's Disease. Arch Med Res 2012;43:600–8. https://doi.org/10.1016/j.arcmed.2012.11.003.

[2] Roberson ED, Mucke L. 100 Years and Counting: Prospects for Defeating Alzheimer's Disease. Science (80-) 2006;314:781–4. https://doi.org/10.1126/science.1132813.

[3] Querfurth HW, LaFerla FM. Alzheimer's Disease. N Engl J Med 2010;362:329–44. https://doi.org/10.1056/NEJMra0909142.

[4] Hebert LE, Weuve J, Scherr PA, Evans DA. Alzheimer disease in the United States (2010-2050) estimated using the 2010 census. Neurology 2013;80:1778–83. https://doi.org/10.1212/WNL.0b013e31828726f5.

[5] Alzheimer's Association. 2019 Alzheimer's disease facts and figures. Alzheimer's Dement 2019;15:321–87. https://doi.org/10.1016/j.jalz.2019.01.010.

[6] Fung TC, Olson CA, Hsiao EY. Interactions between the microbiota, immune and nervous systems in health and disease. Nat Neurosci 2017;20:145–55. https://doi.org/10.1038/nn.4476.

[7] Bhattacharjee S, Lukiw WJ. Alzheimer's disease and the microbiome. Front Cell Neurosci 2013;7:1–4. https://doi.org/10.3389/fncel.2013.00153.

[8] Zhuang Z-Q, Shen L-L, Li W-W, Fu X, Zeng F, Gui L, et al. Gut Microbiota is Altered in Patients with Alzheimer's Disease. J Alzheimer's Dis 2018;63:1337–46. https://doi.org/10.3233/JAD-180176.

[9] Pistollato, F., Sumalla Cano, S., Elio, I., Masias Vergara, M., Giampieri, F., and Battino, M. (2016). Role of gut microbiota and nutrients in amyloid formation and pathogenesis of Alzheimer disease. Nutr. Rev. 74, 624–634. doi:10.1093/nutrit/nuw023

[10] Liu P-P, Xie Y, Meng X-Y, Kang J-S. History and progress of hypotheses and clinical trials for Alzheimer's disease. Signal Transduct Target Ther 2019;4:29. https://doi.org/10.1038/s41392-019-0063-8.

[11] J. Q. Trojanowski et al., "Update on the biomarker core of the Alzheimer's Disease Neuroimaging Initiative subjects," Alzheimer's Dement., vol. 6, no. 3, pp. 230–238, May 2010, doi: 10.1016/j.jalz.2010.03.008.

[12] R. C. Petersen, "Mild Cognitive Impairment," N. Engl. J. Med., vol. 364, no. 23, pp. 2227–2234, Jun. 2011, doi: 10.1056/NEJMcp0910237.