Journal of Trace Elements in Medicine and Biology 43 (2017) 202–208

Contents lists available at ScienceDirect

Journal of Trace Elements in Medicine and Biology

journal homepage: www.elsevier.com/locate/jtemb

Pathobiochemistry

Comparative study on serum levels of macro and trace elements in schizophrenia based on supervised learning methods

Tong Lin[a,1], Tiebing Liu[b,1], Yucheng Lin[a], Lailai Yan[c], Zhongxue Chen[d], Jingyu Wang[c,*]

[a] The Key Laboratory of Machine Perception (Ministry of Education), School of EECS, Peking University, Beijing 100871, China
[b] Civil Aviation Medicine Center, Civil Aviation Administration of China, Beijing 100123, China
[c] Center of Medical & Health Analysis, School of Public Health, Peking University, Beijing 100191, China
[d] Department of Epidemiology and Biostatistics, School of Public Health, Indiana University Bloomington, Bloomington 47405, USA

ARTICLE INFO

Article history:
Received 19 December 2016
Received in revised form 22 February 2017
Accepted 21 March 2017

Keywords:
Schizophrenia
Macro elements
Trace elements
Supervised learning methods

ABSTRACT

The etiology and pathophysiology of schizophrenia (SCZ) remain obscure. This study explored the associations between SCZ risk and serum levels of 39 macro and trace elements (MTE). A 1:1 matched case-control study was conducted among 114 schizophrenia patients and 114 healthy controls matched by age, sex and region. Blood samples were collected to determine the concentrations of 39 MTE by ICP-AES and ICP-MS. Both supervised learning methods and classical statistical testing were used to uncover the difference of MTE levels between cases and controls. The best prediction accuracies were 99.21% achieved by support vector machines in the original feature space (without dimensionality reduction), and 98.82% achieved by Naive Bayes with dimensionality reduction. More than half of MTE were found to be significantly different between SCZ patients and the controls. The presented investigation showed that there existed remarkable differences in concentrations of MTE between SCZ patients and healthy controls. The results of this study might be useful to diagnosis and prognosis of SCZ; they also indicated other promising applications in pharmacy and nutrition. However, the results should be interpreted with caution due to limited sample size and the lack of potential confounding factors, such as alcohol, smoking, body mass index (BMI), use of antipsychotics and dietary intakes. In the future the application of the analyses will be useful in designs that have larger sample sizes.

1. Introduction

Schizophrenia is a severe and costly mental disorder that affects 0.5–1% of the population worldwide [1]. It is common for people with SCZ to suffer from problems such as long-term unemployment, poverty, homelessness [2] and a higher suicide rate [3]. The average life expectancy of people with this disorder is ten to twenty five years less than that of people without this disorder [4].

The real cause of schizophrenia is mysterious and mostly believed to be a combination of environmental and genetic factors [5]. In recent years, many molecular signatures were discovered as significant biomarkers in schizophrenia patients' serum. In 2011 Schwarz et al. claimed that his study was the first to identify a biological signature for schizophrenia in blood serum [6]. In the same year, Yang et al. discovered that an array of serum markers, including glycerate, eicosenoic acid, β-hydroxybutyrate, pyruvate, cysteine and urine β-hydroxybutyrate, could achieve an area under the receiver operating characteristic curve (AUC) of 1 in both the training set and the test set [7]. In 2014, Tregellas et al. showed that intrinsic hippocampal activity could serve as a biomarker for cognition and symptoms in schizophrenia [8]. Neuroimaging biomarkers were also introduced by Tregellas for early drug development of schizophrenia in the same year [9]. In 2015, Chiappelli et al. introduced myo-inositol as a potential biomarker for depression in schizophrenia [10]. However, the mechanisms of these biomarkers are still unknown.

MTE play essential roles in the biological processes [11,12]. A number of studies have shown that changes of MTE levels might be linked to the etiology and pathophysiology of some psychiatry diseases, including schizophrenia [13,14]. However, there have existed numerous debates regarding the link between MTE levels and schizophrenia risk [15–17]. In addition, a number of MTE,

* Corresponding author at: Peking University, Center of Medical & Health Analysis, School of Public Health, No.38 Xue-Yuan Road, Haidian District, Beijing 100191, China.
E-mail addresses: lintong@pku.edu.cn (T. Lin), ltbing@yeah.net (T. Liu), xklyc@pku.edu.cn (Y. Lin), mike30103@aliyun.com (L. Yan), zc3@indiana.edu (Z. Chen), wjy@bjmu.edu.cn (J. Wang).
[1] These two authors contributed equally to this work.

http://dx.doi.org/10.1016/j.jtemb.2017.03.010
0946-672X/

such as molybdenum, neodymium, nickel, praseodymium, rubidium, antimony, stannum, strontium, thorium, titanium, thallium, uranium and vanadium, have not been included or completely discussed. The underlying interactions among these dozens of elements can be complex, traditional single variable analysis or correlation analysis may not be applicable to discover accurate predictions. Recently machine learning techniques, such as support vector machines (SVM) and feature selection methods, are gaining popularity in this field for handling high-dimensional input features and yielding better diagnostic capability. In this study, we conduct a 1:1 matched case-control study to probe the differences of 39 MTE levels between SCZ patients and healthy controls using supervised learning methods and classical statistical testing approaches.

2. Materials and methods

All experimental procedures were performed in accordance with the principles of the Declaration of Helsinki and later amendments. The study protocol was approved by the Ethics Review Committee of Health Science Center, Peking University (IRB00001052-12065). Each of the study participants was informed about the objectives of the study, and a written consent was obtained. If the independent capacity of any subject was doubted, the written consent of his or her primary caregiver was obtained simultaneously.

2.1. Study subjects

Initially, 130 clinically stable schizophrenia patients and 130 age-, sex- and region-matched healthy controls without psychiatric disorders, as evaluated by structured clinical interview for ICD-10, were recruited from consecutive admissions at the psychiatric department in the 261 Hospital of People's Liberation Army, Beijing, China from November 2012 to April 2013. The schizophrenia group was diagnosed according to the ICD-10 diagnosis of schizophrenia (F20) by clinically trained and experienced psychiatrists. The inclusion criteria were as follows: age between 18 and 65, no occupational exposure in heavy industry in the past, no acute infectious and traumatic diseases and no use of hormone therapy. After the recruitment, all patients and controls were screened for chronic physical illness (diabetes, kidney failure, or hepatic disease) from a clinical examination. Due to the potential confounding effects on MTE levels, participants who had evidence of chronic physical illness and were receiving mineral or vitamin supplements that might have influence on MTE levels were excluded, according to consultation with an internal medicine specialist. This screening procedure was done before data collection, and consequently a total of 114 schizophrenia patients and 114 healthy controls were included in the final sample for the study.

2.2. Sample collection

Venous blood sample (about 5 mL) was collected from the forearm vein of each participant after an overnight fast into a metal-free plastic tube. All blood samples clotted at room temperature for half an hour and then centrifuged at 3000 rpm for 15 min. Blood samples were separated, stored at $-20\,^{\circ}C$ and also protected from light.

2.3. Elements measurement

Each serum sample was put into a quartz tube and then 1.5 mL purified nitric acid (HNO_3) was added. After a serum sample was predigested at room temperature for 2 h, 0.5 mL H_2O_2 was added to promote further digestion. After that, the sample was digested in a microwave digestion system (MWS-2; Bergholt Co., Germany) and then diluted to 7 mL with deionized water.

Concentrations of calcium (Ca), magnesium (Mg), phosphorus (P) and sodium (Na), sulfur (S) were determined by inductively coupled plasma-atomic emission spectrometry (ICP-AES, American Thermo Electron Corporation iCAP-6300). Levels of other 34 elements, including aluminum (Al), arsenic (As), boron (B), barium (Ba), bismuth (Bi), cadmium (Cd), cerium (Ce), cobalt (Co), chromium (Cr), cesium (Cs), copper (Cu), iron (Fe), gadolinium (Gd), germanium (Ge), mercury (Hg), Lanthanun (La), lithium (Li), manganese (Mn), molybdenum (Mo), neodymium (Nd), nickel (Ni), lead (Pb), praseodymium (Pr), rubidium (Rb), antimony (Sb), selenium (Se), stannum (Sn), strontium (Sr), thorium (Th), titanium (Ti), thallium (Tl), uranium (U), vanadium (V) and zinc (Zn), were measured by inductively coupled plasma-mass spectrometry (ICP-MS, American PerkinElmer ELAN DRCII).

2.4. Quality control

Several methods were applied to ensure the accuracy of the determination of macro and trace elements. First, all tubes were made of polypropylene instead of glass materials in order to prevent metal contamination. Second, all reagents were analytical grade, and water was deionized. Third, several Standard Plasma References, including Level I (REF 8883) and Level II (REF 8884), were used. Four, Indium was added into each sample as an internal standard before digestion. Finally, the measurement of element levels was based on the most abundant isotope of each element to avoid interference.

2.5. Data normalization

Each sample contains two demographic variables (age, gender), together with concentrations of the aforementioned 39 elements. After data acquisition, preprocessing is necessary for later use. The first step is digitization, with gender (male or female) represented by 0 or 1. The next step is normalization: mapping all the concentrations of elements into an interval [0,1]. In the same way, ages are linearly transformed from 0 to 100 year-old into [0,1]. After preprocessing, the data of 114 schizophrenia patients and 114 healthy controls is presented as a 228×41 matrix, with each row for one subject. Ground-truth labels are used to stand for the case of schizophrenia or not by +1 or −1, respectively.

2.6. Evaluation protocol

To compare the classification performance of different methods, we use the 10 rounds of 5-fold cross validation [18] to obtain the average classification accuracies, namely the proportion of correct predictions to all the test subjects. Another popular measure is the area under the receiver operator [19], but this measure requires that classifiers change their parameters continuously to yield a function of sensitivity and specificity, which is rather demanding for our developed methods. Running time of cross-validation is used to compare the training and prediction time of each algorithm. All algorithms were implemented in Matlab on a 2.4-GHz i7-CPU machine with 8-GB memory.

2.7. Dimensionality reduction

The main objective of dimensionality reduction is to reduce the computational burden of classifiers and to alleviate the effects of data noise. Six methods are used for this objective, including Fisher feature selection (FFS), principal component analysis (PCA), Fisher discriminant analysis (FDA) with its variant (FDAx, where the

Table 1
Classification accuracies (in percentage) and runtime (in seconds) of 10 rounds 5-fold cross validation.

	Origin	FFS	PCA	FDA	FDAx	LPP	FA
NB	89.78(**0.17**)	91.14(0.20)	93.82(0.17)	98.82(0.10)	76.93(0.17)	94.17(0.14)	86.75(0.13)
LR	98.68(2.22)	95.88(0.60)	96.81(0.73)	98.79(0.68)	98.00(0.73)	95.92(0.73)	97.59(0.92)
NN	98.15(8.72)	96.30(10.01)	97.01(6.51)	98.66(5.62)	97.99(5.69)	96.22(5.69)	97.27(7.73)
AB	98.51(66.70)	94.65(17.90)	91.82(17.75)	98.52(2.43)	98.06(20.88)	73.59(20.88)	90.86(11.74)
SVM	**99.21**(0.23)	96.00(0.14)	96.58(0.14)	98.66(0.07)	97.66(0.18)	95.03(0.18)	96.21(0.08)
RF	97.98(22.31)	94.76(16.45)	94.30(19.08)	98.73(18.03)	98.09(18.74)	95.13(18.74)	94.12(19.22)

Table 2
The projection coefficients w in FDA and the Fisher discriminant ratio F used in Fisher feature selection.

Feature	w	F	Feature	w	F	Feature	w	F	Feature	w	F
Age	−0.63	0	Cr	−0.39	8.5	Mo	−0.72	0.32	Sn	−2.27	73.8
Gender	0.22	0	Cs	1.08	3.5	Na	−1.29	114.3	Sr	1.26	12.7
Al	−0.35	4.3	Cu	0.99	38	Nd	1.64	6.1	Th	0.36	4.8
As	0.92	52.4	Fe	−0.39	6.2	Ni	−0.84	17.1	**Ti**	**−16.1**	**196.9**
B	3.06	91.4	Gd	2.11	9.4	P	1.42	14.4	Tl	3.8	11.9
Ba	−0.94	0	Ge	2.08	2.9	Pb	0.99	7.6	U	−0.94	2.8
Bi	−0.25	22.5	Hg	1.37	2	Pr	−2.63	2.7	V	−1.15	41.3
Ca	**−7.27**	55.1	La	−1.31	0.02	Rb	−1.7	13.0	Zn	4.84	1.5
Cd	1.69	9.2	Li	0.89	25.8	S	0.67	**114.1**			
Ce	0.63	1.7	Mg	4.73	0.83	Sb	3.06	0.12			
Co	−0.58	32.2	Mn	2.87	48.3	Se	−1.56	54.8			s

between-class scatter matrix is simply replaced by the "total" mixture scatter matrix to allow more nonzero eigen-values), locality preserving projection (LPP), and factor analysis (FA) [20–24]. Apart from FFS in which a small set of features are chosen directly, other five methods aim at transforming the original features into a low-dimensional embedding space. The reader can refer to the review and systematic comparison of different methods for dimensionality reduction [25].

2.8. Classifiers

A recent comprehensive performance evaluation on the whole UCI classification datasets [26] showed that the top rank of the best classifiers includes random forest, SVM, neural network, and boosting. In our SCZ diagnosis system, we apply six classifiers: Random Forest [27], Support Vector Machine [28], Neural Network [29], AdaBoost [30], Naive Bayes [31], and Logistic Regression [31].

3. Results

A 228×39 matrix consisting of 114 patients and 114 healthy subjects with 39 feature macro and trace elements from the schizophrenia patient data is used. We test the aforementioned six classifiers on the original feature space as well as on the embedding spaces by using six dimensionality reduction methods. The dimension of the most embedding spaces is set as 10 for a trade-off between accuracy and complexity, except that the FDA algorithm

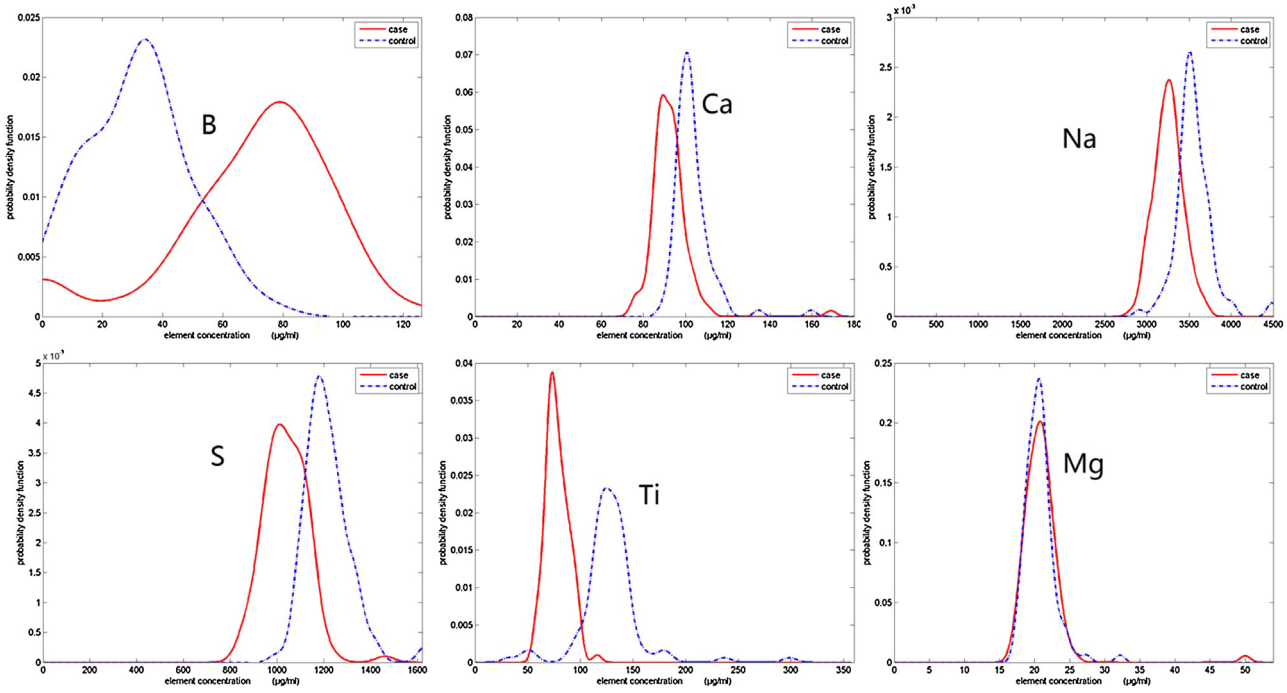

Fig. 1. Concentration distributions of five important elements and one unimportant element (Mg) for SCZ patients and healthy controls.

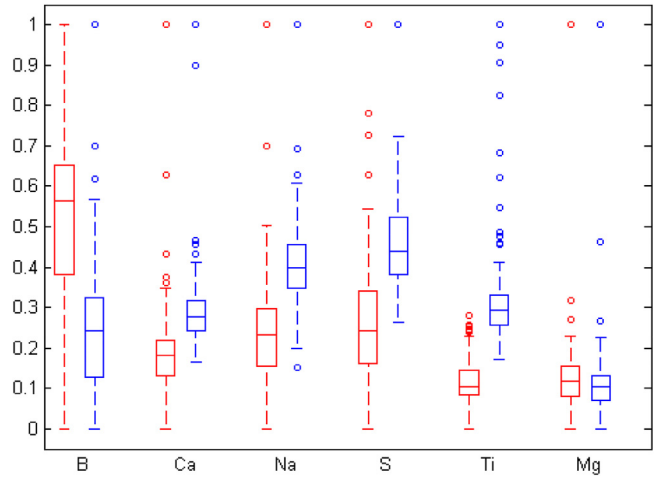

Fig. 2. Concentration distributions of previous six elements shown as bot plots for patients (red) and control (blue). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

can only project into 1D because of its inherent limitation for two-class problems.

The averaged classification accuracies are reported in Table 1 based on 10 rounds 5-fold cross validation. We can see that SVM achieves the best accuracy 99.21% on the original feature vectors (without dimensionality reduction), and Naive Bayes outperforms other classifiers by yielding 98.82% on embedding spaces with dimensionality reduction. In comparison, dimensionality reduction methods, namely FDA and FDAx, based on Fisher discriminant ratios are often more favorable than the other four. The results on running time are also listed in Table 1. The dimensionality reduction step can evidently speedup the running time for each classifier; on the other hand, three classifiers, NB, LR, and SVM, are faster than the remaining ones.

Through dimension reduction, the learned parameters of FDA and Fisher feature selection (FFS) can reflect the correlation or significance of the macro and trace elements to SCZ. Specifically, the projection vector $\boldsymbol{w}$ of FDA can be regarded as linear weights to the original features, thus larger weights means higher contribution. For FFS, the $\boldsymbol{F}$ statistic indicates the discriminant capability of each chemical element: the larger the F is, the greater the difference exists for that chosen element. As shown in Table 2, we can see that Titanium (Ti) is the top element with more discriminant information, with F = 196.9. Besides, other important elements may include Boron (B) (F = 91.4), Calcium (Ca) (F = 55.1), Natrium (Na) (F = 114.3), and sulfur (S) (F = 114.1).

To further examine the discriminant capability, we draw the concentration distributions of these five important elements and of one unimportant element (Mg) for comparison. As shown in Figs. 1 and 2, we can see that the differences between cases and controls in these five elements are significant, whereas the difference in the element Mg is not so much. The first part of Table 3 lists the classification accuracies based on each single element. As we can see, all six classifiers can achieve accuracies more than 90% based on the single most important element Ti. The other four elements are B, Ca, Na and S, which provid accuracies around 80%.

We also investigate whether the relation of any two elements is different between cases and controls. Fig. 3 shows the distributions in normalized concentration for four pairs of elements, including B-Ti, Na-S, Ca-Ti and Ca-Na. We can see that these 2D scatter plots are highly separable, though no straightforward functions can be available to describe the pairwise relationship. The middle part of Table 3 displays the classification accuracies based on a small set of element pairs. It indicates that the pair of B and Ti achieves the best performance for all the classifiers, with classification accuracies ranging from 92.89% to 95.35%. It appears that B and Ti are most diverse elements and they complement each other, though classification performance on single B, with classification accuracies ranging from 72.72% to 79.69%, is not satisfactory.

When considering the scatter plots of any three important elements, the separability can be improved as shown in Fig. 4. Almost

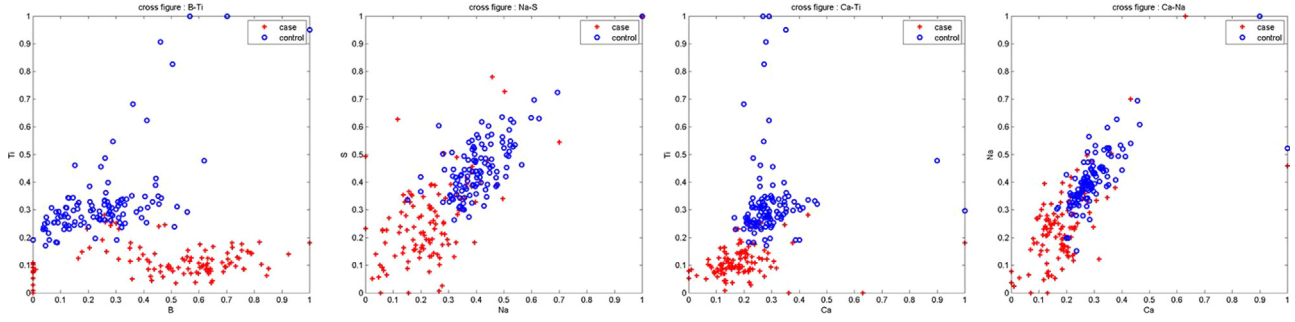

Fig. 3. Distributions in normalized concentration for pairs of elements (B-Ti, Na-S, Ca-Ti and Ca-Na).

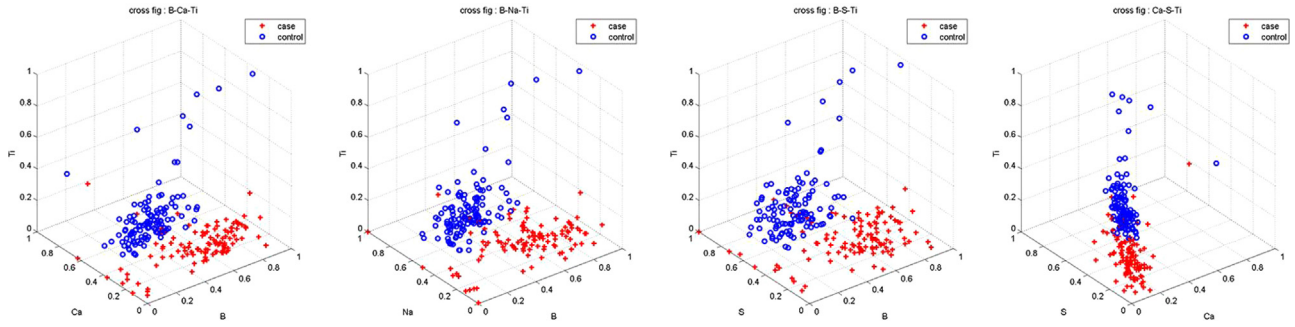

Fig. 4. Distributions in normalized concentration for combinations of three elements (B-Ca-Ti, B-Na-Ti, B-S-Ti, and Ca-S-Ti).

Table 3
Classification accuracies (in percentage) based on 1/2/3 elements.

	B	Ca	Na	S	Ti
NB	77.89	80.31	82.41	79.78	**92.98**
LR	77.76	81.93	81.71	80.22	**92.98**
NN	79.69	81.80	83.07	78.60	**94.04**
AB	79.61	79.87	83.82	76.76	**92.32**
SVM	79.02	82.06	83.25	79.65	**93.11**
RF	72.72	74.47	77.02	73.11	**91.71**

	B + Na	B + S	Ca + S	B + Ti	Ca + Ti
NB	87.02	85.00	80.13	**94.17**	92.68
LR	87.32	84.69	80.57	**95.35**	92.46
NN	88.55	88.55	83.29	**93.51**	92.68
Ada	86.75	84.12	80.13	**92.89**	92.32
SVM	87.46	85.13	79.87	**95.18**	93.33
RF	86.23	87.37	80.88	**94.21**	92.46

	B + Ca + Ti	B + Na + Ti	B + S + Ti	Ca + Na + Ti	Ca + S + Ti
NB	93.99	**94.56**	93.42	91.84	92.63
LR	95.00	**95.18**	94.91	93.11	92.46
NN	93.25	93.86	**93.95**	91.84	93.29
Ada	**94.39**	93.77	94.34	91.40	91.49
SVM	**94.96**	94.91	94.56	93.64	93.38
RF	**94.82**	94.52	94.25	92.94	93.64

any linear classifier can achieve good performance based on these triples of elements. The last part of Table 3 displays classification results based on a small set of element triples. The triple composed of B-Ca-Ti achieves the best for three classifiers and B-Na-Ti achieves the best for two.

As shown in Table 4, the serum levels of Al, As, B, Bi, Ca, Cd, Co, Cu, Fe, Gd, Li, Mn, Na, Ni, P, Rb, S, Se, Sn, Sr, Th, Ti, Tl, and V, are all significantly different between cases and controls (all $P < 0.05$), according to both *T*-test and rank sum test.

4. Discussion

In this article, we present a study of macro and trace elements (MTE) in serum for SCZ patients based on supervised learning methods. In order to make a comparison to machine learning methods, the traditional hypothesis testing approaches, including *T*-test and rank sum test, are applied. As shown in Table 4, at the level of 5%, more than half of MTE are shown to be statistically significantly different between cases and controls. The means and standard deviations differ enormously between the case and control groups over many features. This is mainly because of singular values (or outliers) incurred in measurements to the data. For example, on certain elements, the highest concentration is more than thousand times of the average value, which influenced the test results greatly. However, machine learning methods are much more robust dealing with such "outliers" problem.

The contributions of this study are twofold. First, we identify an array of significant MTE in blood serum samples for distinguishing SCZ patients from healthy subjects. The blood sample acquisition is practical, noninvasive, and easily performed in clinical settings. Second, we use state-of-the-art machine learning algorithms to classify the concentrations of 39 measured MTE. Since the underlying interactions among these dozens of elements can be complex, traditional single variable analysis or correlation analysis may not be applicable to discover accurate predictions. Instead, several dimensionality reduction methods are employed to exclude

Table 4
The results of hypothesis tests. Mean and standard deviation of case and control, and the P-value of *T*-test and rank sum test. (ng/ml, Alpha = 0.05, P-value under alpha is bolded.).

Feature	Case	Control	*T*-test	Rs-test	Feature	Case	Control	*T*-test	Rs-test
age	32.8 ± 11.3	32.9 ± 10.7	0.928	0.778	Mn	7.01 ± 4.86	3.47 ± 2.64	**0.000**	**0.000**
Al	169 ± 212	121 ± 146	**0.040**	**0.000**	Mo	2.00 ± 2.37	2.03 ± 0.97	0.571	**0.004**
As	25.0 ± 11.1	37.3 ± 14.8	**0.000**	**0.000**	Na	3.3e6 ± 2.2e5	3.6e6 ± 1.9e5	**0.000**	**0.000**
B	64.4 ± 30.0	32.4 ± 21.9	**0.000**	**0.000**	Nd	0.20 ± 0.26	0.14 ± 0.09	**0.015**	0.106
Ba	7.68 ± 31.5	5.16 ± 7.03	0.946	0.113	Ni	5.73 ± 7.62	10.0 ± 9.44	**0.000**	**0.000**
Bi	0.26 ± 0.42	0.07 ± 0.13	**0.000**	**0.000**	P	1.2e5 ± 1.7e4	1.1e5 ± 1.5e4	**0.000**	**0.000**
Ca	9.7e4 ± 4.1e4	10e4 ± 1.0e4	**0.000**	**0.000**	Pb	6.31 ± 8.78	3.98 ± 12.2	**0.006**	0.066
Cd	0.17 ± 0.16	0.28 ± 0.47	**0.000**	**0.000**	Pr	0.04 ± 0.07	0.03 ± 0.02	0.102	0.751
Ce	0.37 ± 0.63	0.27 ± 0.52	0.194	0.406	Rb	196 ± 55.7	225 ± 66.1	**0.000**	**0.007**
Co	0.47 ± 0.33	0.74 ± 0.63	**0.000**	**0.000**	S	1.1e6 ± 1.2e5	1.2e6 ± 8.5e4	**0.000**	**0.000**
Cr	4.63 ± 6.31	7.20 ± 50.6	**0.004**	0.058	Sb	3.64 ± 1.25	3.59 ± 1.25	0.729	0.295
Cs	0.98 ± 3.59	0.89 ± 0.72	0.061	**0.000**	Se	65.7 ± 35.5	97.5 ± 37.1	**0.000**	**0.000**
Cu	803 ± 232	1033 ± 320	**0.000**	**0.000**	Sn	0.52 ± 0.38	1.98 ± 2.24	**0.000**	**0.000**
Fe	1340 ± 921	1530 ± 592	**0.014**	**0.000**	Sr	50.9 ± 17.7	67.2 ± 60.3	**0.000**	**0.001**
Gd	0.06 ± 0.04	0.04 ± 0.03	**0.002**	**0.000**	Th	0.50 ± 0.78	0.29 ± 0.76	**0.029**	**0.000**
Ge	3.80 ± 1.39	4.21 ± 2.36	0.090	**0.044**	Ti	79.9 ± 15.8	139 ± 42.4	**0.000**	**0.000**
Hg	3.52 ± 20.5	1.15 ± 3.17	0.161	**0.000**	Tl	0.10 ± 0.10	0.14 ± 0.10	**0.001**	**0.000**
La	0.13 ± 0.26	0.13 ± 0.30	0.879	0.575	U	0.03 ± 0.03	0.15 ± 1.31	0.095	0.236
Li	253 ± 184	175 ± 53.2	**0.000**	**0.000**	V	0.50 ± 0.52	0.91 ± 0.60	**0.000**	**0.000**
Mg	2.2e4 ± 1.4e4	2.1e4 ± 3404	0.361	**0.038**	Zn	881 ± 270	843 ± 279	0.215	**0.006**

seemingly "irrelevant" variables and to speed up the following calculations. Then in the embedded low-dimensional space, classifiers are trained to learn the implicit patterns within the data. Experimental results on a data set composed of 114 schizophrenia patients and 114 healthy controls show that the prediction accuracies in the embedded space and in the original space are adequate to identify SCZ patients convincingly. The key point is that machine learning methods can be promising for analyzing complex patterns in MTE for mental disorder diagnosis. The proposed approach may have other emerging applications beyond diagnosis purposes. One possibility is to modify MTE by pharmacy for enforcing the element concentrations of a patient into a normal interval. Another perspective is to provide a rebalanced diet in nutrition for patients. These new applications will depend on the thorough and deep analysis of MTE in serum between patients and healthy controls.

There were several limitations in this study. First, possible confounding biases were unlikely to be controlled and avoided, due to the absence of some factors, such as the use of antipsychotics, body mass index (BMI), alcohol, dietary intakes and smoking. In the literature, one study did not observe any link between duration of illness, antipsychotic doses, and the alterations of some elements, including zinc, iron, copper, selenium and manganese [32]. Moreover, even if the nutritional and socioeconomic status of cases and controls were comparable, it remains impossible to eliminate the possibility that diverse genetic features might be associated with schizophrenia. Second, our data only presented the alterations of MTE in serum among a relatively small number of participants, further studies with prospective design and larger sample sizes are need.

This study also had several significant merits. The first advantage was the simultaneous measurement of 39 MTE under the identical experimental conditions. Another one was the match on age, sex and region of SCZ patients and controls. Finally, the classification accuracies reported in this article were remarkably higher than most empirical decisions, and the test error rates provided quantitative confidence of the prediction results.

5. Conclusion

This study might be useful for diagnosis and prognosis of SCZ disorder. It also indicated other potential applications in pharmacy and nutrition. The results of this study provide useful information for designing future studies with larger sample sizes. However, the results should be interpreted with caution due to the limited sample size and the lack of controlling for the potential confounding factors, such as the use of antipsychotics, BMI, alcohol, dietary intakes and smoking.

Funding

This work was supported by the National Basic Research 973 Program of China under Grant No. 2011CB302202, the National Natural Science Foundation of China under Grant Nos. 61375051 and 61075119, and the Seeding Grant for Medicine and Information Sciences of Peking University under Grant No. 2014-MI-21.

Author contributions

J.W. and T.B.L. proposed and supervised the project. T.L. and Y.L. designed and carried out the experiments and analyzed the data. L.Y. and T.B.L. contributed to the sample collection, macro and trace elements measurement, quality control, and finding of related literature. Z.C. participated in the discussion and helped to improve the proposed methods. T.L., Y.L. and T.B.L. wrote the manuscript.

Competing financial interests

The authors declare no competing financial interests.

Acknowledgements

We thank Jing Guo, Fangbo Feng, Jinyun Qiu for screening participants, performing cognitive and symptom assessments, and collecting the samples and information of study subjects. We also thank Yaqiong Liu for the technical support of the determination of macro and trace elements.

References

[1] S. Saha, D. Chant, J. Welham, J. McGrath, A systematic review of the prevalence of schizophrenia, PLoS Med. 2 (2005) e141.
[2] P.F. Buckley, B.J. Miller, D.S. Lehrer, D.J. Castle, Psychiatric comorbidities and schizophrenia, Schizophr. Bull. 35 (2009) 383–402.
[3] K. Hor, M. Taylor, Suicide and schizophrenia: a systematic review of rates and risk factors, J. Psychopharmacol. 24 (2010) 81–90.
[4] T.M. Laursen, T. Munk-Olsen, M. Vestergaard, Life expectancy and cardiovascular mortality in persons with schizophrenia, Curr. Opin. Psychiatry 25 (2012) 83–88.
[5] J. van Os, S. Kapur, Schizophrenia, Lancet 374 (2009) 635–645.
[6] E. Schwarz, P.C. Guest, H. Rahmoune, L.W. Harris, L. Wang, F.M. Leweke, M. Rothermundt, B. Bogerts, D. Koethe, L. Kranaster, P. Ohrmann, T. Suslow, G. McAllister, M. Spain, A. Barnes, N.J. van Beveren, S. Baron-Cohen, J. Steiner, F.E. Torrey, R.H. Yolken, S. Bahn, Identification of a biological signature for schizophrenia in serum, Mol. Psychiatry 17 (2012) 494–502.
[7] J. Yang, T. Chen, L. Sun, Z. Zhao, X. Qi, K. Zhou, Y. Cao, X. Wang, Y. Qiu, M. Su, A. Zhao, P. Wang, P. Yang, J. Wu, G. Feng, L. He, W. Jia, C. Wan, Potential metabolite markers of schizophrenia, Mol. Psychiatry 18 (2013) 67–78.
[8] J.R. Tregellas, J. Smucny, J.G. Harris, A. Olincy, K. Maharajh, E. Kronberg, L.C. Eichman, E. Lyons, R. Freedman, Intrinsic hippocampal activity as a biomarker for cognition and symptoms in schizophrenia, The Am. J. Psychiatry 171 (2014) 549–556.
[9] J.R. Tregellas, Neuroimaging biomarkers for early drug development in schizophrenia, Biol. Psychiatry 76 (2014) 111–119.
[10] J. Chiappelli, L.M. Rowland, S.A. Wijtenburg, F. Muellerklein, M. Tagamets, R.P. McMahon, F. Gaston, P. Kochunov, L.E. Hong, Evaluation of myo-Inositol as a potential biomarker for depression in schizophrenia, Neuropsychopharmacology 40 (2015) 2157–2164.
[11] C.G. Fraga, Relevance, essentiality and toxicity of trace elements in human health, Mol. Asp. Med. 26 (2005) 235–244.
[12] A. Alimonti, G. Ristori, F. Giubilei, M.A. Stazi, A. Pino, A. Visconti, S. Brescianini, M. Sepe Monti, G. Forte, P. Stanzione, B. Bocca, G. Bomboi, C. D'Ippolito, V. Annibali, M. Salvetti, G. Sancesario, Serum chemical elements and oxidative status in Alzheimer's disease, Parkinson disease and multiple sclerosis, Neurotoxicology 28 (2007) 450–456.
[13] M. Nechifor, C. Vaideanu, I. Palamaru, C. Borza, I. Mindreci, The influence of some antipsychotics on erythrocyte magnesium and plasma magnesium calcium, copper and zinc in patients with paranoid schizophrenia, J. Am. Coll. Nutr. 23 (2004) 549S–551S.
[14] A. Rahman, M.A. Azad, I. Hossain, M.M. Qusar, W. Bari, F. Begum, S.M. Huq, A. Hasnat, Zinc, manganese, calcium, copper, and cadmium level in scalp hair samples of schizophrenic patients, Biol. Trace Elem. Res. 127 (2009) 102–108.
[15] A. Herran, M.T. Garcia-Unzueta, M.D. Fernandez-Gonzalez, J.L. Vazquez-Barquero, C. Alvarez, J.A. Amado, Higher levels of serum copper in schizophrenic patients treated with depot neuroleptics, Psychiatry Res. 94 (2000) 51–58.
[16] B. Vidovic, B. Dordevic, S. Milovanovic, S. Skrivanj, Z. Pavlovic, A. Stefanovic, J. Kotur-Stevuljevic, Selenium, zinc, and copper plasma levels in patients with schizophrenia: relationship with metabolic risk factors, Biol. Trace Elem. Res. 156 (2013) 22–28.
[17] T.L. Wolf, J. Kotun, J.H. Meador-Woodruff, Plasma copper iron, ceruloplasmin and ferroxidase activity in schizophrenia, Schizophr. Res. 86 (2006) 167–171.
[18] M. Mohri, A. Rostamizadeh, A. Talwalkar, Foundations of Machine Learning, The MIT Press, Cambridge, UK, 2012.
[19] F. Provost, T. Fawcett, Analysis and visualization of classifier performance: comparison under imprecise class and cost distributions, KDD (1997) 43–48.
[20] J.A. Rice, Mathematical Statistics and Data Analysis, 2nd ed., Duxbury Press, New York, USA, 1995.
[21] S. Theodoridis, K. Koutroumbas, Pattern Recognition, 4th ed., ElsevierL, Burlington, Canada, 2008.
[22] X. He, P. Niyogi, Locality preserving projections, Proceedings of Advances in Neural information processing systems (2003) 153–160.
[23] A. Ng, Factor Analysis and Lecture Notes of Machine Learning, Stanford, 2008.
[24] H. Harman, Modern Factor Analysis, University of Chicago Press, Chicago, USA, 1976.
[25] L. Maaten, E. Postma, H. Herik, Dimensionality Reduction: A Comparative Review, TiCC, Tilburg University, 2009 (Technical Report).

[26] C.f.M.L.a.I. Systems, the UC Irvine Machine Learning Repository. <http://archive.ics.uci.edu/ml/> (Accessed 1 March 2016).
[27] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.
[28] V. Vapnik, Statistical Learning Theory, John Wiley & Sons, New York, USA, 1998.
[29] D. Rumelhart, G. Hinton, R. Williams, Learning representations by backpropagating errors, Nature (1986) 533–536.
[30] R. Schapire, Y. Freund, Boosting: Foundations and Algorithms, The MIT Press, Massachusetts, USA, 2012.
[31] S. Shalev-Shwartz, S. Ben-David, Understading Machine Learning, Cambridge University Press, New York, USA, 2014.
[32] M. Yanik, A. Kocyigit, H. Tutkun, H. Vural, H. Herken, Plasma manganese, selenium, zinc, copper, and iron concentrations in patients with schizophrenia, Biol. Trace Elem. Res. 98 (2004) 109–117.