

# A Metabonomics Study on Celiac Disease by CART

Fariba Fathi<sup>1</sup>, Fatemeh Ektefa<sup>2</sup>, Kaveh Sohrabzadeh<sup>3</sup>, Afsaneh Arefi-Oskouie<sup>4</sup>, Mohsen Tafazzoli<sup>1\*</sup>, Kamran Rostami<sup>5</sup>,  
Mohammad-Reza Zali<sup>6</sup>, Mohammad Rostami-Nejad<sup>6,\*</sup>

<sup>1</sup>Department of Chemistry, Sharif University of Technology, Tehran, Iran

<sup>2</sup>Department of Chemistry, Tarbiat Modares University, Tehran, Iran

<sup>3</sup>Department of Electrical Engineer, Payam Nonprofit Higher Education Institution, Golpayegan, Iran

<sup>4</sup>Department of Basic Science Faculty of Paramedical, Shahid Beheshti University of Medical Sciences, Tehran, Iran

<sup>5</sup>Gastroenterology Department, Worcestershire Royal Hospital Worcester, UK

<sup>6</sup>Gastroenterology and Liver Disease Research center, Shahid Beheshti University of Medical Sciences, Tehran, Iran

\*Corresponding author: Tafazzoli@sharif.edu, m.rostami@gmail.com

Received April 22, 2014; Revised April 24, 2014; Accepted April 25, 2014

**Abstract** Celiac disease (CD) is an immune reaction as a consequence of ingestion of gluten. Diagnosis of CD is not easily using the clinical tests. Then, the discovery of appropriate methods for CD diagnosis is necessary. This study was concentrated to seek the metabolic biomarkers causes of CD compare to healthy subjects. In the present study, we classify CD and healthy subjects using classification and regression tree (CART). To find metabolites in serum which are helpful for the diagnosis of CD, the metabolic profiling was employed using the proton nuclear magnetic resonance spectroscopy (<sup>1</sup>H NMR). Based on CART results, it was concluded that just using one descriptor, CD and control groups could be classified separately. The 89 % of data in the test set was predicted correctly by the obtained classification model. Our study indicates that quantitative metabolite analysis of serum can be employed to distinguish healthy from CD subjects.

**Keywords:** nuclear magnetic resonance spectroscopy, celiac disease, classification and regression tree, metabonomics

**Cite This Article:** Fariba Fathi, Fatemeh Ektefa, Kaveh Sohrabzadeh, Afsaneh Arefi-Oskouie, Mohsen Tafazzoli, Kamran Rostami, Mohammad-Reza Zali, and Mohammad Rostami-Nejad, "A Metabonomics Study on Celiac Disease by CART." *International Journal of Celiac Disease*, vol. 2, no. 2 (2014): 44-46. doi: 10.12691/ijcd-2-2-3.

## 1. Introduction

Celiac disease (CD) is a digestive disease interfering with absorption of nutrients from food as consequence of damaging the small intestine [1]. Gluten, a protein found in wheat, rye, and barley and may also be found in everyday products such as medicines, vitamins, and lip balms was recognized as essential factor for CD. Diagnosis of CD is a challenging and confusing problem because some of its signs are similar to those of other diseases including irritable bowel syndrome, iron-deficiency anemia caused by menstrual blood loss, inflammatory bowel disease, diverticulitis, intestinal infections, and chronic fatigue syndrome [1]. Also, CD is suspected in people who have signs or symptoms of malabsorption or malnutrition. Based on the multivariate analysis of complex biological profiles, metabonomics is a powerful technique making disease diagnosis easier and more trustable. Metabonomics, a novel methodology arising from the post genomics era, is described as the quantitative measurement of the time-related multi-parametric metabolic response of living systems to pathophysiological stimuli or genetic modification. In clinical metabonomics study, the discovery of disease

biomarkers are necessary because the exploration of drug mechanism depend on it [2].

In the biological systems, NMR spectroscopy, typically <sup>1</sup>H NMR spectroscopy is a useful technique providing enriched information of the metabolites. <sup>1</sup>H NMR spectroscopy can be available detailed information on the concentration of the low-molecular weight metabolites in body fluids such as serum, plasma and urine in a single experiment [3,4,5].

Problems of classification or categorical dependent variables can be solved by CART. In this case the method builds a decision tree, describing a response variable as a function of different explanatory variables. In this study we evaluated the use of classification and regression tree (CART) [6], a non-parametric statistical technique, to classify an easily interpretable predictive model and seek the significance of metabolic biomarkers to distinguish between CD and healthy subjects.

## 2. Methods and Materials

### 2.1. Data

Thirty adult patients (14 males and 16 females with mean age of 33 ± 10 years (mean ± S.D.)) diagnosed with CD and referred to the Gastroenterology and Liver

Diseases Research Center, Shahid Beheshti University of Medical Sciences, participated in this study. Also 30 healthy subjects (15 males and 15 females with age of  $35 \pm 12$  years (mean  $\pm$  S.D.)) served as control group. In this metabolic profile, the healthy subjects were matched with CD subjects by gender and age [7]. The study content was explained for participants and a written consent was signed by all of them. All of candidates who entered to study had not previous significant medical history including hypertension, diabetes mellitus or hyperlipidemia. Then the serum samples of participants were collected and stored at  $-80^\circ\text{C}$  until the time NMR spectroscopy data analysis.

## 2.2. Data Preprocessing

Processing of  $^1\text{H}$  NMR spectra was performed at 500.13 MHz  $^1\text{H}$  resonance frequency on a Bruker DRX500 spectrometer. An exponential line-broadening of 0.3 Hz was applied to the free induction decay (FID) before Fourier transformation. All acquired spectra were referenced to the  $\text{CH}_3$  chemical shift of lactate at 1.33ppm [8]. The spectra divided into 0.04 ppm regions, so-called bins for reducing data of the  $^1\text{H}$  NMR spectra using ProMetab software (version prometab\_v3\_3) [9] in MATLAB (version 6.5.1, The Math works, Cambridge, U.K.). Each bin was then integrated to obtain the total signal intensity. The region from 0.2–10ppm, except for the region comprising the water signal (4.5– 6.0 ppm), was used for analysis [7]. Before further data analysis, normalization to total intensity of the spectrum was performed. For all the serum samples,  $^1\text{H}$  NMR spectra were also achieve using Carr-Purcell- Meiboom-Gill (CPMG) pulse sequence with suppression of water resonance by Presaturation to remove the broad resonances arising from macromolecules [10].

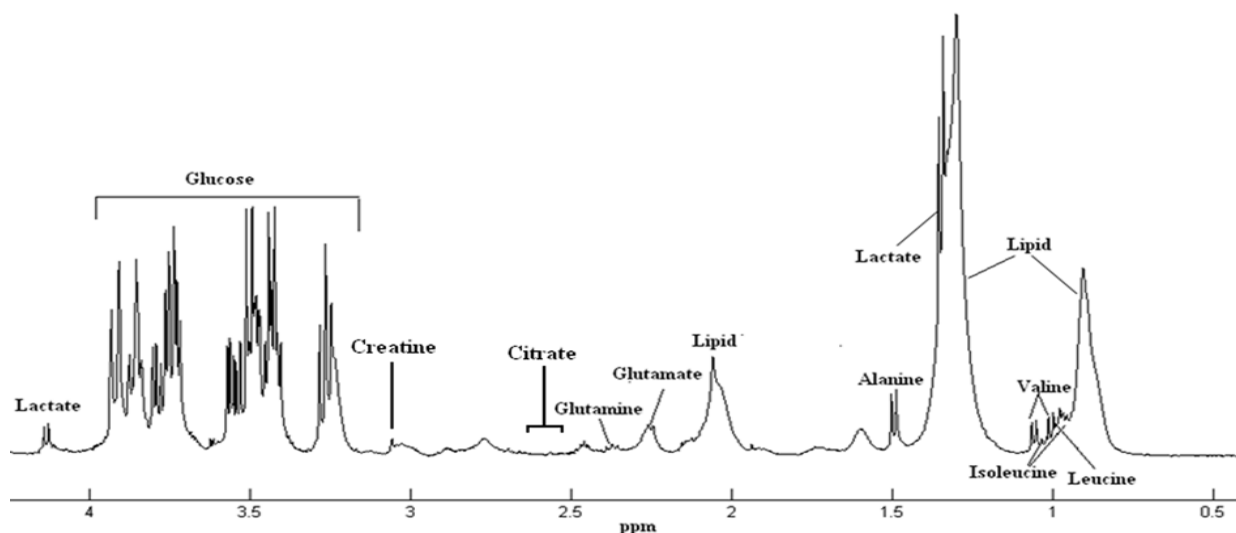
## 2.3. Chemometrics

### 2.3.1. Classification and Regression Tree (CART)

Classification and regression trees (CART) as a non-parametric methodology was introduced firstly by Breiman and colleagues in 1984 [11,12]. CART is a tree-building method and data are split repeatedly into groups [11,12]. In CART method a response  $Y$  is explained by selecting some independent variables  $X$  from a larger set of  $X$  values. The tree is built in a recursive binary way and the branches connect nodes. A node could be divided into two new nodes. Primitive node and the two new ones are called as parent and childnodes. Terminal nodes are nodes without child nodes. Based on the values of its  $X$  variables, a new sample (test set) is allocated to a terminal node and the class of this sample was determined.

## 3. Results

Figure 1 shows human blood serum spectra assigning by comparison of chemical shifts in a representative spectrum to established libraries reported in the literature, the Human Metabolome Data Base (HMDB). To demonstrate the prediction capability of the classes of test set, about 1/3 of the samples (18 samples) are employed as test set not used in the training set [13]. The CART method was employed for classifying the training set (42 samples). Descriptive variables are the integral at the different the chemical shift in NMR spectra while responses are the class numbers of the different samples. Figure 2 shows CART diagram in which just using one descriptor CD and control groups could be classified separately.



**Figure 1.** Typical 500MHz  $^1\text{H}$ NMR spectra of control human blood serum. Eleven metabolites viz., lipid, leucine/isoleucine, valine, lactate, alanine, glutamine, glutamate, citrate, creatine and glucose were identified

The most important metabolite lactate was recognized using CART method. As shown in Figure 2, if the level of metabolite lactate is lower than 0.1026, the result will be belonged to class 1(CD) and respectively if it is above 0.1026 will be belonged to class 2 (control). This result shows the reduction of lactate level in blood serum of patients compared to healthy individuals. Table 1

illustrated characteristics of descriptor chosen based on CART method. Performance of classification model is commonly evaluated using the data in a confusion matrix. This matrix contains information about actual and predicted classifications. Table 2 contains the confusion matrix for the training and test set. Prediction accuracy, sensitivity, and specificity of 80%, 100%, and 89%,

respectively, are given in detecting CD patients of external test set. Table 3 presents other classification parameters such as error rate and non-error rate.

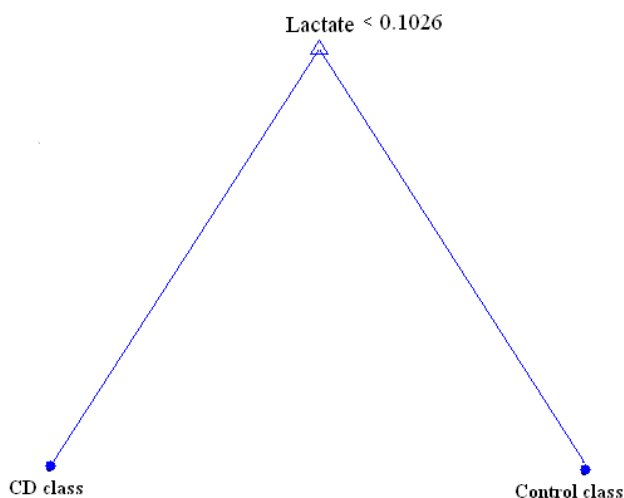


Figure 2. Classification tree viewer

Table 1. Specifications of the selected CART descriptor

Descriptors	Assignment	<sup>1</sup> H chemical shift (ppm)
1) Lactate	βCH3	1.32

Table 2. Confusion matrix for training and test set

	Observed	Predicted	
		CD class	Healthy class
Training set	CD class	21	0
	Healthy class	0	21
Test set	CD class	7	2
	Healthy class	0	9

Table 3. The calculated error and non-error rates of the classification index and the classification performances of training and test sets

	Error rate	Non-error rate	specificity	sensitivity	accuracy
Training set	0	1	1	1	1
Test set	0.11	0.89	0.82	1	0.89

## 4. Discussion

The results of this study show that the level of lactate in blood serum of CD patients is lower than healthy subjects. Glucose and muscle glycogen apply in the cell for the duration of the requirements of high-intensity exercise. The two molecules of pyruvate produce in the last step of glucose breakdown. With absorption of two protons into structure of pyruvate, lactate is produced. It has been estimated that about 50% of the lactate produced during intensive exercise is used directly as a metabolic fuel by muscle, and by the liver to produce blood glucose and glycogen (Cori Cycle). Lactic acid is converted to carbon dioxide (65%) and water, glycogen (20%), Protein (10%), glucose (5%).

Bertini and et al. [14] expressed that lactate should be decreased in CD. According to what mentioned about

pyruvate, the analysis of metabolic profile points to pyruvate as the key to elucidate the origin of this peculiar metabolomics signature of CD. Higher levels of glucose can be related to a variety of causes. Due to metabolic pathway of pyruvate a decrease of pyruvate sera levels is consistent with the fact that in untreated CD patients, glycolysis is somehow impaired by reduction of glucose intake at cellular level or by impairment of one or more steps in the glycolysis process itself. Impairment of glycolysis explains both a lowering of pyruvate and lactate levels and an increase of glucose levels in blood.

In this study, we evaluated the application of CART to build a predictive model to differentiate between CD and healthy controls. Due to the results CART proved to be quite powerful in discriminating between CD and healthy subjects. In conclusion, metabolomics and analysis of these important metabolites in serum is applied widely in early stage of disease. Therefore, further investigations are required to establish its real usefulness in clinical practice.

## References

- [1] Rostami Nejad M, Rostami K, Pourhoseingholi M, Nazemalhosseini E, Dabiri H, Habibi M. Atypical presentation is dominant and typical for Celiac Disease. *J Gastrointest Liver Dis*, 2009; 18: 285-91.
- [2] Nicholson J, Lindon J, Holmes. E. Metabolomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica*, 1999; 29: 1181-9.
- [3] Hong Y-S, Hong KS, Park M-H, Ahn Y-T, Lee J-H, Huh C-S, et al. Metabonomic Understanding of Probiotic Effects in Humans With Irritable Bowel Syndrome. *J Clin Gastroenterol*, 2011(45): 415-25.
- [4] Bezabeh T, Somorjai RL, Smith ICP. MR metabolomics of fecal extracts: applications in the study of bowel diseases. *Magn Reson Chem*, 2009; 47: 54-61.
- [5] Gall GL, Noor SO, Ridgway K, Scovell L, Jamieson C, Johnson IT, et al. Metabolomics of Fecal Extracts Detects Altered Metabolic Activity of Gut Microbiota in Ulcerative Colitis and Irritable Bowel Syndrome. *J Proteome Res*, 2011; 10: 4208-18.
- [6] Breiman L, Friedman J, Olshen R, Stone C. Classification and Regression Trees. Chapman & Hall (Wadsworth I, editor. New York; 1984.
- [7] Fathi F, Tafazzoli M, Mehrpour M, Shahidi G. Study of metabolic profiling Parkinson's disease. *HealthMED*, 2013; 7: 204-10.
- [8] Fathi F, Kyani A, Rostami-Nejad M, Rezaye-Tavirani M, Naderi N, Zali MR, et al. A metabolomics study on Crohn's Disease using Nuclear Magnetic Resonance spectroscopy. *HealthMED*, 2012; 6: 3577-84.
- [9] Viant MR. Improved methods for the acquisition and interpretation of NMR metabolomic data. *Biochem Biophys Res Commun*, 2003; 310: 943-8.
- [10] Zhang G, Hirasaki G. CPMG relaxation by diffusion with constant magnetic field gradient in a restricted geometry: numerical simulation and application. *J Magn Reson*, 2003; 163: 81-91.
- [11] Zhang MH, Xu QS, Daeyaert F, Lewi PJ, Massart DL. Application of boosting to classification problems in chemometrics. *Anal Chim Acta*, 2005; 544: 167-76.
- [12] Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and Regression Trees. Group WI, editor: Belmont; 1984.
- [13] Fathi F, Kyani A, Darvizeh F, Mehrpour M, Tafazzoli M, Shahidi G. Relationship Between Serum Level of Selenium and Metabolites Using <sup>1</sup>HNMR-Based Metabolomics in Parkinson's Disease. *Appl Magn Reson*, 2013; 44: 1-14.
- [14] Bertini I, Calabro A, Carli VD, Luchinat C, Nepi S, Porfirio B, et al. The Metabonomic Signature of Celiac Disease. *J Proteome Res*, 2009; 8: 170-7.