Research article

MRC

Received: 26 September 2013 Revised: 14 February 2014 Accepted: 20 March 2014 Published online in Wiley Online Library: 22 April 2014

(wileyonlinelibrary.com) DOI 10.1002/mrc.4074

^{1}H NMR based metabolic profiling in Crohn's disease by random forest methodology

Fariba Fathi,[a] Laleh Majari-Kasmaee,[a] Ahmad Mani-Varnosfaderani,[b] Anahita Kyani,[c†] Mohammad Rostami-Nejad,[d] Kaveh Sohrabzadeh,[e††] Nosratollah Naderi,[d] Mohammad Reza Zali,[d] Mostafa Rezaei-Tavirani,[f†††] Mohsen Tafazzoli[a]* and Afsaneh Arefi-Oskouie[f]*

The present study was designed to search for metabolic biomarkers and their correlation with serum zinc in Crohn's disease patients. Crohn's disease (CD) is a form of inflammatory bowel disease that may affect any part of the gastrointestinal tract and can be difficult to diagnose using the clinical tests. Thus, introduction of a novel diagnostic method would be a major step towards CD treatment.

Proton nuclear magnetic resonance spectroscopy (^{1}H NMR) was employed for metabolic profiling to find out which metabolites in the serum have meaningful significance in the diagnosis of CD. CD and healthy subjects were correctly classified using random forest methodology. The classification model for the external test set showed a 94% correct classification of CD and healthy subjects. The present study suggests Valine and Isoleucine as differentiating metabolites for CD diagnosis. These metabolites can be used for screening of risky samples at the early stages of CD diagnoses.

Moreover, a robust random forest regression model with good prediction outcomes was developed for correlating serum zinc level and metabolite concentrations. The regression model showed the correlation (R^2) and root mean square error values of 0.83 and 6.44, respectively. This model suggests valuable clues for understanding the mechanism of zinc deficiency in CD patients.

Keywords: nuclear magnetic resonance spectroscopy; metabonomics; zinc; Crohn's disease; random forest

Introduction

Crohn's disease (CD) is one of the two major subtypes of inflammatory bowel disease (IBD) that cause chronic inflammation of the intestinal tract.[1–3] While the pathophysiology of IBD is not fully understood, it has been widely accepted that multiple components, including environmental factors, diet, smoking habits, hormone levels, drug usage, and genetics contribute to the occurrence and perpetuation of this disease. Although the prevalence and incidence of IBD are stabilizing in high-incidence areas such as northern Europe and North America, in low-incidence areas such as southern Europe, Asia, and the developing world, they continue to rise.[4]

Crohn's disease often mimics other symptoms, hence correct identification of CD in some cases may be complicated.[5] To reach the correct diagnosis, clinical tests including endoscopic, histological, and radiologic techniques are applied. These methods can be time consuming and costly. With respect to these problems, metabonomics is an important technique for identification of biomarkers for early diseases detection.[6]

Metabonomics is defined as 'the quantitative measurement of the dynamic multi-parametric response of living systems to pathophysiological stimuli or genetic modification'.[7] If biological variations in the target group are meaningfully different from those in the control group, quantitative analysis of metabolite can provide important results.[8]

Proton nuclear magnetic resonance spectroscopy (^{1}H NMR) is one of the most commonly applied techniques to obtain vital information from complex and unprocessed biological samples. It provides quantitative and reproducible information with little

* Correspondence to: Mohsen Tafazzoli, Department of Chemistry, Sharif University of Technology, P.O. Box 11155-9516, Tehran, Iran. E-mail: tafazzoli@sharif.edu

** Correspondence to: Afsaneh Arefi Oskouie, Department of Basic Science Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, P.O. Box: 19395-4618.Tehran, Iran. E-mail: a_arefioskouie@yahoo.com

† Present address: Department of Medicinal Chemistry, College of Pharmacy and Translational Oncology Program, University of Michigan, Ann Arbor, MI

†† Present address: Department of Electrical Engineer, Payam Nonprofit Higher Education Institution, Golpayegan, Iran

††† Present address: Proteomics Research Center, Facultyof Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran

a Department of Chemistry, Sharif University of Technology, Tehran, Iran

b Department of Chemistry, Faculty of Sciences, Tarbiat Modares University, Tehran, Iran

c Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran

d Gastroenterology and Liver Disease Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran

e Department of Electrical Engineer, Golpayegan Payam University, Tehran, Iran

f Department of Basic Science Faculty of Paramedical, Shahid Beheshti University of Medical Sciences, Tehran, Iran

sample preparation, and hence it is widely used to build metabolic profiles in diverse metabolic studies.[9,10] Many gastrointestinal diseases have been diagnosed by investigation of biofluid and tissue samples using ^{1}H NMR-based metabonomics.[11–13]

NMR spectroscopy in combination with suitable statistical analysis have been successfully used to distinguish patients with IBD from healthy subject through profiling of fecal extracts,[14] biopsy,[15] and urine samples.[16,17] Investigation of the serum metabolic profile in CD subjects is rare. So far, the metabolite profile of CD patients' serum has been investigated in only one study.[18] Schicho and co-authors performed metabolite analysis of serum and plasma on IBD subjects. They obtained regular one-dimensional proton NMR spectra using a standard pulse sequence (Bruker pulse program prnoesy1d).[18] In this study, we chose Carr-Purcell-Meiboom-Gill (CPMG) spin echo pulse sequent as a different method for metabolic profiling of serum in CD subjects.

Zinc is an essential trace element with various biological functions in a large variety of enzymes, depending on the structural and/or catalytic roles it plays. Zinc participates in free radicals scavenging. As a consequence, it may stop the progress of the gastrointestinal disease and halting the inflammatory process. Also, it plays an important role as a powerful anti-inflammatory agent to treat gut inflammation. Decline in Cu–Zn super oxide dismutase activity was reported in CD.[19] Finally, this trace element is lost through diarrhea, and zinc deficiency, while unusual, can occur in individuals who have IBD, particularly the patients with chronic diarrhea.[20,21]

The primary goal of this study was to classify the control cohorts, and the subjects identified clinically, endoscopically, and histologically to have active CD. Then, a regression model between the serum zinc level and metabolites was developed. Random forest (RF) was employed as a powerful classification method; it was used to inspect the differences in the serum zinc levels of healthy and CD samples. Based on the results, RF modeling can be a versatile alternative technique for analyzing the metabonomics data although extensive studies will be needed to verify our present findings.

Materials and Methods

Sample collection

Twenty-six adult patients (with mean age of 34 ± 11 years), diagnosed with CD, were recruited from Gastroenterology and Liver Disease Research Center, Shahid Beheshti University of Medical Sciences. These patients had been diagnosed with CD by experienced gastroenterologists on the basis of radiographic, experimental, and often colonoscopy criteria. Twenty-nine subjects (with mean age of 35 ± 12 years) were enrolled as the control group. Healthy control subjects were matched for gender and age to CD subjects. None of the participants in this study had any other significant past medical history such as hypertension, diabetes mellitus, or hyperlipidemia.

Serum samples were drawn from peripheral veins of patients and healthy subjects in the morning after a 12 h fast. Whole blood samples were collected in vacutainer tubes containing no anticoagulant. These vials were shaken thoroughly and incubated in upright position at room temperature for 30–45 min to allow coagulation. The clotted samples were then centrifuged for 15 min at 2500 rpm. Next, the sera were carefully aspirated and collected in fresh polypropylene tubes, up to three-quarter of tubes capacity. Any turbid sample was centrifuged and aspirated again to separate insoluble particles. The sera were stored, for further analysis at −80 °C.

Atomic absorption spectrometry

Serum zinc level was measured using an atomic absorption spectrophotometer (PERKLIN ELMER 400). Serum was diluted with distilled water (dilution ratio of 1 : 5). The source of light was monochromatic light with wavelength of 213.9 nm and slit-width of 1 mm. Acetylene gas was used in the burner of atomic absorption spectrophotometer. Standard solutions (Sigma Aldrich) containing 25, 50, and 100 mg dl^{-1} of zinc were used for standardization and calibration. The diluted samples were analyzed in triplicates in serial order. Atomic absorption spectrometry measurements show that CD group's zinc serum level was lower than the control group. The mean observed serum zinc level in CD and control groups were (70 ± 6 μg/l) and (92 ± 9 μg/l), respectively (p-value < 0.001).

^{1}H NMR spectroscopy

^{1}H NMR spectroscopy experiments were performed on a 500 MHz Bruker DRX spectrometer. The spectrometer was operating at 500.13 MHz. Five-millimeter high-quality NMR tubes (Sigma Aldrich, RSA) were used. For the NMR spectrometer, NMR lock signal was provided applying 100 μl of D_2O (Deuterium oxide, 99.9%D, Aldrich Chemicals Company). Combination of high molecular weight components causes broad resonances, on which super imposed sharper resonances from the low molecular weight species are arising (e.g., amino acids and carboxylic acids). CPMG experiment results in the suppression of the broader elements and thus enhances visualization of the low molecular weight metabolites and assuages the broad signals of protein and lipoprotein molecules.[22] CPMG spin echo pulse sequent was employed to record 1D ^{1}H NMR spectra of the samples. Spectra were recorded at 298 K, and other acquisition parameters were spectral width: 8389.26 Hz; time domain points: 32 K; number of scans: 154; acquisition time: 2 s; spectrum size: 32 K; and line broadening: 0.3 Hz.

Data pre-processing

All ^{1}H NMR spectra of serum samples were manually phased and baseline-corrected applying XWINNMR (version 3.5, Bruker Spectrospin Ltd). Using XWINNMR, peaks in the serum spectra were referenced to the chemical shift of lactate at $\delta = 1.33$. ^{1}H NMR spectra processing was performed using PROMETAB software (version prometab_v3_3) in MATLAB (version 6.5.1, The Mathworks, Cambridge, UK). Using this software, the region 0.2-10.0 ppm of the CPMG spectra was reduced into integrated bins of equal width (0.04 ppm).

The spectral region between 4 and 5.5 ppm, corresponding to water signal, was excluded. Prior to further data analysis, the integral values of each spectrum were normalized to a constant sum of all the integrals in that spectrum to decrease any significant concentration differences between samples.[23,24] Preceding multivariate analysis, the data have been mean centered using the procedure that is explained in the literature.[25] Then, the metabolites were assigned based on previous studies.[26–28]

Statistical analysis

Random forest

Random forest [29,30] is a collection of hundreds of identically distributed decision trees.[31] These trees are grown using a classification algorithm such as classification and regression trees. A

basic RF is formed through random selection of a small group of input variables to split on. This random selection is carried out at each node, and the size of the group is fixed while the forest is growing. Each tree in the RF is grown on a bootstrap replicate of the learning sample. Generally, the bootstrap samples are selected from two thirds of the original samples. The remaining one third are called out-of-bag samples (OOB samples). The OOB cases are employed to get a run-time impartial estimate of the classification and regression errors as trees are added to the forest. To classify an input, first each tree classifies it individually. Then, the forest chooses the class in which the given input has the majority vote. Next, the proportion of votes for each class is calculated. When the test set is entered to the forest, these proportions are also computed. The margin of a case is the proportion of votes for the true class minus the maximum proportion of votes for the other classes.

Random forest classifiers have some advantages over simple discriminatory techniques such as linear discriminant analysis. The RF models can handle large datasets without direct variable deletion. Moreover, the RF classifiers use majority voting strategy for decision making. If a classifier makes a mistake, other classifiers may detect the miss classified sample. The RF models can determine the importance of independent variables in modeling procedure. It helps for identification of discriminatory variables and biomarkers. In addition, the RF classifiers can approximate the missing data values in large datasets. Regarding the mentioned properties, the RF method has been used in this work for analysis of the collected data.

In this study, we used the RF package in MATLAB (version 6.5.1, The Mathworks, Cambridge, U.K.) to perform statistical computing and create graphics, for both classification and regression models. The detailed description about RF technique can be found in reference.[32]

Results

Classification using RF

The most important metabolites in serum were selected based on previous studies.[26–28] The studied metabolites are amino acids (alanine, glutamine, leucine/isoleucine, lysine and valine), organic acids (lactate and creatine), lipid, and glucose (Fig. 1).

In order to classify CD and healthy subjects, the data set was divided into two parts, training and test sets. The training set was used to build a model and identify the most relevant metabolites. In order to test the predictive ability of the classification model, test set was employed. Approximately, 30% of the patient and normal samples have been randomly selected as test set. Consequently, the training and test sets were composed of 39 and 16 ^{1}H NMR spectra, respectively. In order to reduce the risk of over-fitting, the test set was not used to make the model. Samples of the training set were classified using RF in which 500 trees were grown. The OOB data were used to estimate the prediction accuracy of classification. Figure 2 presents the OOB error rate. The frequencies of selection of ^{1}H NMR chemical shifts by 500 grown trees are shown in Fig. 3. This frequency profile shows that

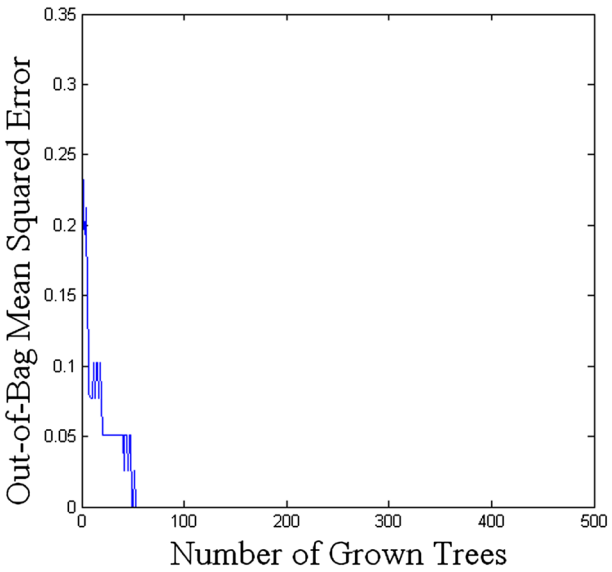

Figure 2. Plot of out-of-bag (OOB) error for random forest classification of Crohn's disease and control groups. The OOB data were used to estimate the prediction accuracy of classification.

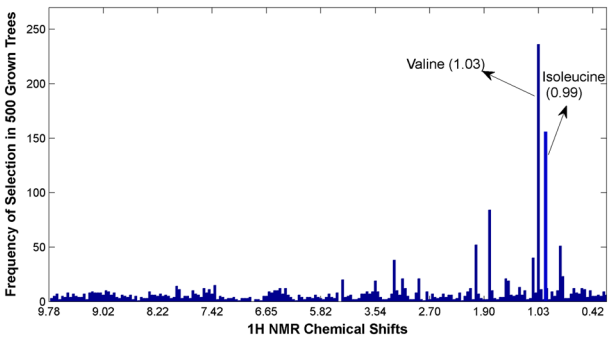

Figure 3. The frequencies of selection of ^{1}H NMR chemical shifts after 500 trees were grown by random forest methodology.

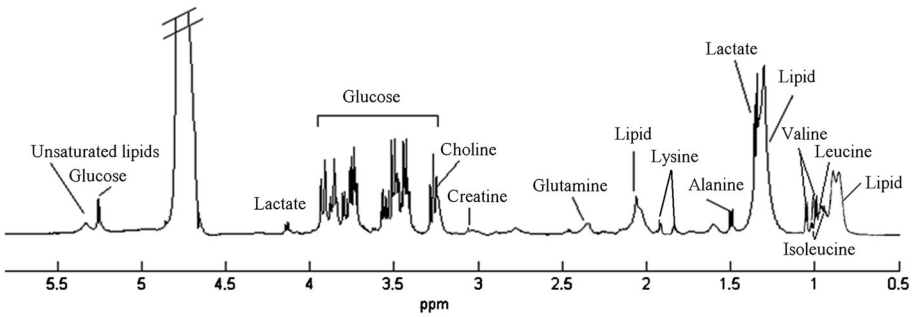

Figure 1. Typical 500 MHz ^{1}H NMR spectrum of control human blood serum.

two chemical shifts of 0.99 and 1.03 have considerable impact for discriminating patient and normal samples. ^{1}H chemical shifts at 1.03 and 0.99 ppm are assigned to δCH_3 (valine) and $\beta CH3$ (isoleucine), respectively. The distribution of the area under peak values for Valine and Isoleucine metabolites in normal and patient samples are shown in Fig. 4a and 4b. As can be seen in these figures, the distributions of these metabolites completely differ for normal and patient samples (p-value < 0.0001). CD group's valine and isoleucine levels were lower and higher than those of healthy cohort, respectively. The data in Figs 3 and 4 show that the RF methodology could extract important metabolites from large number of them and can help for pattern recognition purposes in Metabolomic studies.

A confusion matrix, including knowledge about the number of correct and incorrect predictions, was compared with the real outcomes by a classification model. Performance of a classification model is commonly evaluated using the data in this matrix. Table 1 shows the confusion matrix for a two-class classifier. The summary of the classification parameters is also shown in Table 2. As shown in Table 2, RF model has an accuracy of 0.94 in detecting CD patients in the external test set. These results show that RF classification model is successful in CD diagnosis.

The area under ROC curve is used to measure the quality of the classification models. A random classifier has an area under the curve (AUC) of 0.5, whereas AUC for a perfect classifier is equal to 1. In practice, most of the classification models have an AUC between 0.5 and 1.[33]

The obtained values of AUC for the training and test sets are 1 and 0.94, respectively. The high AUC score of the proposed model for the samples in the external test set is another evidence that RF model has high capability to detect CD.

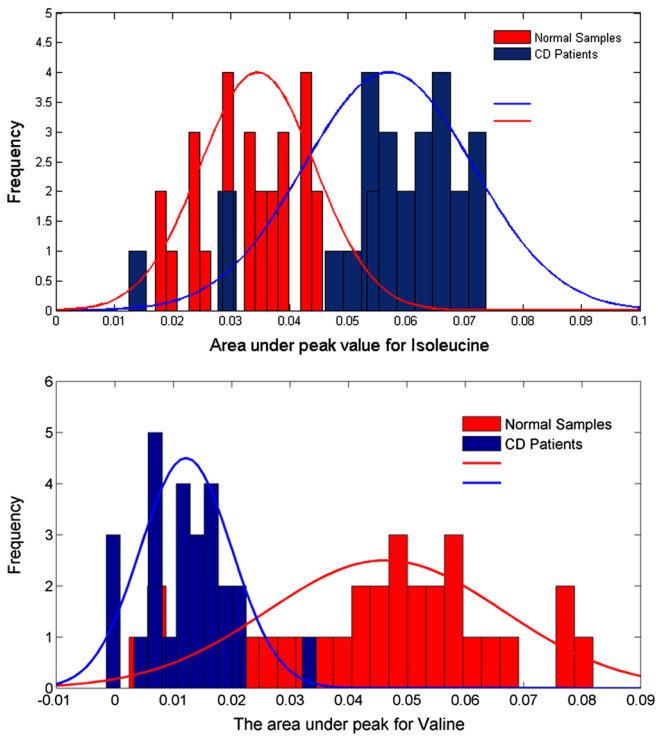

Figure 4. The distribution of the area under peak values for (a) Valine and (b) Isoleucine metabolites in normal and patient sample.

Regression model of blood zinc

In order to predict zinc level based on NMR data, a regression model between the serum zinc level and metabolites was developed. Then, the effective metabolites on zinc level in CD subjects were identified.

Random forest regression modeling was also performed using NMR spectra and serum zinc level. Similar to the classification process, data were divided into training and test sets. The test set contained about 30% of the samples. Then, the orthogonal signal correction[34,35] was performed on the training set to remove the bilinear components in X matrix (NMR spectra) that are orthogonal to Y. This approach makes a signal correction that does not remove information from X.

To build the RF model, zinc level in serum was used as dependent variable. The model was constructed by 500 trees. Figure 5 illustrates the convergence of the RF algorithm. This figure depicts the OOB error as a function of the number of trees that were used for the training procedure. As Fig. 5 represents, the OOB error became most stable when the number of trees reached about 200 and the RF model has reached its optimal classification error.

The model was first validated internally using the same training set that had been employed for the model generation (39 samples). The plot of predicted RF values of zinc concentration against the experimental ones is shown in Fig. 6. Correlation (R^2) value and root mean square error (RMSEs) of 0.94 and 4.10 were obtained for the training data set, respectively. The correlation coefficient (R) measures the adequacy of the model for predicting the dependent variable in a regression analysis. R^2, ($0 \le R^2 \le 1$), is the square of the correlation between experimental and predicted response values. Here, R^2 value is very close to 1, indicating that the RF model explains majority of variability in data. RMSE is the standard deviation of the differences between predicted and experimental values.

The residuals of calculated zinc concentration values are plotted against the experimental ones in Fig. 7. The propagation of the residuals in both sides of zero line shows that no systematic error exists in the development of RF model. In order to examine the predictive power of the RF model, the zinc concentration values of test data were predicted using the proposed model.

Table 1. Confusion matrix for training and test set

		Predicted	
	Observed	CD class	Healthy class
Training set	CD class	18	0
	Healthy class	0	21
Test set	CD class	7	1
	Healthy class	0	8

CD, Crohn's disease.

Table 2. 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	0.06	0.94	0.88	1	0.94

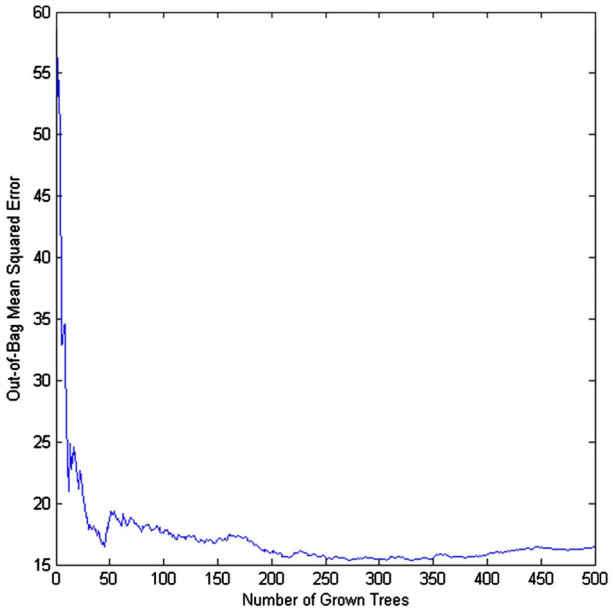

Figure 5. Plot of out-of-bag (OOB) error for random forest regression of serum zinc levels. The OOB error became most stable when the number of trees reached about 200.

The test data set included 16 samples and had no role in the model-building phase. The correlation (R^2) value and RMSEs are 0.83 and 6.44, respectively. Therefore, when applied to new samples, this RF model has good prediction capability. The experimental and predicted zinc concentration values for the training and test sets are given in Table 3.

The RF model was further evaluated through applying the Y-randomization test. Several random shuffles of Y (zinc concentration) were chosen and features were selected using RF feature importance plot and then modeling process was performed for all the cases. All sets showed the low R^2 values for both training and test data indicate that the good statistical results of RF model are not because of a chance correlation or sample dependence of the training set.[36]

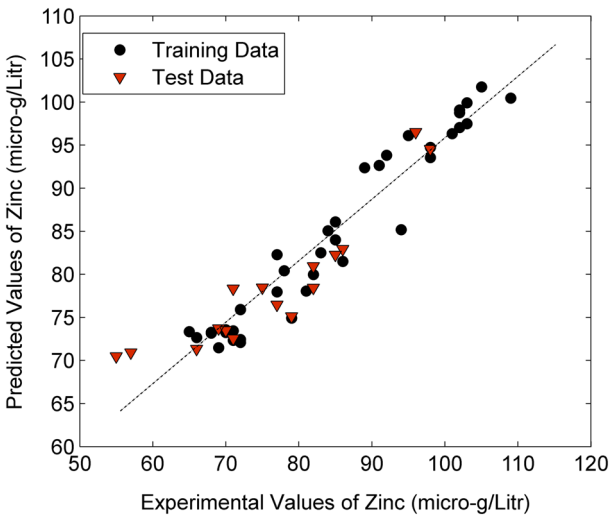

Figure 6. Plot of predicted values of serum zinc levels using random forest model against the experimental ones for the training and test sets.

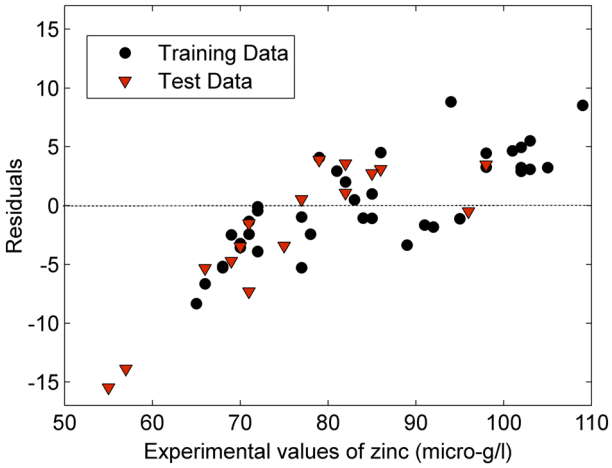

Figure 7. Plot of residuals against the experimental values of serum zinc levels for the training and test sets.

By applying RF to the data set, as a regression method, two important metabolites (glutamine and lysine) were identified. The corresponding ^{1}H chemical shifts were at 2.41, 1.88, and 1.72 ppm. The ^{1}H chemical shift at 2.41 ppm is assigned to γCH_2 (glutamine) and the ^{1}H chemical shifts at 1.88 and 1.72 ppm are related to βCH_2 and δCH_2 (lysine). The data in this section shows that the level of glutamine and lysine are important factors related to serum zinc level. This information can help for unraveling the mechanism of zinc deficiency in CD patients. We should emphasize that this finding shed some light on the effect of zinc serum level in CD patients; however, making a serious decision on this important relationship needs more experiments and proofs in clinical studies.

Discussion

Based on previous studies, the significance of the selected metabolites selected by the RF classification model and their concentration changes in CD patients is scrutinized subsequently.

As discussed in the Results section, RF classification confirms that in CD patients, isoleucine and valine levels are higher and lower than healthy subjects, respectively. The catabolism of isoleucine and valine initiates in the muscles and yields NADH, which can be utilized for ATP generation.[37] CD patients are at high risk of developing nutritional status impairment, so the amino acid level changes in CD appear to be rational.[38] Decreased function of the intestinal mucosa correlates with malnutrition. Therefore, the assessment of nutritional status and energy requirements plays an important role in the management and follow-up of CD.[38] Some of the increased amino acids were also reported to be increased in fecal extracts.[14]

Based on Schicho and co-authors' findings, CD have an impact on amino acid metabolism; in comparison with the control cohort, CD patients had higher isoleucine and lower valine levels.[18] Our observations are in agreement with Schicho et al. study.

In order to investigate the effect of metabolites on serum zinc level, RF regression modeling was performed. Zinc is the second most prevalent trace element in the human body. In recent years, the scientific community have registered an explosion of interest about the role of zinc in human diseases.[39] Studies have confirmed an obvious decrease in the Cu/Zn super oxide dismutase expression within the intestinal mucosa of CD patients. This

Table 3. The serum Zinc levels (experimental and predicted) values for the training and test sets

	No	Zinc (observed) μg/l	Zinc (predicted) μg/l
Training set	55	94	85
	11	72	76
	37	103	97
	36	101	96
	27	77	82
	22	72	72
	43	91	92
	25	70	73
	23	71	73
	52	79	75
	38	82	80
	4	71	72
	31	98	94
	7	68	73
	14	72	72
	51	86	81
	1	77	78
	5	72	75
	12	69	71
	15	66	72
	17	68	73
	18	70	73
	20	78	80
	24	65	73
	26	81	78
	29	83	83
	30	85	84
	32	102	99
	34	103	100
	35	85	86
	39	105	102
	40	95	96
	41	102	99
	42	89	92
	44	109	100
	45	92	93
	48	98	95
	49	102	97
Test set	50	84	85
	54	85	82
	8	69	73
	6	71	78
	19	77	76
	3	75	78
	10	66	71
	28	82	81
	46	98	95
	16	55	70
	2	70	73
	9	57	70
	47	82	78
	33	96	97
	21	79	75
	53	86	83
	13	71	73

finding highlighted the role of zinc deficiency in IBD.[40] Also, the disturbances in the zinc-related antioxidant cascade in the gut of CD patients have been verified by analysis.[41] It is conceivable that in CD patients, zinc deficiency could occur as a result of both malabsorption and intestinal loss of cells and plasma.

As mentioned in the Results section and based on RF regression results, there is a considerable dependency between glutamine and lysine concentrations and zinc level in the serum. The most abundant amino acid in the human body is glutamine. It plays extremely vital role in the functioning of most cells within the body such as immune cells. Glutamine is mainly utilized by the immune cells and also contributes to the proliferation of these cells.[42] CD is a disease of immune system, and CD patients are unable to generate enough glutamine to encounter their immune system's increased glutamine requirements. So, CD patients encounter chronic glutamine deficiency. This amino acid may release as much as one-third of the glutathione (GSH) when the body is under stress. GSH is the most abundant antioxidant in the body, and provides extra fuel to the injury site in muscles. Zinc deficiency causes decrease in blood GSH levels.[43]

Glutamine reduces oxidative stress and cytokine production during intestinal inflammation. Thus, this amino acid suppresses the inflammatory response.[44] Glutamine is as a major respiratory fuel for enterocytes and gut-associated immune cells.

Aiken *et al*. declared that a significant proportion of serum zinc exists as a complex form with amino acids.[45] Lysine is an essential amino acid essential for a variety of metabolic roles and can be a ligand for zinc. Formation of zinc-lysine complex increases the absorption of zinc in the small intestine.[46]

The findings of this study may open doors for further investigations regarding zinc deficiency. More clinical analysis for further research is required.

In conclusion, this study shows that quantitative metabolite profile of CD patients' serum can be used to distinguish between healthy and CD subjects. Furthermore, our findings confirm previous studies, and these results are of great importance to detect biomarkers in the future. In the second part of the study, as a novel approach, NMR data and RF were employed to indicate the biomarkers and biologically important variables that are correlated with zinc level in CD subjects' serum. Generally, the proposed approach in this work is an alternative and less invasive method to detect CD compared with clinical tests.

Acknowledgements

We gratefully acknowledge financial support from Iran National Science Foundation, Sharif University of Technology, and Shahid Beheshti University of Medical Sciences.

References

[1] D. Baumgart, S. Carding. *Lancet* **2007**, *369*, 1627–1640.
[2] D. Baumgart, W. Sandborn. *Lancet* **2007**, *369*, 1641–1657.
[3] R. Xavier, D. Podolsky. *Nature* **2007**, *448*, 427–434.
[4] E. V. J. Loftus. *Gastroenterology* **2004**, *126*, 1504–1517.
[5] B. H. Stephen, W. Sandborn. *Am. J. Gastroenterol.* **2001**, *96*, 635–643.
[6] J. B. German, B. D. Hammock, S. M. Watkins. *Metabolomics* **2005**, *1*, 3–9.
[7] J. Nicholson, J. Lindon, E. Holmes. *Xenobiotica* **1999**, *29*, 1181–1189.
[8] J. Nicholson, J. Connelly, J. Lindon, E. Holmes. *Nat. Rev. Drug Discov.* **2002**, *1*, 153–161.
[9] F. Fathi, A. Kyani, F. Darvizeh, M. Mehrpour, M. Tafazzoli, G. Shahidi. *Appl. Magn. Reson.* **2013**, *44*, 721–734.

[10] M. Mehrpour, A. Kyani, M. Tafazzoli, F. Fathi, M.-T. Joghataie. *Magn. Reson. Chem.* **2013**, *51*, 102–109.
[11] Y.-S. Hong, K. S. Hong, M.-H. Park, Y.-T. Ahn, J.-H. Lee, C.-S. Huh, J. Lee, I.-K. Kim, G.-S. Hwang, J. S. Kim. *J. Clin. Gastroenterol.* **2011**, *11*, 415–425.
[12] T. Bezabeh, R. L. Somorjai, I. C. P. Smith. *Magn. Reson. Chem.* **2009**, *47*, 54–61.
[13] G. L. Gall, S. O. Noor, K. Ridgway, L. Scovell, C. Jamieson, I. T. Johnson, I. J. Colquhoun, E. K. Kemsley, A. Narbad. *J. Proteome Res.* **2011**, *10*, 4208–4218.
[14] J. R. Marchesi, E. Holmes, F. Khan, S. Kochhar, P. Scanlan, F. Shanahan, I. D. Wilson, Y. Wang. *J. Proteome Res.* **2007**, *6*, 546–551.
[15] T. Bezabeh, R. L. Somorjai, I. C. Smith, A. E. Nikulin, B. Dolenko, C. N. Bernstein. *Am. J. Gastroenterol.* **2001**, *96*, 442–448.
[16] N. S. Stephens, J. Siffledeen, X. Su, T. B. Murdoch, R. N. Fedorak, C. M. Slupsky. *J. Crohns Colitis.* **2012**, *7*, 42–48.
[17] H. R. Williams, I. J. Cox, D. G. Walker, B. V. North, V. M. Patel, S. E. Marshall, D. P. Jewell, S. Ghosh, H. J. Thomas, J. P. Teare, S. Jakobovits, S. Zeki, K. I. Welsh, S. D. Taylor-Robinson, T. R. Orchard. *Am. J. Gastroenterol.* **2009**, *104*, 1435–1444.
[18] R. Schicho, R. Shaykhutdinov, J. Ngo, A. Nazyrova, C. Schneider, R. Panaccione, G. G. Kaplan, H. J. Vogel, M. Storr. *J. Proteome Res.* **2012**, *11*, 3344–3357.
[19] L. Lih-Brody, S. R. Powell, K. P. Collier, G. M. Reddy, R. Cerchia, E. Kahn, G. S. Weissman, S. Katz, R. A. Floyd, M. J. Mckinley, S. E. Fisher, G. E. Mullin. *Dig. Dis. Sci.* **1996**, *41*, 2078–2086.
[20] G. Sturniolo, M. Molokhia, R. Shields, L. Turnberg. *Gut* **1980**, *21*, 387–391.
[21] N. Solomons, I. Rosenberg, H. Sandstead, K. Vokhactu. *Digestion* **1977**, *16*, 87–95.
[22] G. Zhang, G. Hirasaki. *J. Magn. Reson.* **2003**, *163*, 81–91.
[23] E. Holmes, A. W. Nicholis, J. C. Lindon, S. Ramos, M. Spraul, P. Neidig, S. C. Connor, J. Connelly, S. J. P. H. Damment, J. K. Nicholson. *NMR Biomed.* **1998**, *11*, 235–244.
[24] N. J. Waters, E. Holmes, A. Williams, C. J. Waterfield, R. D. Farrant, J. K. Nicholson. *Chem. Res. Toxicol.* **2001**, *14*, 1401–1412.
[25] L. Eriksson, E. Johansson, N. Kettaneh-Wold, S. Wold. UMETRICS **1999**, 213–225.
[26] J. K. Nicholson, P. J. Foxall, M. Spraul, R. D. Farrant, J. C. Lindon. *Anal. Chem.* **1995**, *67*, 793–811.
[27] M. Fan. *Prog. Nucl. Magn. Reson. Spectrosc.* **1996**, *28*, 161–219.
[28] J. C. Lindon. *Annu. Rep. NMR Spectrosc.* **1999**, *38*, 1–88.
[29] L. Breiman. *Mach. Learn.* **2001**, *1*, 5–32.
[30] R. Diaz-Uriarte, S. A. D. Andres. *BMC Bioinformatics* **2006**, *7*, 97–106.
[31] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer, New York, **2001**.
[32] L. Breiman. *Mach. Learn.* **2001**, *45*, 5–32.
[33] T. Fawcett. *Pattern Recogn. Lett.* **2006**, *27*, 861–874.
[34] S. Wold, H. Antti, F. Lindgren, J. Ohman. *Chemom. Intell. Lab. Syst.* **1998**, *44*, 175–185.
[35] J. Sjöblom, O. Svensson, M. Josefson, H. Kullberg, S. Wold. *Chemom. Intell. Lab. Syst.* **1998**, *44*, 229–244.
[36] J. G. Topliss, R. P. Edwards. *J. Med. Chem.* **1979**, *22*, 1238–1244.
[37] M. W. King. Introduction to amino acid metabolism.
[38] E. Capristo. *Eur. Rev. Med. Pharmacol. Sci.* **1998**, *4*, 111–113.
[39] G. Faa, V. M. Nurchi, A. Ravarino, D. Fanni, S. Nemolato, C. Gerosa, P. V. Eyken, K. Geboes. *Coord. Chem. Rev.* **2008**, *252*, 1257–1269.
[40] C. H. Cho, L. Y. Y. Fong, P. C. C. Ma, C. W. Ogle. *Pharmacol. Biochem. Behav.* **1987**, *26*, 293–297.
[41] L. Kruidenier, I. Kuiper, W. V. Duijn, M. A. Mieremet-Ooms, R. A. V. Hogezand, C. B. Lamers, H. W. Verspaget. *J. Pathol.* **2003**, *201*, 17–27.
[42] M. S. Ardawi, E. A. Newsholme. *Biochem. J.* **1983**, *212*, 835–842.
[43] B. J. Mills, R. D. Lindeman, C. A. Lang. *J. Nutr.* **1981**, *111*, 1098–1102.
[44] N. A. Kretzmann, H. Fillmann, J. L. Mauriz, C. A. Marroni, N. Marroni, J. Gonzalez-Gallego, M. J. Tunon. *Inflamm. Bowel Dis.* **2008**, *14*, 1504–1513.
[45] S. P. Aiken, N. M. Horn, N. R. Saunders. *J. Physiol.* **1992**, *445*, 69–80.
[46] R. A. Disilvestro, R. J. Cousins. *Annu. Rev. Nutr.* **1983**, *3*, 261–288.