UNIVERSITY OF WEST LONDON
The Career University

UWL REPOSITORY

repository.uwl.ac.uk

Identification of metabolic kinetic patterns in different brain regions using metabolomics methods coupled with various discriminant approaches

Zheng, D, Li, Z, Li, S, Li, X, Kamal, GM, Liu, C, Manyande, Anne ORCID logoORCID: https://orcid.org/0000-0002-8257-0722, Xu, F, Bao, Q and Wang, J (2021) Identification of metabolic kinetic patterns in different brain regions using metabolomics methods coupled with various discriminant approaches. Journal of Pharmaceutical and Biomedical Analysis, 198. p. 114027. ISSN 0731-7085

http://dx.doi.org/10.1016/j.jpba.2021.114027

This is the Accepted Version of the final output.

UWL repository link: https://repository.uwl.ac.uk/id/eprint/7762/

Alternative formats: If you require this document in an alternative format, please contact: open.research@uwl.ac.uk

Copyright: Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy: If you believe that this document breaches copyright, please contact us at open.research@uwl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Rights Retention Statement:

Journal Pre-proof

Identification of metabolic kinetic patterns in different brain regions using metabolomic methods coupled with various discriminant approaches

Danhao Zheng, Zhao Li, Shuang Li, Xihai Li, Ghulam Mustafa Kamal, Chaoyang Liu, Anne Manyande, Fuqiang Xu, Qingjia Bao, Jie Wang

PII: S0731-7085(21)00139-4

DOI: https://doi.org/10.1016/j.jpba.2021.114027

Reference: PBA 114027

To appear in: *Journal of Pharmaceutical and Biomedical Analysis*

Received Date: 26 November 2020

Revised Date: 8 March 2021

Accepted Date: 12 March 2021

Please cite this article as: Zheng D, Li Z, Li S, Li X, Kamal GM, Liu C, Manyande A, Xu F, Bao Q, Wang J, Identification of metabolic kinetic patterns in different brain regions using metabolomic methods coupled with various discriminant approaches, *Journal of Pharmaceutical and Biomedical Analysis* (2021), doi: https://doi.org/10.1016/j.jpba.2021.114027

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier.

Journal Pre-proof

Identification of metabolic kinetic patterns in different brain regions using metabolomic methods coupled with various discriminant approaches

Danhao Zheng [a, b, 1]; Zhao Li [a, b, 1]; Shuang Li [a]; Xihai Li [c], Ghulam Mustafa Kamal [d], Chaoyang Liu [a, b]; Anne Manyande [e], Fuqiang Xu [a, b, f]; Qingjia Bao [g, h*]; Jie Wang [a, b, i*]

[a] Key Laboratory of Magnetic Resonance in Biological Systems, State Key Laboratory of Magnetic Resonance and Atomic and Molecular Physics, National Center for Magnetic Resonance in Wuhan, Wuhan Institute of Physics and Mathematics, Innovation Academy for Precision Measurement Science and Technology, Chinese Academy of Sciences-Wuhan National Laboratory for Optoelectronics, 430071 Wuhan, P. R. China;

[b] University of Chinese Academy of Sciences, Beijing 100049, P.R. China;

[c] Fujian Key Laboratory of Integrative Medicine on Geriatrics, Academy of Integrative Medicine, Fujian University of Traditional Chinese Medicine, Fuzhou, 350122, P.R. China

[d] Department of Chemistry, Khwaja Fareed University of Engineering and Information Technology Rahim Yar Khan 64200 Pakistan

[e] School of Human and Social Sciences, University of West London, London, UK;

[f] Center for Excellence in Brain Science and Intelligent Technology, Chinese Academy of Sciences, Shanghai 200031, P.R. China;

[g] Wuhan United Imaging Life Science Instrument Co., Ltd, Wuhan 430206

[h] Weizmann Institute of Science, Tel Aviv-Yafo, Israel 76001

[i] Hebei Provincial Key Laboratory of Basic Medicine for Diabetes, 2nd Hospital of Shijiazhuang, Shijiazhuang, Hebei, 050051, P.R. China.

[1:] These authors contributed equally to this work

[*:] Corresponding author

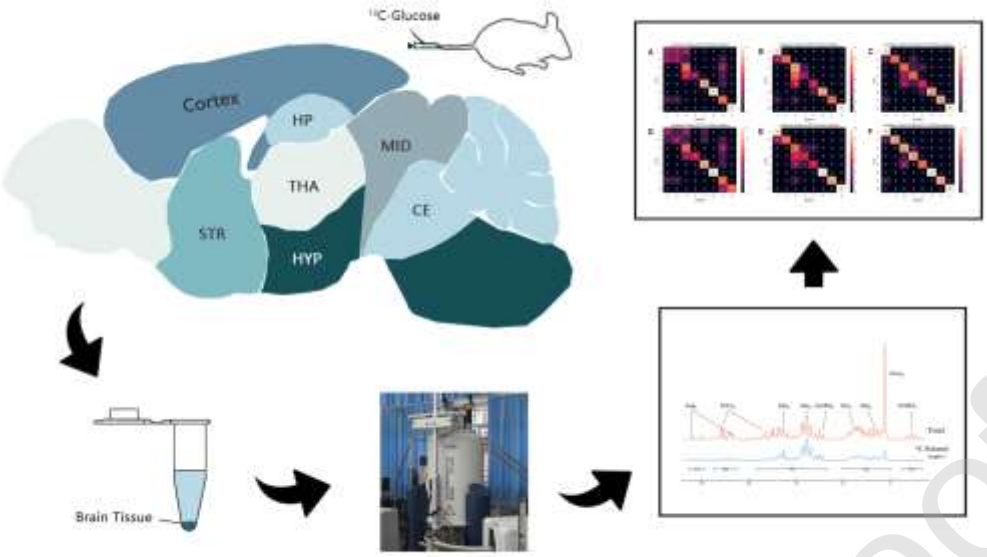
Jie Wang: E-mail: jie.wang@wipm.ac.cn; Tel: +86-27-87187359; Fax: +86-27-87199543.

Qingjia Bao: E-mail: qingjia.bao@weizmann.ac.il; Tel: 027-87199686.

Graphical Abstract

Journal Pre-proof

Journal Pre-proof



Highlights

- NMR spectroscopic was utilized to measure the enrichment ratio of brain regions.
- Metabolic kinetics was used as the feature to construct classification models.
- Multiple discriminative methods were used to construct classification models.
- Differences in metabolic kinetics among these brain regions were analyzed.

Abstracts

Metabolomics is widely used as a powerful technique for identifying metabolic patterns and functions of organs and biological systems. Normally, there are multiple groups/targets involved in data processed by discriminant analysis. This is more common in cerebral studies, as there are always several brain regions involved in neuronal studies or brain metabolic dysfunctions.

Journal Pre-proof

Furthermore, neuronal activity is highly correlated with cerebral energy metabolism, such as oxidation of glucose, especially for glutamatergic (excitatory) and GABAergic (inhibitory) neuronal activities. Thus, regional cerebral energy metabolism recognition is essential for understanding brain functions. In the current study, ten different brain regions were considered for discrimination analysis. The metabolic kinetics were investigated with ^{13}C enrichments in metabolic products of glucose and measured using the nuclear magnetic spectroscopic method. Multiple discriminative methods were used to construct classification models in order to screen out the best method. After comparing all the applied discriminatory analysis methods, the boost-decision tree method was found to be the best method for classification and every cerebral region exhibited its own metabolic pattern. Finally, the differences in metabolic kinetics among these brain regions were analyzed. We, therefore, concluded that the current technology could also be utilized in other multi-class metabolomics studies and special metabolic kinetic patterns could provide useful information for brain function studies.

Keywords: Metabolomics; Metabolic kinetics; POCE; Brain; Decision tree

1. Introduction

Metabolomics has emerged as a very powerful and rapidly developing technique for the identification and quantification of metabolites such as amino acids, hormones and signaling molecules, as well as their interactions to study the metabolism and function of organs and biological systems [1]. With advancements in analytical techniques, for example, nuclear magnetic resonance spectroscopy (NMR), liquid or gas chromatography and mass spectrometry, researchers

Journal Pre-proof

are now capable of efficiently screening key metabolites in blood samples, cerebrospinal fluid, urine, or postmortem tissues [2]. This development in metabolomics has contributed significantly to our understanding of the fundamental mechanisms of the physiology, pathobiology, drug toxicology and functional genomics in furthering clinical research [3].

Among the various tissue/fluid samples, a remarkable example is that of brain tissues. The central nervous system (CNS), considered as the most intricate network in structure and function of the organism [4], consumes nearly 20% of the body's energy in the resting state. The brain plays the most dominant role in the CNS and is responsible for advanced life activities comprising of the human language, consciousness, sensory information, emotional reactions and voluntary movements. A variety of activities of daily living are associated with different brain regions and closely related to metabolic connections of different groups of neurons or between neurons and glial cells [5]. Under different pathological conditions, the levels of small metabolites in different brain regions show corresponding changes. For example, neuroapoptosis occurs in the cerebral cortex and striatum of patients with Huntington's disease [6]. This results from abnormal energy metabolism and high levels of glutamate accumulation as a consequence of its reduced uptake [7]. Another example is the substantial loss of dopaminergic neurons in substantia nigra pars compacta (SNpc) and ventral tegmental area (VTA) of Parkinson's patients [8]. Studies have shown that output signals of dopamine cells from the midbrain, especially the substantia nigra pars compacta (SNpc) to the striatum, particularly the dorsal striatum, are capable of triggering locomotion in mice, which can cause Parkinson's metabolic changes [9]. Thus, the systematic study of the brain metabolome is conducive to deepening our understanding of the pathogenesis of brain diseases thereby providing us with a strong basis for precision therapy and drug development.

Journal Pre-proof

Early metabolomics research mainly focused on concentrations of small molecules in the brain tissue, but the concentration could not solely reveal the changing trends in metabolites from a dynamic perspective. The rate of change in metabolite levels serves as a measure for distinguishing different brain regions. Glucose, undoubtedly, is the main substrate for energy metabolism in the brain. By intravenously injecting glucose labeled with ^{13}C isotope and later detecting all the ^{13}C-labeled molecules *via* NMR spectroscopy [5], the pathways of energy generation in each brain region can be determined over a period of time. Thus, detailed kinetics of brain metabolites can be studied. This makes it possible for a machine learning algorithm to extract features of different brain regions and provide a robust predictive model.

Binary classification models, such as partial least squares discriminant analysis (PLS-DA) or orthogonal partial least squares discriminant analysis (OPLS-DA), are used for exploratory analysis of multidimensional metabolome datasets by reducing data dimensions while preserving most of the variation [10]. However, for brain samples with multiple groups of signature metabolites, this unique statistical modeling does not work effectively and a new methodology suitable for exploring the diversity of biomarkers is needed. Ensemble learning algorithms, such as Decision Tree, Adaptive Boosting and Random Forest, can prove to be a better choice for the identification and prediction of metabolic characteristics of different brain regions. In this study, POCE-NMR (Proton observed carbon editing nuclear magnetic resonance, ^{1}H-[^{13}C]-NMR) sequence was utilized to acquire the ^{13}C isotopic labelled information from different brain regions in rats. With the help of different statistical classification methods, brain regions characterized with different molecular metabolism could be identified, which has tremendous potential for future metabolomics studies in the brain or any other organs. The special metabolic kinetic

Journal Pre-proof

patterns identified in the current study could provide useful information for brain function studies. Furthermore, the optimized and studied classification methods could also potentially be applied in other multiple classes of metabolomic studies.

2. Materials and Methods

2.1 Animals

All experimental protocols were approved by the Animal Care and Use Committee in Innovation Academy for Precision Measurement Science and Technology, the Chinese Academy of Sciences (APM20043A). To save the research funding and animals, a total of 26 alcohol preferring rats (P-rats) were used and the breeding pairs were kindly offered as a gift by Prof. Lawrence (University of Melbourne). The animals were born and kept in SPF (Specific pathogen free) animal housing. On postnatal day (PND) 21, animals were ablactated and separated by sex (three males/females per cage). Only male P-rats (~200-250*g*) were chosen as the test group. During the preparation process, all rats were allowed free access to chow and water. The colony room was illuminated on a 12h light/dark cycle (lights on: 8:00-20:00) at a temperature of 23±1˚C and relative humidity of 45%. To minimize stress and familiarization with human contact, animals were handled twice and weighed on a daily basis.

2.2 Infusion of [1-^{13}C] glucose

To minimize the influence of intrinsic glucose levels in the animal's body, the rats were fasted overnight (for around 12 h) before the [1-^{13}C] glucose infusion. On the day of the experiment, animals were firstly weighed and then given anesthesia through inhalation with 4.0-5.0% isoflurane mixed with air, followed by a decrease to 1.0-1.5% to maintain the anesthetic

Journal Pre-proof

state. After an adequate level of anesthesia was provided, all rats were catheterized in the lateral tail vein *via* a PE50 cannula (Instech, PA, USA) with a needle tip and fastened with paper adhesive tape. The other side of the canula was attached to a syringe filled with saline [11]. After that the animal was moved to a new cage for recovery. The syringe was replaced with a new syringe containing [1-^{13}C]-glucose solution (pH 7.2; 0.75 M) which was held with an infusion pump (Fusion100, Chemyx, TX, USA), and the animals gradually awoke from anesthesia. After 20 minutes recovery, [1-^{13}C]-glucose solution was gradually infused for 30 minutes through the tail vein catheterization following a former protocol [12]. During the infusion periods, the animal was allowed free movement.

2.3 Brain tissue collection

As soon as the infusion was completed, the animal was anesthetized with an overdose of isoflurane. Then the anesthetized animal was euthanatized by the head-focused microwave irradiation machine (1kW, Tangshan Nanosource Microwave Thermal Instrument Manufacturing Co. Ltd., Hebei, P. R. China) [13]. The heated brain was manually dissected into 10 different brain regions: cerebellum (CE), midbrain (MID), thalamus (THA), hypothalamus (HYP), hippocampus (HP), striatum (STR), frontal cortex (FC), occipital cortex (OC), parietal cortex (PC), and temporal cortex (TC) [1, 14]. The dissected tissue was weighed, frozen in liquid nitrogen and stored at -40 ˚C until further processing.

2.4 Metabolites extraction from the tissue

The frozen brain tissue extraction was completed using a Tissuelyser (Qiagen TissueLyser, Retsch GmBH, Germany) following a former protocol [13, 14]. Briefly, a tungsten bead was placed in the EP tube with the brain tissue and 200 µl HCl/Methanol (0.1 M) (HCl: Methanol,

Journal Pre-proof

1:100) was added and the mixture homogenized for 90 s. The shaking frequency was set to 20 Hz. Then 800 μl 60% ethanol was added into the tissue mixture and the homogenization procedure was repeated. After that, the mixture was centrifuged at 14,000 *g* for 10 min at 4 °C and the supernatant was collected. The precipitate was re-extracted using 800μl 60% ethanol twice and the supernatant was collected each time. The final supernatant was concentrated using a rotary evaporator for removing organic solvents (Thermo Scientific 2010, Germany). Then the remaining water solution was finally lyophilized with a freezing vacuum dryer (Thermo Scientific, Germany) for 24 h. During the sample extraction, seven samples were lost due to the tungsten bead breaking the EP tube. The dried product was dissolved in 600 μl phosphate buffer (pH=7.2) (600μl D_2O maintaining 0.2 M Na_2HPO_4/NaH_2PO_4, 5mM TSP (2,2,3,3-d(4)-3-(Trimethylsilyl) propionic acid sodium salt) and 2.5 mM [2-^{13}C]-Glycine). At the end, the mixture was centrifuged (4 °C, 14,000 g, 10 min) and the supernatant (550 μL) transferred into a 5 mm nuclear magnetic resonance tube for further NMR analysis.

2.5 Acquisition of NMR spectra

All NMR spectra were acquired at 298 K by POCE (Proton observed carbon editing, ^{1}H-[^{13}C]-NMR) sequence using a 500 MHz liquid State NMR spectrometer (Bruker Biospin, Germany) [15, 16]. The POCE method has been widely used in ^{13}C enrichment studies by metabolic researchers using different kinds of animal models[5]. In brief, after a spin-echo sequence and a decoupling pulse, the proton signals associated with ^{13}C (^{1}H-[^{13}C]) were decoupled and mixed with a proton signal connected to ^{12}C (^{1}H-[^{12}C]). The proton signals were added together and a new spectrum was generated (^{1}H (^{12}C+^{13}C)). After that, a 180° pulse was applied to the ^{13}C in the spin-echo sequence and the signals of ^{1}H-[^{13}C] were decoupled and inverted. A

Journal Pre-proof

different spectrum was thus generated with proton signals associated with ^{12}C and ^{13}C positions (^{1}H (^{12}C-^{13}C)). The total concentration of metabolites was obtained from related proton signals in the first spectrum and the dual concentrations of ^{13}C-labelled metabolites was obtained from the subtraction of these two different spectra during the POCE study.

For the NMR spectra acquisition, the following parameters were applied: NS (Number of Scans): 32; Number of Dummy scans: 8; Sweep width: 16 ppm; Repetition time: 20 ms; Acquisition data: 16K; Echo time: 8 ms.

2.6 NMR analysis

The processing of NMR spectra was completed in Topspin (Version 2.1, Bruker Biospin) and a home-made software NMRSpec [17]. All POCE spectra were imported into Topspin for phase and baseline correction and a different spectrum was also obtained by subtracting the ^{1}H-(^{12}C+^{13}C) and ^{1}H-(^{12}C-^{13}C) spectra. Then, the spectra were loaded into NMRSpec for peak alignment and integration. The ^{13}C enrichments of the metabolites were mainly located in the spectral areas of δ2.86-2.79 (P1), δ2.75-2.65 (P2), δ2.56-2.27(P3), δ2.21-2.00 (P4) and δ1.95-1.87 ppm (P5). Apparently, the concentration of lactate was seriously influenced by the euthanatizing method and its concentration was not stable during the *in vitro* investigation. So, the spectral region related to lactate was discarded during the following statistical classification analysis.

To avoid the influence of noise, the five different regions (P1-P5) with ^{13}C signals were automatically extracted into even buckets (δ=0.004 ppm) and the area of each bucket was automatically calculated in NMRSpec for further statistical analysis. Then the ^{13}C enrichment was calculated with the ratio difference of the spectrum and the sum spectrum. The ratio was collected for further metabolomics studies. The ^{13}C enrichment in a special carbon position of a metabolite

Journal Pre-proof

was automatically obtained from the quotient of the relative peak areas in the POCE spectra in NMRSpec.

To study enrichment rates of metabolites in each brain region, one-way analysis of variance (ANOVA) and post hoc least significant difference were used for enrichment comparison of different brain regions.

2.7 Metabolomics studies with different classification methods

In the current study, there were ten different brain regions involved and a total of 253 samples were measured with NMR spectroscopy. To screen the metabolic kinetic patterns among these different brain regions, the metabolomics study was utilized. However, it is very difficult to use a classical metabolomics study, such as PLS-DA or OPLS-DA, which is mainly utilized for binary classification. To screen metabolic kinetic patterns and find out the most efficient classification tool, there are several machine learning methods available for the classification of multiple groups involved, such as decision tree (DT), multi-layer perceptron (MLP), support vector machine (SVM), Random Forest (RF) and Boost- decision tree (Boost-DT) *etc*. These methods have widely been used for multiple groups classification studies [18]. The machine learning models are implemented on the sklearn toolkit for the Python 3.6 environment. For statistical classification analysis, the data set was randomly divided into the training set (193 spectra) for generating classification models and the rest of the data (60 spectra) was set as the test data for the validation of the constructed models.

DT is an algorithm that classifies and regresses the data through a set of if-then rules, including a series of branch algorithms such as ID3, C4.5 and classification and regression tree (CART). Considering the saving of computing resources, we used CART here and pruned it by

Journal Pre-proof

adjusting the maximum depth and the minimum sample of leaf nodes to prevent it from overfitting. MLP is a kind of artificial neural network (ANN), which fits the targeted mapping relationship through the full connection between hidden layers [19]. We set up two hidden layers, the number of hidden units was 50 and 10 respectively, the optimizer used stochastic gradient descent (SGD) and the initial learning rate was 0.001. SVM is widely used in binary classification problems. In order to deal with multi-classification tasks, we used a one-against-one strategy [20] and adjusted the penalty term to make the SVM achieve the best performance.

3. Results and discussion

3.1 Metabolic enrichments in brain extracts

To obtain total concentrations and metabolic kinetic differences of small molecules in the ten brain regions (253 samples), POCE NMR pulse sequence was utilized to measure the metabolites in brain extracts. In order to reduce the prediction errors, five consecutive regions which contained ^{13}C labeled compounds were selected and calculated for further analysis. A typical form of POCE NMR spectrum with chemical shifts ranging from 2.87 to 1.85 ppm is illustrated in Fig. 1. The spectral peaks under different chemical shifts reflect the concentration levels of different metabolites. The top curve represents the total concentrations of the compounds in a specific brain region and the bottom curve was obtained by subtracting the two ^{1}H-NMR spectra in the POCE data. The resultant spectrum is related to the total concentrations of ^{13}C-labeled metabolites (multiplied by three for better display).

The selected chemical shift sections mainly contained the ^{13}C enrichment information of five compounds in different positions: aspartate (Asp), N-acetylaspartate (NAA), glutamine (Gln),

Journal Pre-proof

glutamate (Glu), and γ-aminobutyric acid (GABA). Glutamate, GABA and Asp were the most important neurotransmitters in the brain, which were closely correlated with cerebral energy metabolism. For example, [1-^{13}C]-glucose was transported across the blood-brain barrier and the membranes of neurons (Glutamatergic neuron or GABAergic neuron) or astrocytes. The ^{13}C label of the [1-^{13}C]-glucose was first transferred to [4-^{13}C]-glutamate. Through the glutamate-glutamine or glutamine-GABA neurotransmitter cycles among neurons and astrocytes compartments, the [4-^{13}C]-glutamate was transferred to [4-^{13}C]-glutamine or [2-^{13}C]-GABA. Then the following metabolites were gradually formed. This cerebral nervous system metabolism is consistent with our ^{13}C enrichment information (Fig. 1). Thus, the metabolic kinetic information was very important for the activities of the neurons and astrocytes. It could reflect the state of the brain associated with the specific function of the brain region. Thus, the metabolic kinetic patterns for these different brain regions were analyzed and discussed in the following sections.

3.2 Metabolomics studies with different discriminate methods

In the current study, there were ten different brain regions involved. The classical metabolomics method mainly focused on binary classification models. Thus, it could not be utilized here to screen the major metabolic kinetic patterns. Furthermore, it was also very difficult to ascertain the best method for better classification performance. To screen out the best approach, there were several multiple classifiers employed, such as the decision tree (DT), multi-layer perceptron (MLP) and support vector machine (SVM), random forest (RF) and Boost-DT.

At first, three basic classifiers were utilized for constructing the discriminant models, such as DT, MLP and SVM. There are five different regions mentioned in the NMR spectra with measurable ^{13}C labeled metabolites (Fig. 1). Thus, they were used to predict the classification

Journal Pre-proof

performance of these different classification models one by one. The results were collected and are displayed in Table 1. It can be seen that part 3 (p3, δ2.56-2.27) has the highest predictability from the test set among these three basic classifiers. This indicates that it may be the major characterization for these different brain regions compared to the other four different parts.

To further verify the prediction results of the models with the combination of multiple parts, the following combination groups were applied to construct different classification models: p1+p2, p3+p4+p5, p2+p3+p4+p5 and p1+p2+ p3+p4+p5. Results of these different models are illustrated in Table 1. It is clear to see that the combination of p2+p3+p4+p5 shows the highest predictability for both training and test sets which are possibly related to the section of p1 containing more noise. At the same time, DT is the best among the five basic classifiers utilized. However, the predictability of the test set was much lower (0.467), which is not promising for the application of this metabolomics method.

To avoid the error of a single DT, two other methods of ensemble learning with increased capability of classification were further adopted, such as the RF and Boost-DT. Between these two methods, the RF classifier could randomly generate multiple DTs using the training set and finally determining the class according to the predicted average. Similarly the Boost-DT uses the DT obtained in the previous step as the initial DT, trains a series of DTs in an iterative manner and updates their weights and finally uses the weighted majority vote to determine the class [21]. For both methods, the best performance was achieved by adjusting the number of trees.

Results of the RF and Boost-DT are demonstrated in Table 1. It is evident that the RF and Boost-DT could obtain much better performance than a single DT in classification. This shows that the ensemble learning method could compensate the error of a single DT in the current study.

Journal Pre-proof

Furthermore, the method of Boost-DT is based on optimizing the result of the DT, thus the test accuracy score is higher than the RF, which randomly generated DT. However, some scores in the training set (1.0) are much higher than the test set (0.333), which may be due to over-fitting the training set caused by its small size and many features. Furthermore, the combinational parameters (p2+p3+p4+p5) also have the best predictability with the Boost-DT method.

In addition, the confusion matrices of the predicted results of the test set for these five different methods with combinational parameters (p2+p3+p4+p5) are illustrated (Fig. 2). Results show that the regions of the PC and TC had the worst classification performance compared to any other method. For the best Boost-DT method, the correct predicted rate for region TC was only 16.7% for the test set and 50% of the samples were predicted in the region of the OC. Among the different brain regions, such as the PC, TC, FC and OC which belong to the whole cortex, their metabolic compounds and neuronal compositions should be similar. Furthermore, there was no obvious border separating these four different regions, especially for PC and TC. These four different regions were manually separated which could generate some systematic errors and cause overfitting during model construction. To avoid this, the brain regions of PC and TC were removed for further analysis.

The above results demonstrate that the Boost-DT with three DT-based classifiers could directly utilize the values of the ratio of spectra to efficiently discriminate these different brain regions. This could make it easier to screen metabolic kinetics with the greatest difference and thus make it convenient for researchers to obtain the major characteristics among different brain regions.

3.3 Analysis of eight brain regions with DT-related methods

Journal Pre-proof

Journal Pre-proof

From the above analysis, regions of the PC and TC had the worst classification performance and were excluded from the dataset for further discriminant analysis. The remaining eight brain regions were used to construct different classification models with the optimized DT-related methods (DT, RF, and Boost-DT). In the end, there were 155 samples left in the training set and 48 samples in the test set.

In this step, only the optimized combinational parameters and all other parameters were utilized to construct the classification models. As an example, a new DT model with whole parameters are illustrated (Fig. 3). The shape of the tree (different branches) is very similar to the previous DT. However, there were fewer branches involved and the tree structure is much shallower which indicates that removing brain regions (PC and TC) could effectively simplify the decision-making process. Finally, the predicted accuracy scores for the training set and the test set of these three DT related classifiers can be found in Table 2, which are much improved compared to previous results. Based on comparisons of these two different combinational parameters in these three different discrimination methods, the p1 part posed a negative impact on the classification results, so it was excluded from the analysis. The confusion matrices of all classifiers are shown (Fig. 4). For models with the optimized combinational parameters (p2+p3+p4+p5), the predictive accuracies of the training set of the random forest and boost-DT reached 100% for both models and the predictive accuracy scores of the test set for the RF and Boost-DT improved further (77.1% and 87.5%). Thus, the results of the Boost-DT show that it had the best-predictability (87.5%) compared with the other different discriminant approaches. It is the best method for screening major characteristics of the metabolic kinetics of different brain regions. Thus, the different brain regions not only have different metabolite concentrations, but they also

Journal Pre-proof

have different metabolic kinetic patterns which are highly associated with the function of the brain. To further investigate the metabolic kinetic information of different brain regions, the ^{13}C metabolic patterns for these ten different brain regions were also investigated.

3.4 ^{13}C labeled metabolic patterns in different brain regions

To clarify the significance of nodes of the classification tree from a bio-metabolic perspective, we investigated the differences among the enrichment rates of the metabolites in each brain region. As shown in the figure (Fig. 5 and Fig .6), the overall difference in brain regions of [3-^{13}C]-glutamate (Glu_3) and [4-^{13}C]-glutamate (Glu_4) was not noticeable. The enrichment rate of Glu_3 was generally maintained around 5%. Among others, compared with OC, HYP (p=0.016), HP (p=0.033), STR (p=0.014), and CE (p =0.032) there was a significant difference. The enrichment rate of Glu_4 was generally maintained between 10% and 20% and there were no significant statistical differences among various brain regions. However, the enrichment rate of [3-^{13}C]-glutamate and [3-^{13}C]-glutamine (Glx_3) had a more notable difference. In the MID group, the enrichment rate of Glx_3 was significantly higher than that in the HYP (p=0.022), HP (p=0.002), STR (p=0.001), PC (p=0.027), TC (p=0.004) and CE (p=0.041). The OC group was significantly higher than the HP (p=0.011), STR (p=0.005) and TC (p=0.021). While the THA group was also significantly increased compared to the HP (p=0.027), STR (p=0.014) and TC (p=0.046).

As for [2-^{13}C]-GABA ($GABA_2$), the OC and CE groups were found to be significantly higher than others. In addition, compared with the HYP, FC (p=0.009) and THA (p=0.044), they were significantly increased. Moreover, the enrichment rate of the [3-^{13}C]-GABA ($GABA_3$) in THA, OC and CE were significantly raised compared to the other groups. There was, nonetheless, no significant statistical differences between the THA, OC and CE.

Journal Pre-proof

The enrichment rate of [3-^{13}C]-aspartate (Asp_3) varied within the group. Nevertheless, the THA group was still significantly elevated compared to all the other groups (p=0.000). Similarly, the HYP group was significantly higher than the FC (p=0.005), PC (p=0.001), OC (p=0.001), TC (p=0.003), MID (p=0.002) and CE (p=0.000). However, CE group was significantly lower than the HP (p=0.030) and STR (p=0.019).

4. Discussion

4.1 Distribution of glutamate in different brain regions

Glutamate is considered to be the main mediator of excitatory signals in the central nervous system of mammals and may be involved in most aspects of normal brain function, including cognition, memory and learning. The brain contains a large amount of glutamate and its content varies from 5 to 15 mmol/kg depending on the brain regions [22]. Moreover, its content is highly sensitive to changes in energy. It achieves dynamic balance through rapid release and utilization from cells. Endogenous glutamate is more evenly distributed than endogenous GABA [12]. Generally, in brain regions with high GABA levels, glutamate content is not too high, due to the glutamate decarboxylase in the GABAergic neurons which converts glutamate to GABA to meet the functional needs and this conversion is not limited by the concentrations [23]. Furthermore, the conversion rate was also related with the passage rate of glucose in different brain regions. Although it is unknown in the current study, it should be another important parameter for the different metabolite enrichments among various brain regions.

4.2 Distribution of GABA in different brain regions

GABA has always been regarded as the main inhibitory neurotransmitter in the adult

Journal Pre-proof

mammalian brain. It plays a basic role in controlling neuronal excitability, information processing, and neuroplasticity [24]. In mammalian brains, GABA is mainly synthesized from glutamate under the catalysis of two glutamate decarboxylases (GAD), GAD65 and GAD67 and is transported by the vesicular neurotransmitter transporter VGAT and then released. The surrounding neurons and glial cells re-uptake GABA through the activity of the GABA transporter (GAT) [25].

In our study, the cortex region, especially the occipital lobe and cerebellar cortex contained higher levels of GABA enrichment, which is consistent with histological staining results of previous studies. The cerebellar cortex is divided into three layers, which are the molecular layer, Purkinje cell layer and granular cell layer respectively from the surface to the interior. The cortex contains five classes of neurons, including the stellate cell, basket cell, Purkinje cell, Golgi cell and granulosa cell. Among them, except for granulosa cells, all major neuron types (four out of five) of the cerebellar cortex are GABA immunoreactive and the cerebellar cortex has been confirmed to contain high-density $GABA_A$ and $GABA_B$ receptors [26], which provide solid support for our experimental results. For the GABA immunostaining experiment on the cerebral cortex of adult rats, the results show that, compared with the frontal cortex (FC) and parietal cortex (PC), the occipital cortex (OC) had a higher neuron density and accordingly, the GABAergic neurons of the occipital lobe also had an increased population. However, the proportions of GABAergic neurons in the three cortical layers were basically the same, all were around 15% [27]. We hold that, apart from the proportion of the GABA, the difference in the density of neurons might also be a reason for the difference in enrichment rates.

It was also worth noting that in the staining experiment, except for the reticular nucleus, most

Journal Pre-proof

of the thalamus only had fewer of the GABA immunostaining signals, while the hypothalamus had a wider staining positive area [28]. In our previous study on metabolites' concentration, we also demonstrated that the hypothalamus of adult rats had higher GABA levels, which is contrary to our result of the enrichment rate. One possible explanation could be that the enrichment rate was not only related to the concentration of chemical metabolites, but also to the number and activity of neurons and GABAergic cells in the thalamus which might have increased activity.

5. Conclusion

In summary, the metabolic kinetic patterns of various brain regions were investigated using discriminant methods combined with the enrichment rates of multiple metabolites. The THA region was quite different from other brain regions. The regions of the CE and OC had higher concentrations of ^{13}C-labeled GABA metabolites, suggesting that these regions of the brain had distribution of more GABAergic neurons. These are consistent with our result that the classifiers were unlikely prone to errors in these three brain regions. However, there was only a trend, but no significant difference between the HYP, HP and STR on any metabolites, which might be the reason why the classifiers performed not so well in these brain regions. Except for the OC region, the differences between the other three regions of the cortex were not significant, resulting in the consistency of the cortex and the imprecision of manual segmentation, which is also consistent with our classification results. The current technology could also be applied in other multiple classes' metabolomics studies and the special metabolic kinetic patterns could provide useful information for brain function research.

Journal Pre-proof

Authors' contributions

J. Wang, Q. Bao, F. Xu and C. Liu contributed to experimental design. D. Zheng, Z. Li, S. Li and X. Li contributed to animal experiment data acquisition and data analysis. D. Zheng, Z. Li, A. Manyande, J. Wang and Q. Bao contributed to data analysis, result interpretation, and writing. G. Kamal and J. Wang contributed new reagents or analytic tools. All authors have read, revised, and approved the final manuscript.

Author Statement

J. Wang, Q. Bao, F. Xu and **C. Liu** contributed to experimental design.
D. Zheng, Z. Li, S. Li and **X. Li** contributed to animal experiment data acquisition and data analysis.
D. Zheng, Z. Li, A. Manyande, J. Wang and **Q. Bao** contributed to data analysis, result interpretation, and writing.
G. Kamal and **J. Wang** contributed new reagents or analytic tools.
All authors have read, revised, and approved the final manuscript.

Conflict of interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Acknowledgement

This work was supported by grants from the National Natural Science Foundation of China (31771193, 31970973), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB32030200), and the Youth Innovation Promotion Association of Chinese Academy of Sciences (Y6Y0021004).

Journal Pre-proof

Journal Pre-proof

Data availability statement

All data generated or analyzed during this study are included in this article.

Journal Pre-proof

Journal Pre-proof

References

[1] J. Wang, H.L. Zeng, H. Du, Z. Liu, J. Cheng, T. Liu, T. Hu, G.M. Kamal, X. Li, H. Liu, F. Xu, Evaluation of metabolites extraction strategies for identifying different brain regions and their relationship with alcohol preference and gender difference using NMR metabolomics, Talanta, 179 (2018) 369-376.

[2] O. Beckonert, H.C. Keun, T.M. Ebbels, J. Bundy, E. Holmes, J.C. Lindon, J.K. Nicholson, Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts, Nat Protoc, 2 (2007) 2692-2703.

[3] J.K. Nicholson, J. Connelly, J.C. Lindon, E. Holmes, Metabonomics: a platform for studying drug toxicity and gene function, Nat Rev Drug Discov, 1 (2002) 153-161.

[4] L. Fan, Mapping the Human Brain: What Is the Next Frontier?, The Innovation, 2 (2021) 100073.

[5] R.A. de Graaf, G.F. Mason, A.B. Patel, K.L. Behar, D.L. Rothman, In vivo 1H-[13C]-NMR spectroscopy of cerebral metabolism, NMR Biomed, 16 (2003) 339-357.

[6] W. Auerbach, M.S. Hurlbert, P. Hilditch-Maguire, Y.Z. Wadghiri, V.C. Wheeler, S.I. Cohen, A.L. Joyner, M.E. MacDonald, D.H. Turnbull, The HD mutation causes progressive lethal neurological disease in mice expressing reduced levels of huntingtin, Hum Mol Genet, 10 (2001) 2515-2523.

[7] J.C. Lievens, B. Woodman, A. Mahal, O. Spasic-Boscovic, D. Samuel, L. Kerkerian-Le Goff, G.P. Bates, Impaired glutamate uptake in the R6 Huntington's disease transgenic mice, Neurobiol Dis, 8 (2001) 807-821.

Journal Pre-proof

[8] S. Lei, R. Powers, NMR Metabolomics Analysis of Parkinson's Disease, Curr Metabolomics, 1 (2013) 191-209.

[9] M.W. Howe, D.A. Dombeck, Rapid signalling in distinct dopaminergic axons during locomotion and reward, Nature, 535 (2016) 505-510.

[10] K.M. Mendez, S.N. Reinke, D.I. Broadhurst, A comparative evaluation of the generalised predictive ability of eight machine learning algorithms across ten clinical metabolomics data sets for binary classification, Metabolomics, 15 (2019) 150.

[11] L. Wu, Z. Niu, X. Hu, H. Liu, S. Li, L. Chen, D. Zheng, Z. Liu, T. Liu, F. Xu, A. Manyande, J. Wang, H. Xia, Regional cerebral metabolic levels and turnover in awake rats after acute or chronic spinal cord injury, FASEB J, 34 (2020) 10547-10559.

[12] J. Wang, L. Jiang, Y. Jiang, X. Ma, G.M. Chowdhury, G.F. Mason, Regional metabolite levels and turnover in the awake rat brain under the influence of nicotine, J Neurochem, 113 (2010) 1447-1458.

[13] J.Y. Yang, H.L. Guo, D.D. Sun, J. Duan, X.P. Rao, F.Q. Xu, A. Manyande, Y.Q. Tang, J. Wang, F. Wang, Elevated glutamate, glutamine and GABA levels and reduced taurine level in a schizophrenia model using an in vitro proton nuclear magnetic resonance method, Am J Transl Res, 11 (2019) 5919-5931.

[14] T. Liu, Z. Li, J. He, N. Yang, D. Han, Y. Li, X. Tian, H. Liu, A. Manyande, H. Xiang, F. Xu, J. Wang, X. Guo, Regional Metabolic Patterns of Abnormal Postoperative Behavioral Performance in Aged Mice Assessed by (1)H-NMR Dynamic Mapping Method, Neurosci Bull, 36 (2020) 25-38.

[15] D.L. Rothman, K.L. Behar, H.P. Hetherington, J.A. den Hollander, M.R. Bendall, O.A.

Journal Pre-proof

Journal Pre-proof

Petroff, R.G. Shulman, 1H-Observe/13C-decouple spectroscopic measurements of lactate and glutamate in the rat brain in vivo, Proc Natl Acad Sci U S A, 82 (1985) 1633-1637.

[16] D.L. Rothman, K.L. Behar, H.P. Hetherington, R.G. Shulman, Homonuclear 1H double-resonance difference spectroscopy of the rat brain in vivo, Proc Natl Acad Sci U S A, 81 (1984) 6330-6334.

[17] Y. Liu, J. Cheng, H. Liu, Y. Deng, J. Wang, F. Xu, NMRSpec: An integrated software package for processing and analyzing one dimensional nuclear magnetic resonance spectra, Chemometrics and Intelligent Laboratory Systems, 162 (2017) 142-148.

[18] L. Lebanov, L. Tedone, A. Ghiasvand, B. Paull, Random Forests machine learning applied to gas chromatography - Mass spectrometry derived average mass spectrum data sets for classification and characterisation of essential oils, Talanta, 208 (2020) 120471.

[19] Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, L.D. Jackel, Backpropagation Applied to Handwritten Zip Code Recognition, Neural Comput, 1 (1989) 541-551.

[20] Y. Liu, Y.F. Zheng, One-against-all multi-class SVM classification using reliability measures, Ieee Ijcnn, DOI (2005) 849-854.

[21] T. Hastie, S. Rosset, J. Zhu, H. Zou, Multi-class AdaBoost, Statistics and Its Interface, 2 (2009) 349-360.

[22] A. Schousboe, V. Lisy, L. Hertz, Postnatal alterations in effects of potassium on uptake and release of glutamate and GABA in rat brain cortex slices, J Neurochem, 26 (1976) 1023-1027.

[23] J.Y. Wu, E. Roberts, Properties of brain L-glutamate decarboxylase: inhibition studies, J

Journal Pre-proof

Neurochem, 23 (1974) 759-767.

[24] F. Fonnum, R.L. Karlsen, D. Malthe-Sørenssen, K.K. Skrede, I. Walaas, Localization of Neurotransmitters, Particularly Glutamate, in Hippocampus, Septum, Nucleus Accumbens and Superior Colliculus, Development and Chemical Specificity of Neurons1979, pp. 167-191.

[25] D.F. Owens, A.R. Kriegstein, Is there more to GABA than synaptic inhibition?, Nat Rev Neurosci, 3 (2002) 715-727.

[26] N.G. Bowery, A.L. Hudson, G.W. Price, GABAA and GABAB receptor site distribution in the rat central nervous system, Neuroscience, 20 (1987) 365-383.

[27] C. Beaulieu, Numerical data on neocortical neurons in adult rat, with special reference to the GABA population, Brain Res, 609 (1993) 284-292.

[28] O.P. Ottersen, J. Stormmathisen, Glutamate-Containing and Gaba-Containing Neurons in the Mouse and Rat-Brain, as Demonstrated with a New Immunocytochemical Technique, J Comp Neurol, 229 (1984) 374-392.

Journal Pre-proof

Journal Pre-proof

Figure captions:

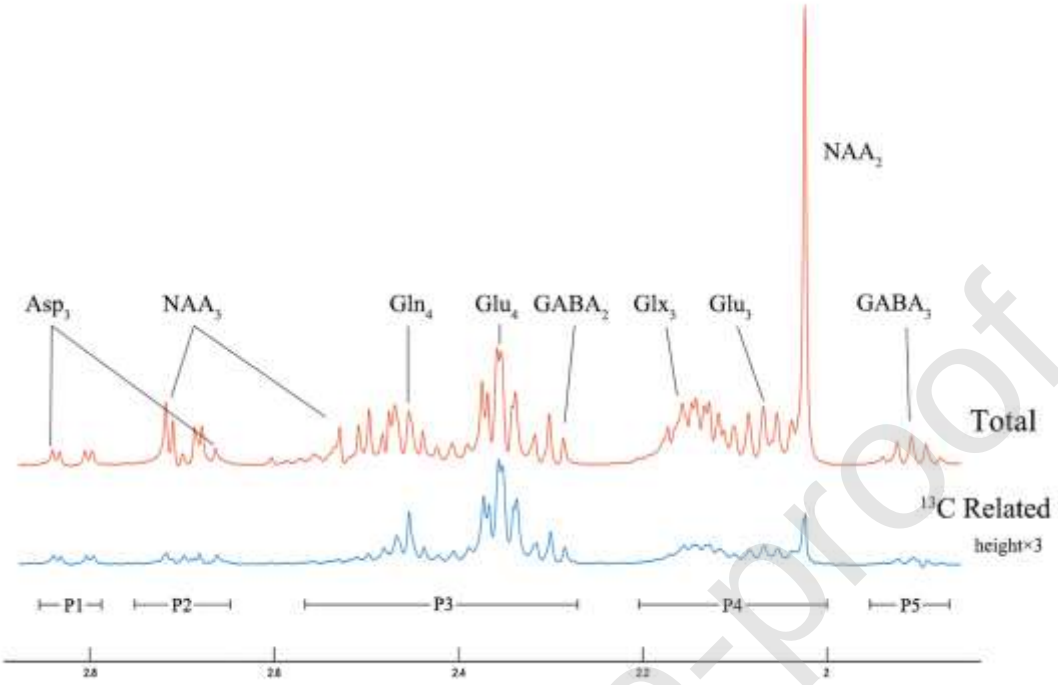


Fig. 1: NMR spectra at the chemical shift of 2.87~1.85 ppm for total metabolites ($^{12}C+^{13}C$, red) and ^{13}C related metabolites (2*^{13}C, and then height*3, blue) by the POCE (^{1}H-observed/^{13}C-edited) NMR sequence. The ^{13}C-labeled metabolites are mainly located in the areas of δ2.86-2.79(P1), δ2.75-2.65(P2), δ2.56-2.27(P3), δ2.21-2.00(P4) and δ1.95-1.87(P5). *Note: The isotope-labeled carbon atom is denoted by the subscript. Asp: Aspartate; NAA: N-acetylaspartate; Gln: glutamine; Glu: glutamate; GABA: γ-aminobutyric acid; Glx: glutamine and glutamate.*

Journal Pre-proof

Journal Pre-proof

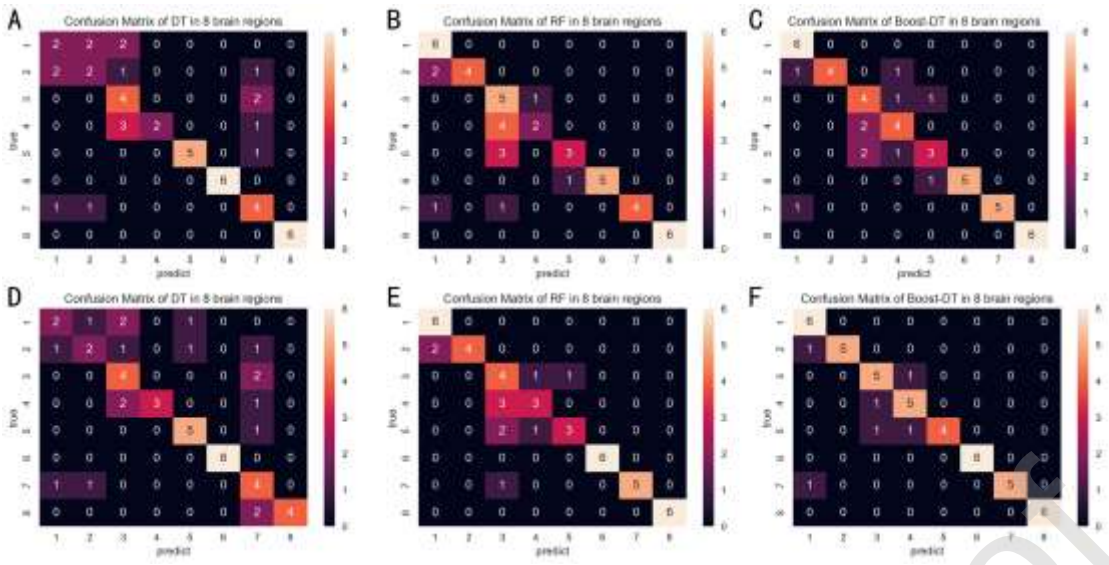


Fig. 2: Confusion matrices of the results of the test set (n=6) with combinational areas of P2+P3+P4+P5 using five different classifiers. The horizontal axis represents the test results of 5 classifiers and the vertical axis represents the correct results. (A) Decision Tree (DT). (B) Multi-layer Perceptron (MLP). (C) Support Vector Machine (SVM). (D) Random Forest (RF). (E) Boost-DT. *Note: The numerals from 1 to 10 of the axes symbolize the 10 different brain regions. 1: Thalamus (THA); 2: Hypothalamus (HYP); 3: Hippocampus (HP); 4: Striatum (STR); 5: Frontal Cortex (FC); 6: Parietal Cortex (PC); 7: Occipital Cortex (OC); 8: Temporal Cortex (TC); 9: Midbrain (MID); 10: Cerebellum (CE).*

Journal Pre-proof

Journal Pre-proof

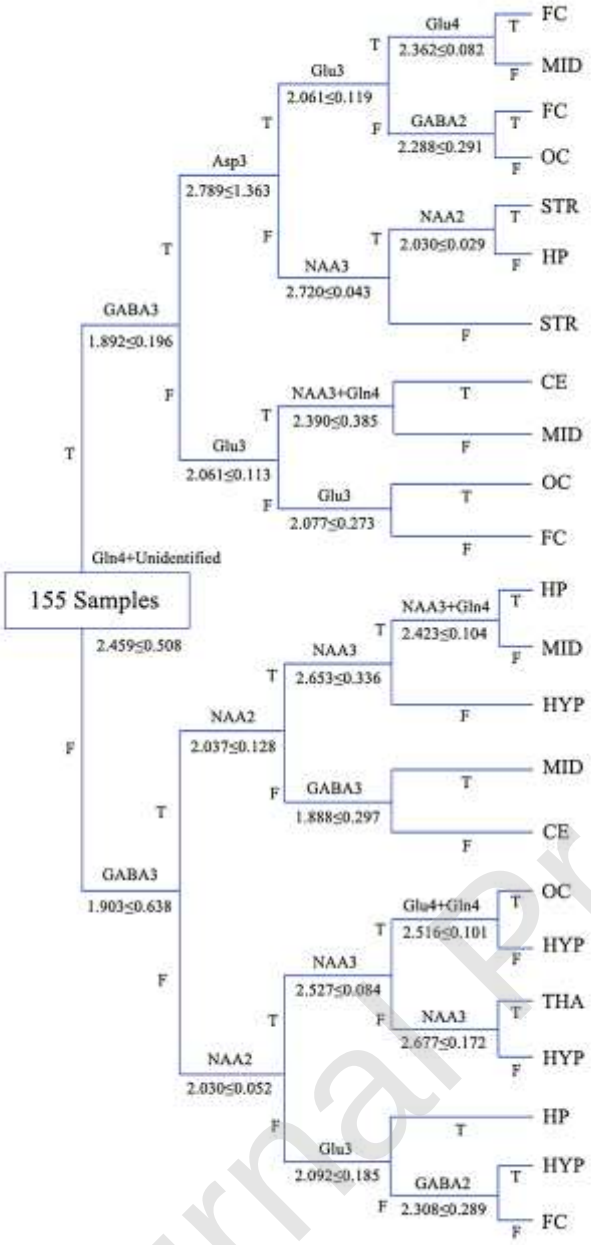


Fig. 3: One of the models of Decision Tree (DT) of the training sets (n=155) with 8 brain regions. The left side of the conditional statement is the chemical shift of the integral value to be compared. “T” means true and “F” means false.

Journal Pre-proof

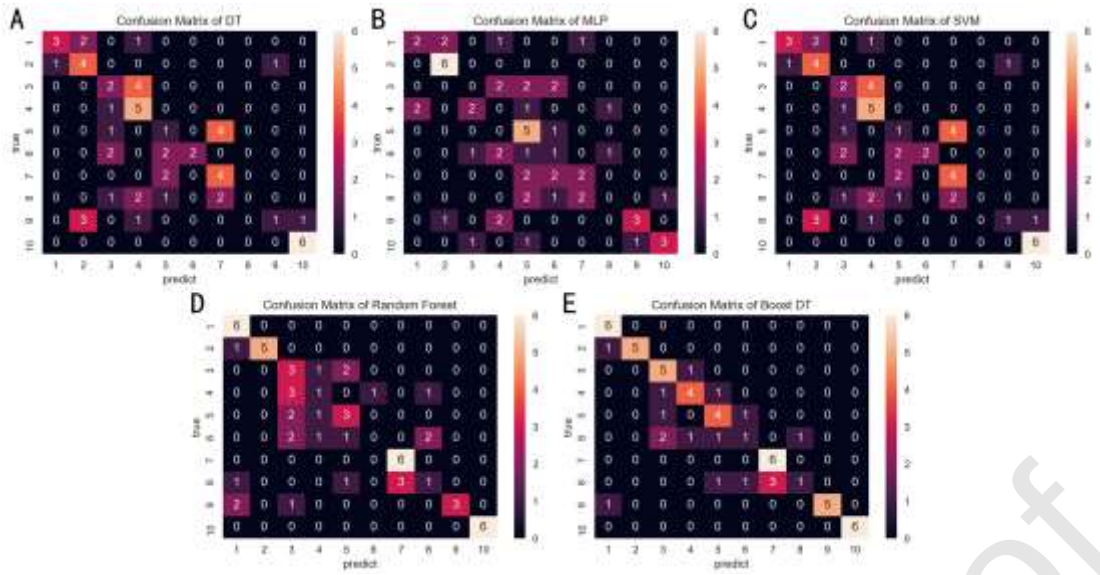


Fig. 4: Confusion matrices of 8 brain regions with spectrum of all (P1+P2+P3+P4+P5) or P2+P3+P4+P5 areas by 3 DT-based classifiers. All areas combined (P1+P2+P3+P4+P5): (A) Decision Tree; (B) Random Forest; (C) Boost-DT. Part 2,3,4 and 5 combined: (D) Decision Tree; (E) Random Forest; (F) Boost-DT. *Note: The numerals from 1 to 8 of the axes symbolize the 8 different brain regions. 1: Thalamus (THA); 2: Hypothalamus (HYP); 3: Hippocampus (HP); 4: Striatum (STR); 5: Frontal Cortex (FC); 6: Occipital Cortex (OC); 7: Midbrain (MID); 8: Cerebellum (CE).*

Journal Pre-proof

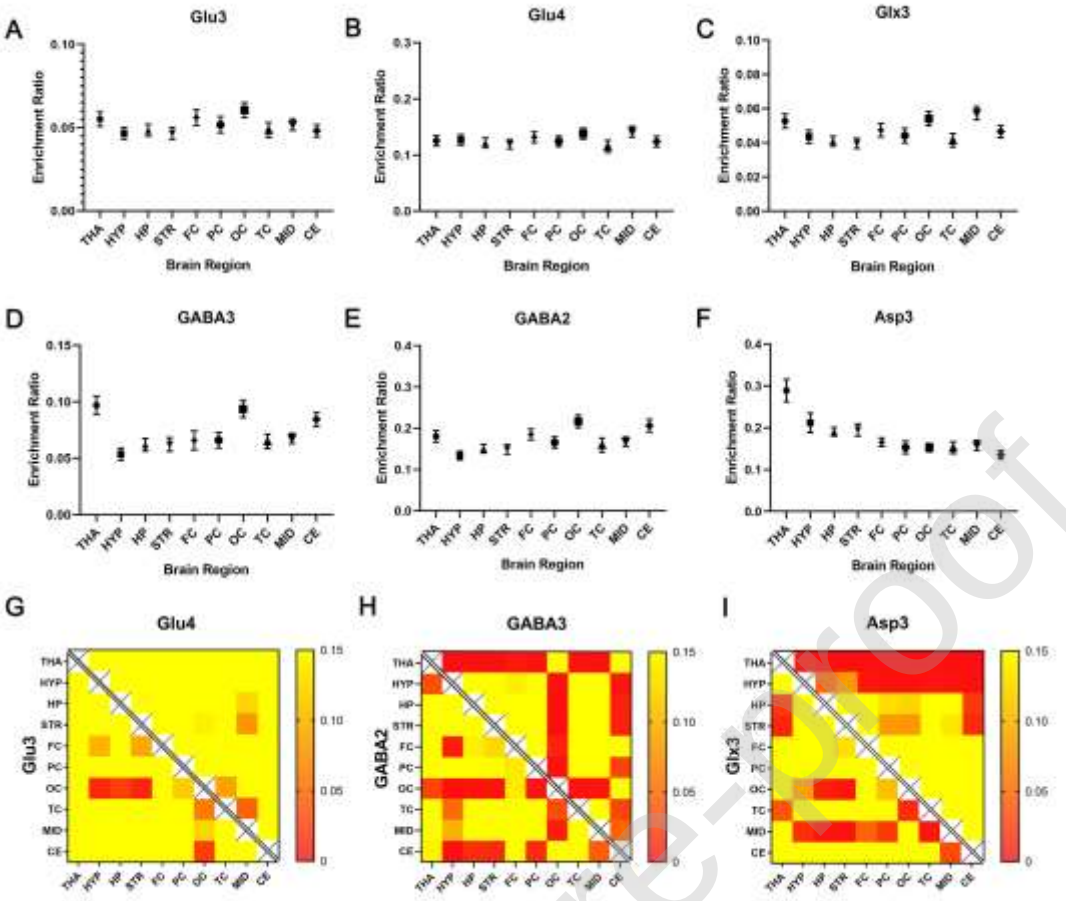


Fig. 5: The enrichment ratio of metabolites: (A) *Glu3*, (B) *Glu4*, (C) *Glx3*, (D) *GABA2*, (E) *GABA3 and* (F) *Asp3* in 10 different brain regions. (G)~(I): The p-values of ANOVA (Least-Significant Difference).

Journal Pre-proof

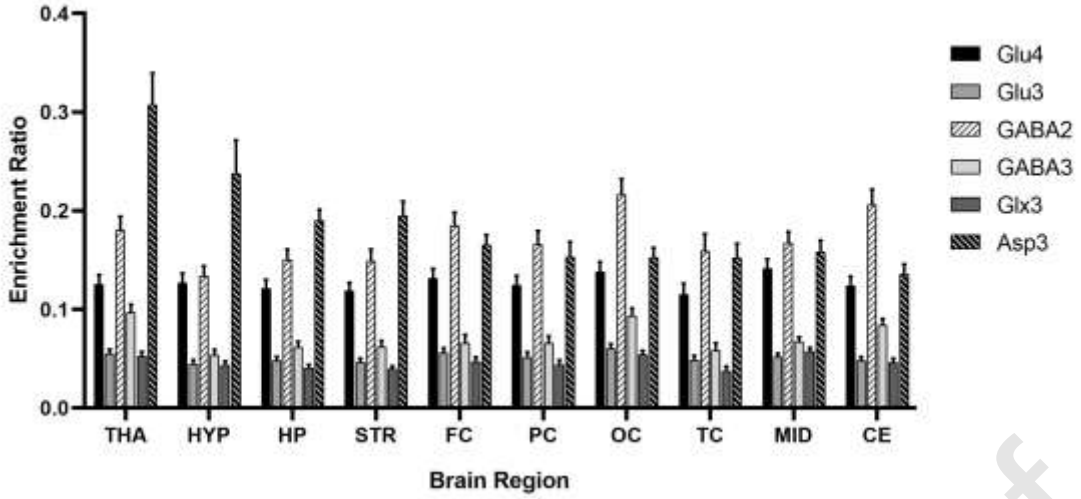


Fig. 6: The enrichment of metabolites *Glu3, Glu4, Glx3, GABA2, GABA3* and *Asp3* in 10 different brain regions.

Journal Pre-proof

Journal Pre-proof

Table 1. The classification performance of a single spectrum and combinations of spectrum by different classifiers, training accuracy/test accuracy

	CART	MLP	SVM	RF	Boost-CART
Part 1 (P1)	0.645/0.200	0.312/0.150	0.197 /0.133	0.980/0.300	1.000/0.317
Part 2 (P2)	0.617/0.233	0.259/0.217	0.197/0.133	1.000/0.400	1.000/0.400
Part 3 (P3)	0.819/0.533	0.306/0.233	0.829/0.433	1.000/0.550	1.000/0.600
Part 4 (P4)	0.829/0.450	0.288/0.233	0.585/0.233	0.969/0.483	1.000/0.583
Part 5 (P5)	0.661/0.333	0.283/0.264	0.560/0.350	0.959/0.417	1.000/0.517
P1+P2	0.745/0.200	0.8340.150	0.264/0.133	0.803/0.317	1.000/0.333
P3+P4+P5	0.829/0.450	0.551/0.483	0.902/0.367	1.000/0.600	1.000/0.700
P2+P3+P4+P5	0.927/0.467	0.601/0.367	0.912/0.383	1.000/0.617	1.000/0.717
ALL	0.917/0.450	0.560/0.283	0.896/0.233	1.000/0.667	1.000/0.667

Table 2. The classification performance of all spectrum and p2+p3+p4+p5 by DT-based classifiers.

	DT	RF	Boost-DT
All	74.8%/64.6%	99.4%/72.9%	100%/79.1%
p2+p3+p4+p5	82.6%/62.5%	100%/77.1%	100%/87.5%

Note: The percentage before "/" was the predicted accuracy of the training set, and that after "/" was the predicted accuracy of the test set,

Journal Pre-proof