Identifying Methamphetamine Abstainers With Convolutional Neural Networks and Short-Time Fourier Transform

Xin Lai1†, Qiuping Huang2,3†, Jiang Xin1†, Hufei Yu1, Jingxi Wen1, Shucai Huang2,3,4, Hao Zhang1, Hongxian Shen2,3* and Yan Tang1*

1 School of Computer Science and Engineering, Central South University, Changsha, China, 2 National Clinical Research Center for Mental Disorders, Department of Psychiatry, The Second Xiangya Hospital of Central South University, Changsha, China, 3 Institute of Mental Health of Central South University, Chinese National Technology Institute on Mental Disorders, Hunan Key Laboratory of Psychiatry and Mental Health, Hunan Medical Center for Mental Health, Changsha, China, 4 The Fourth People’s Hospital of Wuhu, Wuhu, China

Few studies have investigated the functional patterns of methamphetamine abstainers. A better understanding of the underlying neurobiological mechanism in the brains of methamphetamine abstainers will help to explain their abnormal behaviors. Forty-two male methamphetamine abstainers, currently in a long-term abstinence status (for at least 14 months), and 32 male healthy controls were recruited. All subjects underwent functional MRI while responding to drug-associated cues. This study proposes to combine a convolutional neural network with a short-time Fourier transform to identify different brain patterns between methamphetamine abstainers and controls. The short-time Fourier transformation provides time-localized frequency information, while the convolutional neural network extracts the structural features of the time–frequency spectrograms. The results showed that the classifier achieved a satisfactory performance (98.9% accuracy) and could extract robust brain voxel information. The highly discriminative power voxels were mainly concentrated in the left inferior orbital frontal gyrus, the bilateral postcentral gyri, and the bilateral paracentral lobules. This study provides a novel insight into the different functional patterns between methamphetamine abstainers and healthy controls. It also elucidates the pathological mechanism of methamphetamine abstainers from the view of time–frequency spectrograms.

Keywords: methamphetamine abstainers, deep learning, short-time Fourier transform, functional magnetic resonance imaging, drug cues

INTRODUCTION

Methamphetamine (MA) is a highly addictive stimulant with a continuously increased production and that is abused globally. MA addiction can lead to anxiety, depression, and psychosis (Zweben et al., 2004), causing structural and functional changes in the brain (Salo and Fassbender, 2012). Brain imaging provides valuable information on the neurobiological effects of drug abuse and helps explain the causes and mechanisms of vulnerability to drug abuse (Weinstein et al., 2016); however, there is limited neuroimaging research on the structural and functional recovery of the brain of long-term abstaining MA-dependent individuals. To overcome this barrier, we collected...
neuroimaging data from abstaining MA-dependent individuals (for at least 14 months) during the recovery period. Brain differences between healthy controls (HCs) and abstaining MA-dependent individuals were then studied to evaluate the risk of relapse and to understand the neurological impact of MA abuse.

Functional MRI (fMRI) is often used to study the functional characteristics of various brain regions in specific task states (Cassidy et al., 2018; Dvornek et al., 2018; Gui and Gui, 2020; Yousefnezhad et al., 2020; Yotsutsuji et al., 2021). Traditional statistical methods, such as the two-sample t-test, have played a significant role in locating abnormal brain regions associated with psychiatric disorders by looking at the brain activation levels during a task performance (Jiang et al., 2013). Usually, hypothesis testing may not fully capture the underlying group differences when there is a non-linear relationship (Smucny et al., 2021); hence, as an alternative, researchers have proposed to study neuronal oscillations or frequency responses associated with physiological functions (Han et al., 2011; Tang et al., 2016), where frequency analysis has been widely used to reveal the pathophysiology of mental diseases (Chang and Glover, 2010; Bolton and Van De Ville, 2019). These studies have highlighted the importance of studying the intrinsic brain activity within specific frequency bands for the resting-state fMRI. In this study, because long-term abstaining MA-dependent individuals show little difference from HCs using the traditional statistical methods, we propose to identify potential brain abnormality by exploring the spatiotemporal brain activation patterns within specific frequency bands. Short-time Fourier transform (STFT), a commonly used analysis method in physiological signal processing to extract information in the time–frequency domain (Sato et al., 2020), was used to study the changing spectra over time of MA-dependent individuals and HCs. STFT was used to analyze the dynamic changes of the frequency and phase information of a non-stationary signal (Subbarao and Samundiswary, 2016; Seeliger et al., 2018), which has many successful applications in pattern recognition, such as in speech (Takaki et al., 2019) and action (Klejmova and Pomenkova, 2017).

Regarding classifiers, we first resorted to traditional machine learning algorithms, such as support vector machine (SVM) and logistic regression (LR), which have shown excellent performance in the individual-level disease diagnosis (Pfister et al., 2011; Huang et al., 2015, 2016; Wang et al., 2015). However, we found it to be challenging to incorporate prior knowledge to extract biologically meaningful information from subtle changes, especially when it comes to features in the STFT spectrogram. Recently, deep convolutional neural networks (CNNs) have outperformed the traditional machine learning algorithms in capturing subtle changes in features in the network structure. CNN can extract non-linear network structures, can realize the approximation of complex functions, can characterize the distributed representation of input data, and can demonstrate the powerful ability to learn the critical features of datasets. It has been widely used in various fields, such as in the medical field (Fan et al., 2017). It is a very promising avenue offering better results compared with other conventional machine learning or statistical methods.

When granted the above considerations, we proposed an STFT-based CNN model to explore the abnormal brain regions in MA abstainers under the stimulation of drug-associated cues. Our model converted information on the brain area activation in fMRI into the STFT spectrograms, which were then fed into a CNN to generate the recognition results. Finally, we performed a 10-fold cross-validation to eliminate the interference of overfitting.

**METHODS**

**Participants**

The study included 42 male MA-dependent individuals (aged 19–45 years) currently with a long-term abstinence status (at least 14 months) and 32 male HCs of similar age and education. Enrolment criteria for the two group is shown in Table 1. Exclusion criteria for all the subjects included (1) hallucinations, delusions, depression, anxiety, and other psychiatric symptoms; (2) a previous history of other axis I disorders (such as schizophrenia, depression, bipolar disorder, and mania); (3) a previous history of medical or physical therapy affecting brain functions, prescribed by psychiatry, neurology, or other specialties within the last 3 years; (4) a previous history of brain tumor, brain trauma, and other organic brain diseases; (5) a history of seizures, such as epilepsy, coma, high fever, and convulsions; (6) metabolic and endocrine diseases, cardiovascular diseases, and other physical diseases that affect brain functions; (7) metal or electromagnetic implants in the body, claustrophobia, and other conditions that are not suitable for the magnetic resonance examination; and (8) homosexuality. The detailed exclusion criteria are described in Chen et al. (2020).

This study was approved by the ethical review board of the Second Xiangya Hospital of Central South University, which evaluated the study specifically related to the participation of individuals and conditions for the use of incarcerated individuals in research. Participants could decline their involvement in the study if they had any concerns, and all the participants provided voluntary written informed consent.

**Experimental Design and Procedures**

A block design was adopted with a total of six sessions; each session contained one 20-s task block showing MA cue images and one 15-s rest block showing a crosshair. The MA cue...
images contained MA-related contents, e.g., people using MA, instruments used to consume MA, etc. Each block contained five unique images presented for 3 s with a 1-s interstimulus interval. Thus, a total of 30 images for each cue condition were presented during the fMRI scan. The images and blocks within sessions were randomly arranged.

**Image Acquisition and Preprocessing**

All the fMRI scan images were collected at the Medical Imaging Department of the Second Xiangya Hospital of Central South University using a 3.0 Tesla Siemens MRI scan system. The subjects were required to lie flat on an examination bed equipped with a magnetic resonance scanner during the examination, wearing sponge earplugs and noise-proof headphones to reduce noise. An elastic sponge was used to fix the sides of the head and to reduce head movement. The fMRI data were synchronously collected while the subjects viewed the images. The parameters of magnetic resonance data acquisition were as follows: TR = 2000 ms, TE = 20 ms, the field-of-view (FOV) = 220 mm, matrix = 64 × 64, flip angle = 80°, voxel size = 3.4 × 3.4 × 3.4 mm³, slice thickness = 4 mm, and number of slices = 36. Interval scanning was employed, i.e., alternating scans of the even- and odd-numbered layers. A total of 60 time points were collected.

DPARSF software (Chao-Gan and Yu-Feng, 2010) was used to preprocess the task fMRI data such as slicing time, head motion correction, spatial normalization with a 3 × 3 × 3 mm³ EPI template, and spatial smoothing with Gaussian kernel (FWHM = 6 mm). Participants met the standard by limiting their head motion within 2.5 mm. The number of displayed images of the drug-associated cues was 60.

**Development of the Discriminate Model**

We divided the subjects into 10 groups, and 90% of the subjects were used for model training and the remaining 10% were used for model testing. We developed a data-driven classifier to further explore the group differences using a CNN model, which consists of four steps: feature selection, feature transformation, model identification, and cross-validation (Figure 1).

**Feature Selection**

Considering the characteristics of huge voxels in the brain image, feature selection was used to obtain the locations of some significant voxels. To validate our algorithm, we only used the training data to build the generalized linear model (GLM) (first-level analysis). A fixed-effect boxcar waveform was convolved with the hemodynamic response function to produce a matrix to model categorical BOLD responses (Penny et al., 2003). In this procedure, a high-pass filter of 1/128 Hz was used to remove the low-frequency noise, and an AR (1) model was used to correct for temporal autocorrelations. To account for the residual motion artifacts, we included six motion regressors in our first-level model. Following the abovementioned procedure, neural activities associated with the drug cues were found. Thereafter, a two-sample t-test (Jiang et al., 2013) was conducted to compare abstaining MA-dependent individuals with HC in the training data set to select the locations of voxels. When the two-sample t-test was used on the training data, there was no significant difference (p < 0.001 uncorrected) in the brain voxels. In this study, we only kept the locations of voxels with t-values greater than three and cluster sizes greater than five. The chosen number of voxels was denoted by £N, and we obtained the locations of these £N voxels through feature selection.

**Feature Transformation**

Functional MRI time series is often performed by extracting the values of voxels after preprocessing. In this study, STFT was performed on the fMRI time series to investigate the time–frequency information for each subject using the
Allen et al.,

The network structure of convolutional neural network (CNN).

To train the CNN model, we used cross-entropy as the loss function, which can be formulated as given by Equation (2).

\[
L = - \left[ y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i) \right]
\]

Cross-Validation

Because of the limited number of samples in this study, we used a cross-validation strategy to estimate the generalized performance of our classifier. Here, the performance of the classifier in the model for spectrograms was represented as \( \text{Acc} \). We used permutation tests to assess the statistical significance of the cross-validation results. For permutation testing, the classification labels of the training data were randomly permuted 10,000 times. Cross-validation was then performed on every permuted training set. \( \text{Acc}_0 \) was defined as the accuracy rate obtained by the classifier trained on the real class labels. When \( \text{Acc}_0 \) exceeded the 95% (\( P < 0.05 \)) CI of the classifier trained on the randomly relabeled class labels, it was assumed that the classifier had reliably learned the relationship between the data and the labels. For any value of the estimated \( \text{Acc}_0 \), the \( p \)-value represented the probability of observing a classification prediction rate of no less than \( \text{Acc}_0 \).

Considering that a single subject produced many time–frequency spectrograms probably belonging to different categories, we used Equation (3) to determine the category of a subject

\[
\text{Pre}_i = \frac{T_i}{T_i + F_i},
\]

where \( \text{Pre}_i \) represents the probability that subject \( i \) was successfully identified as an MA abstainer, \( T_i \) represents the number of time–frequency spectrograms for subjects that had been judged to be MA abstainers, and \( F_i \) represents the number of time–frequency spectrograms for subject \( i \) that were judged to be normal. When \( \text{Pre}_i \) was <0.5, we assumed that subject \( i \) was normal; otherwise, subject \( i \) was an MA abstainer. The average classification accuracies of the subjects were taken as the final result.
Comparisons With Some Methods
To justify the effectiveness of the proposed method, some existing methods were tested on our dataset as comparisons. First, the traditional statistical method, a two-sample t-test, popular for data analysis in neuroimaging studies (Jiang et al., 2013), was compared with the proposed method. After data preprocessing, we divided all data into two groups (MA abstainers vs. matched controls) and used the two-sample t-test to identify the brain regions that showed the statistically significant MA-related differences (uncorrected \( p < 0.001 \)). Second, considering that LR and SVM are widely used machine learning methods in psychiatry (Zhou et al., 2020), we used them as classifiers taking the STFT spectra as inputs. In the parameter setting of SVM, we set the penalty coefficient C to 1.0 and employed a Gaussian kernel function. In LR, we used L2 regularization with the penalty coefficient C equal to 1.0. We used the default values for other parameters. The classification accuracies of SVM and LR were reported after a 10-fold cross-validation.

RESULTS
Classification Results
The detailed classification results of the cross-validation method used to verify the model recognition effect are shown in Table 2, where \( \text{Acc}_{\text{CNN}} \), \( \text{Acc}_{\text{SVM}} \), and \( \text{Acc}_{\text{LR}} \) represent the classification accuracy of spectrograms with CNN, SVM, and LR, respectively. We took the average of the model classification accuracy obtained during model verification. The average accuracy rate in our model was 93.4%, which was much better than that obtained using SVM and LR, whose average classification accuracies were only 76.2 and 72.0%, respectively.

Permutation tests revealed that the proposed classifier learned the relationship between the data and the labels with an accuracy higher than 95%.

The recognition accuracy rate in our model, calculated using Equation (3) in “Cross-validation” section, ranged between 88.9 and 100% with an average value of 98.9%.

When we used the traditional statistical method (two-sample t-test) to detect the abnormal brain region, we could not find any statistically significant MA-related difference in this dataset.

Brain Regions With a High Discriminative Power
Because the performance of the classifier was tested with a cross-validation strategy, the selected voxels might be different in separate iterations. The voxels that were included across all iterations were reported. Surprisingly, most of the voxels presented a clustered distribution. The difference distribution was mainly concentrated in the left inferior orbital frontal gyrus, the bilateral postcentral gyri, and the bilateral paracentral lobules. These areas mainly control the movements and emotions of the people.

DISCUSSION
In this study, we selected 42 MA abstainers and detected the differences in the brain activation regions when they saw drug cues. The result is shown in Figure 3. We achieved an average accuracy rate of 98.9% using STFT and CNN. The left inferior orbital frontal gyrus, the bilateral postcentral gyrus, and the bilateral paracentral lobule gyri were associated with drug cues in MA abstainers.

Neuroimaging techniques have a high potential to detect brain deficits and correlations between the deficient brain regions and the cognitive–behavioral performance in MA abstainers. However, for MA abstainers, partial functions of the brain return to normal, and group-level statistical methods such as the two-sample test cannot detect any statistically significant MA-related difference in this dataset (Allen et al., 2014). However, using STFT and CNN, we found significant MA-related differences with the average accuracy rate reaching 98.9% in the cross-validation. The STFT method was used to analyze the time–frequency changes in the brain voxel signals. STFT is a time–frequency analysis technique suited to non-stationary signals and provides time-localized frequency information for situations in which the frequency components of a signal vary over time. Thus, this might be ascribed to the essentially non-linear neural dynamics of time–frequency changes underlying the brain activity. A similar conclusion was also obtained for MA abusers using electroencephalogram (EEG) (Khajehpour et al., 2019). In addition, without the pre-engineered features, CNN can conduct local perception and extract the spatial structural features of the time–frequency spectrograms. CNN methods have the potential to scale well and substantially improve the classification performance compared with SVM and LR methods (Abrol et al., 2021). Our findings highlight the presence of non-linearities and time–frequency changes in neuroimaging data that CNN can exploit as discriminative representations to characterize the MA abstainers.

Numerous MRI studies have documented that addictive drugs cause volume and tissue composition changes in the left inferior orbital frontal gyrus, which is associated with a longer duration of use of the MA (Volkow et al., 2015). Changes in the left inferior orbital frontal gyrus are likely associated with the cognitive and decision-making problems of the abusers. Moreover, this impairment of cognitive and altered decision-making in MA abstainers may result in relapse (Mizoguchi and Yamada, 2019), although this region may partially recover from long-term abstinence (Chang et al., 2005). In this study, we still detected different time–frequency spectrograms using the CNN model, i.e., MA abstainers were still influenced by the drug-associated

| Table 2 | A 10-fold cross-validation accuracy of support vector machine (SVM) and convolutional neural network (CCN). |
|---|---|---|---|---|---|---|---|---|---|
| Fold | \( \text{Acc}_{\text{CNN}} \) | \( \text{Acc}_{\text{SVM}} \) | \( \text{Acc}_{\text{LR}} \) | Fold | \( \text{Acc}_{\text{CNN}} \) | \( \text{Acc}_{\text{SVM}} \) | \( \text{Acc}_{\text{LR}} \) |
| 1 | 0.981 | 0.625 | 0.758 | 6 | 0.923 | 0.714 | 0.738 |
| 2 | 0.962 | 0.875 | 0.693 | 7 | 0.894 | 1.000 | 0.762 |
| 3 | 0.946 | 0.750 | 0.749 | 8 | 0.892 | 0.714 | 0.654 |
| 4 | 1.000 | 0.628 | 0.754 | 9 | 0.936 | 0.571 | 0.715 |
| 5 | 0.975 | 0.571 | 0.697 | 10 | 0.833 | 0.875 | 0.678 |
cues, a deficit that is related to dysfunctions of the left inferior orbital frontal gyrus (Volkow et al., 2015).

The paracentral lobule controls the motor and sensory innervations, and the postcentral region is located in the somatosensory cortex. When these regions are impaired, the executive control systems may be affected as demonstrated by Volkow et al. (2015) by specific impairments within the executive brain networks in MA addicts during the exposure to drug-associated cues. Khajehpour et al. (2019) also reported that MA abusers differed in the gamma band in the paracentral lobule. It may be speculated that the substance-dependent individuals are unable to control their addiction-related behaviors (Khajehpour et al., 2019).

This study not only demonstrated a high classification accuracy of the STFT–CNN classifier from a drug–cue functional integration viewpoint but also elucidated the pathological mechanisms of the MA abstainers in a non-linear time–frequency characteristic. In the future, we will test this method on a larger independent dataset to confirm our findings.

**DATA AVAILABILITY STATEMENT**

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

**REFERENCES**

Abrol, A., Fu, Z., Salman, M., Silva, R., Du, Y., Plis, S., et al. (2021). Deep learning encodes robust discriminative neuroimaging representations to outperform standard machine learning. *Nat. Commun.* 12:353. doi: 10.1038/s41467-020-20655-6

Allen, E. A., Damaraju, E., Plis, S. M., Erhardt, E. B., Eichele, T., and Calhoun, V. D. (2014). Tracking whole-brain connectivity dynamics in the resting state. *Cereb. Cortex* 24, 663–676. doi: 10.1093/cercor/bht352

Bolton, T. A. W., and Van De Ville, D. (2019). “Time-frequency characterization of resting-state brain function reveals overlapping components with specific topology and frequency content,” in *Proceedings of the 2nd International Conference on Image and Graphics Processing—ICIGP ’19* (Singapore), 84–88.

Cassidy, B., Bowman, F. D., Rae, C., and Solo, V. (2018). On the reliability of individual brain activity networks. *IEEE Trans. Med. Imaging* 37, 649–662. doi: 10.1109/TMI.2017.2774364

Chang, C., and Glover, G. H. (2010). Time-frequency dynamics of resting-state brain connectivity measured with fMRI. *Neuroimage* 50, 81–98. doi: 10.1016/j.neuroimage.2009.12.011

**ETHICS STATEMENT**

The studies involving human participants were reviewed and approved by the ethical review board of the Second Xiangya Hospital of Central South University. The patients/participants provided their written informed consent to participate in this study.

**AUTHOR CONTRIBUTIONS**

XL, YT, and JX contributed to the conception of the study. XL, HY, JW, and JX designed the study. QH and SH collected the data. YT and XL performed data analysis and drafted the manuscript. HZ and HS modified the manuscript. All authors contributed to the article and approved the submitted version.

**FUNDING**

YT was supported by grant MIMS20-08 from the Research Fund of the Guangxi Key Lab of Multi-source Information Mining and Security. HS was supported by National Natural Science Foundation of China (No. 81971249), National Research Program of China (No. 2016YFC0800908-Z02), and National Natural Science Foundation of Hunan Province (No. 2020JJ4782).
Chang, L., Cloak, C., Patterson, K., Grob, C., Miller, E. N., and Ernst, T. (2005). Enlarged striatum in abstinent methamphetamine abusers: a possible compensatory response. *Biol. Psychiatry* 57, 967–974. doi: 10.1016/j.biopsych.2005.03.039

Chao-Gan, Y., and Yu-Feng, Z. (2010). DPARSF: a MATLAB toolbox for “Pipeline” data analysis of resting-state fMRI. *Front. Syst. Neurosci.* 4:13. doi: 10.3389/fnsys.2010.00013

Chen, S., Huang, S., Yang, C., Cai, W., Chen, H., Hao, W., et al. (2020). Neurofunctional differences related to methamphetamine and sexual cues in men with shorter and longer term abstinence methamphetamine dependence. *Int. J. Neuropsychopharmacol.* 23, 135–145. doi: 10.1093/ijnpyz069

Dvornek, N. C., Yang, D., Ventola, P., and Duncan, J. S. (2018). Learning generalizable recurrent neural networks from small task-fMRI datasets. *Med. Image Comput. Comput. Assist. Interv.* 11072, 329–337. doi: 10.1007/978-3-030-09931-1_38

Fan, L., Xia, Z., Zhang, X., and Feng, X. (2017). "Lung nodule detection based on 3D convolutional neural networks," in 2017 International Conference on the Frontiers and Advances in Data Science (FADS) (Xi'an), 7–10.

Gui, S., and Gui, R. (2020). Utilizing wavelet deep learning network to classify non-stationary signals using frequency slice wavelet transform, " in 2016 10th International Conference on Intelligent Systems and Control (ISCO) (Coiimbatore), 1–6.

Huang, X., Zhao, G., and Pietikäinen, M. (2015). "Facial micro-expression recognition using spatiotemporal local binary pattern with integral projection," in 2015 IEEE International Conference on Computer Vision Workshop (ICCVW) (Santiago), 1–9.

Khajehpour, H., Makkiabadi, B., Ekhtiari, H., Bakht, S., Noroozi, A., and Weizman, A. (2016). Brain imaging studies on the cognitive, pharmacological and neurobiological effects of cannabis in humans: evidence from studies of adult users. *Curr. Pharm. Des.* 22, 6366–6379. doi: 10.2174/1381612822666160822151323

Lai, L., Huang, J., Xin, W., Wen, H., Huang, Z., Zhang, S., Shen, S., and Tang, X. (2017). Identification of a time-varying curve in spectrogram. *Radioengineering* 26, 291–298. doi: 10.13164/rev.2017.0291

Mizoguchi, H., and Yamada K. (2019). Methamphetamine use causes cognitive impairment and altered decision-making. *Neurochem Int.* 124, 106–113. doi: 10.1016/j.neuint.2018.12.019

Penny, W., Kiebel, S., and Friston, K. (2003). Variational Bayesian inference for fMRI time series. *Neuroimage* 19, 727–741. doi: 10.1016/S1053-8119(03)00071-5

Pfister, T., Xiaobai, L., Zhao, G., and Pietikainen, M. (2011). "Recognising spontaneous facial micro-expressions," in 2011 International Conference on Computer Vision (Barcelona), 1449–1456.

Sato, A., Masui, T., Yogo, A., Ito, T., Hirakawa, K., Kanawaku, Y., et al. (2020). Time-frequency analysis of serum with proton nuclear magnetic resonance for diagnosis of pancreatic cancer. *Sci. Rep.* 10:21941. doi: 10.1038/s41598-020-79087-3

Subbarao, M. V., and Samundiswary, P. (2016). “Time-frequency analysis of non-stationary signals using frequency slice wavelet transform," in 2016 10th International Conference on Intelligent Systems and Control (ISCO) (Coiimbatore), 1–6.

Copyright © 2021 Lai, Huang, Xin, Yu, Wen, Huang, Zhang, Shen and Tang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.