Check for updates

OPEN ACCESS

EDITED BY Jingying Chen, Central China Normal University, China

REVIEWED BY Hongliang Bi, Northwestern Polytechnical University, China Youhua Xia, Guangming Laboratory, China

*CORRESPONDENCE Hengjin Ke hengjin.ke@whu.edu.cn

†These authors have contributed equally to this work and share first authorship

RECEIVED 03 April 2023 ACCEPTED 29 May 2023 PUBLISHED 22 June 2023

CITATION

Chen B, Yin B and Ke H (2023) Interpretation of deep non-linear factorization for autism. *Front. Psychiatry* 14:1199113. doi: [10.3389/fpsyt.2023.1199113](https://doi.org/10.3389/fpsyt.2023.1199113)

COPYRIGHT

© 2023 Chen, Yin and Ke. This is an open-access article distributed under the terms of the [Creative Commons Attribution License](http://creativecommons.org/licenses/by/4.0/) [\(CC BY\).](http://creativecommons.org/licenses/by/4.0/) 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.

[Interpretation of deep non-linear](https://www.frontiersin.org/articles/10.3389/fpsyt.2023.1199113/full) factorization for autism

Boran Chen^{1†}, Bo Yin^{1†} and Hengjin Ke^{1,2*}

¹Computer School (Huangshi Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence), Hubei Polytechnic University, Huangshi, China, ²Computer School, Wuhan University, Wuhan, China

Autism, a neurodevelopmental disorder, presents significant challenges for diagnosis and classification. Despite the widespread use of neural networks in autism classification, the interpretability of their models remains a crucial issue. This study aims to address this concern by investigating the interpretability of neural networks in autism classification using the deep symbolic regression and brain network interpretative methods. Specifically, we analyze publicly available autism fMRI data using our previously developed Deep Factor Learning model on a Hibert Basis tensor (HB-DFL) method and extend the interpretative Deep Symbolic Regression method to identify dynamic features from factor matrices, construct brain networks from generated reference tensors, and facilitate the accurate diagnosis of abnormal brain network activity in autism patients by clinicians. Our experimental results show that our interpretative method effectively enhances the interpretability of neural networks and identifies crucial features for autism classification.

KEYWORDS

interpretation, autism, fMRI, deep symbolic regression, brain network, factorization

1. Introduction

In the rapidly evolving landscape of deep learning, remarkable strides have been made in recent years, evincing exceptional outcomes across an expansive array of applications, such as medical $(1-3)$, vehicular technology (4) , image (5) , and videos (6) . This remarkable feat has been accomplished through the innovative merging of more layers in neural networks, thereby generating complex non-linearities, which have yielded unparalleled results in perception tasks, thereby spurring the interest of researchers across the globe $(1, 7)$ $(1, 7)$. Within the realm of autism spectrum disorder (ASD), the prospect of early detection using deep learning technology presents an immensely auspicious prerequisite for timely intervention and treatment. The far-reaching implications of this breakthrough are staggering and hold the key to revolutionizing the diagnosis and management of autism, potentially unlocking a new era of insights into this complex neurological condition. The ramifications of this cutting-edge technology have the potential to ripple across the entire medical field, yielding transformative applications in other areas of medicine and beyond.

Despite the impressive achievements of deep learning techniques in various applications, the underlying rationales and conclusions remain shrouded in mystery, necessitating the need for greater interpretability from a more detailed and concrete perspective [\(8\)](#page-6-4). The interpretability of neural networks can provide valuable insights into refining the neural network design, drawing more meaningful conclusions, and deepening trust in the neural network. Within the realm of autism classification, the credibility of the neural network hinges upon its ability to accurately identify the essential characteristics of autism, leading to precise classifications [\(9\)](#page-6-5). Failure to analyze essential characteristics may lead to decisions based on external factors, noise, or interference, which cannot meet the high standards of medicine due to excessive false positives. Therefore, it is imperative to develop

interpretability for neural networks to shed light on the black box of neural networks, thereby enhancing their credibility, reliability, and applicability in various domains [\(10\)](#page-6-6).

At present, the interpretability of functional magnetic resonance imaging (fMRI) data primarily relies on the elucidation of features and classification models. In this regard, the literature has generated noteworthy contributions in both domains, which are summarized as follows. For example, in the early stage of model interpretability upon fMRI data, interpretation of weight vectors of linear models in multivariate neuroimaging was proposed to determine the origin of cognitive functions and associated neural processes [\(11\)](#page-6-7). From then on, the saliency features activated by the classifier were utilized to interpret reliable biomarkers associated with identifying ASD [\(12\)](#page-6-8). Subsequently, interpretability could aid in detecting patterns in fMRI data that indicate the presence of autism [\(13\)](#page-6-9), leading to more accurate diagnoses and personalized treatment plans. Recently, the integrated Gradients (IG) and Deep LIFT techniques were utilized to identify the correlations between brain regions that contribute most to the classification task [\(14\)](#page-6-10). Nowadays, a hybrid deep learning framework was proposed to improve classification accuracy and interpretability simultaneously [\(15\)](#page-6-11). By unlocking the power of interpretability in neural networks, we can harness the full potential of these technologies and ensure that they are not only effective but also trustworthy and ethically sound.

However, there exists a lack of research when it comes to ascertaining the interpretability of deep neural networks for fMRI data in autism. The present state of research is marked by a qualitative interpretation, single attention, and limited scope, particularly as it pertains to static aspects. However, it must be acknowledged that there is a palpable absence of interpretability of model dynamics in theory. The intrinsic dynamics of data and the interpretability of models are intimately linked. Comprehending the inherent laws and structure of data can aid in the creation of more precise and dependable analytical models. Concurrently, to truly grasp the driving mechanism behind the data, the model itself must attain a certain level of interpretability. As such, characterizing the intrinsic dynamics of data is a vital objective in data analysis and modeling. To accomplish this goal, various methods have been proposed, including deep learning, symbolic regression, and deep symbolic regression. Deep learning, a technique based on artificial neural networks, can learn the intrinsic structure and patterns of data through multiple layers of nonlinear transformations [\(16\)](#page-6-12) and has achieved remarkable success in fields such as image recognition, speech recognition, and natural language processing. Symbolic regression, a technique based on evolutionary algorithms, can learn interpretable mathematical expressions from data by searching the possible function space to find the function expression that best fits the data [\(17\)](#page-6-13). Deep symbolic regression amalgamates the strengths of deep learning and symbolic regression to learn interpretable mathematical expressions from data and unveil the dynamic laws behind it. By integrating deep learning and symbolic regression, it can effectively circumvent overfitting and underfitting issues while enhancing the interpretability of the model [\(18\)](#page-6-14).

To this end, as a result of our antecedent endeavors, we have managed to successfully extract the intrinsic non-linear factors inherent within fMRI via fusing the deep learning and

tensor factorization, namely Deep Factor Learning model on a Hibert Basis tensor (HB-DFL). Moreover, our team has endeavored to construct a classifier that can attain the paramount goal of accomplishing accurate discrimination of Attention Deficit Hyperactivity Disorder (ADHD) [\(19\)](#page-6-15) and Parkinson's [\(20\)](#page-6-16). This study will extend our previous works on quantifying the correlation between attention features and models, elucidating the dynamic interpretation of models to ensure precise discrimination of autism.

2. Methodology

The workflow of the entire solution is illustrated in [Figure 1.](#page-2-0) It is a complex and convoluted process that involves multiple steps of preprocessing, including skull stripping and correction, to ensure the accuracy of the results. The fMRI frames are then sliced and fed into HB-DFL [\(20\)](#page-6-16), a multi-branch convolutional neural network, where they are mapped to a low-rank space in various directions, generating non-linear factors. To generate a reference tensor equivalent to the initial tensor being decomposed, a predefined "constant" kernel tensor is used to perform tensor product operations with the non-linear factors. The difference between the reference tensor and the initial tensor being decomposed serves as the driving force for HB-DFL backpropagation. HB-DFL is then trained using the mean square error loss function until the backpropagation algorithm converges. The desired non-linear factors are then outputted, providing a powerful tool for analyzing fMRI data [\(19,](#page-6-15) [20\)](#page-6-16).

This section aims to shed light on the intricate dynamics and relationships between the extracted factors involved in Section 2.1, while providing a comprehensive interpretation of the model in terms of the brain network (Section 2.2).

2.1. Deep symbolic regression

To better understand the intricate inner workings of complex data, a approach has been extended from deep symbolic regression [\(18\)](#page-6-14). It involves combining deep learning with symbolic regression techniques, which can yield several benefits. This method capitalizes on the advanced high-dimensional and nonlinear processing abilities of deep learning while simultaneously leveraging the compactness, interpretability, and generalization properties of symbolic models. Ultimately, this integrated approach enables efficient pattern discovery in high-dimensional spaces in an end-to-end manner. We achieve the extension of deep symbolic regression by adding more operators, such as differential operators and divide operator. The architecture is illustrated in [Figure 2.](#page-2-1)

Firstly, the factor matrix learned by HB-DFL must be reorganize into a triplet format consisting of (i, j, value) tuples. The row and column indices, represented by i and j, respectively, specify the position of each element within the matrix, while the value component holds the actual value of the element. Once the matrix has been transformed into this format, it can be input into a deep symbolic regression network.

In addition, each layer of the network contains not only regular neurons (g = W \bullet x) but also operators f: ID, +, -, *, /, (\bullet)², e^x ,

 $sin(\bullet)$, $cos(\bullet)$, relu, and differential operators: y' etc.

$$
f(\mathbf{g}) = \begin{bmatrix} f_1(g_1) \\ f_2(g_2) \\ \vdots \\ f_{n_h}(g_{n_g-1}, g_{n_g}) \end{bmatrix}
$$
 (1)

Then, design the optimization function [\(18\)](#page-6-14):

$$
\mathcal{L} = |y_i - \hat{y}_i|_F^2 + \lambda L_{\frac{1}{2}}^* \tag{2}
$$

where $|\bullet|_F^2$ denotes the Frobenius Norm. The y_i and \hat{y}_i represent the ground truth and predicted label, respectively. $L_{\frac{1}{2}}^{*}$ is the regularization term:

$$
L_{\frac{1}{2}}^{*}(w) = \begin{cases} |w|^{1/2} & |w| \ge a \\ \left| -\frac{w^{4}}{8a^{3}} + \frac{3w^{2}}{4a} + \frac{3a}{8} \right|^{1/2} & |w| < a \end{cases}
$$
(3)

where a is a predefined constant threshold with the same setting following the strategies of (18) $(a = 0.01)$.

Deep Symbolic Regression networks share the same training technique with other deep neural networks, relying on the familiar backpropagation algorithm. Despite this, these networks exhibit a distinguishing trait they build a regression framework that correlates elements of the factor feature matrix with their corresponding subscripts. In other words, given any input row coordinate i and column coordinate j , the network's output forecasts the element's value \hat{y}_i , and the residual error between the predicted and actual value y_i determines the update of network parameters.

Finally, based on the combination and arrangement of operators, as well as a predetermined coefficient threshold (>0.01) , the algorithm converges and outputs a mathematical formula in a forward direction according to the network structure.

2.2. Brain network

The present study employs a multi-step approach to investigate brain connectivity in disease states. Firstly, temporal information is captured by splitting the tensor into 3D fMRI frames, and factor matrices are obtained using the HB-DFL method previously developed by the authors [\(20\)](#page-6-16). Subsequently, fMRI frames are reconstructed through a tensor product of factor matrices and Hilbert basis tensors, which are then merged to form a 4D reference tensor. The regions of interest (ROIs) are defined using the automated anatomical labeling (AAL) method, and subsequently parsed into eight distinct subnetworks based on underlying anatomical structures, namely the Default mode network (DMN), Auditory network (AN), Visual network (VN), Soma movement network (SMN), Bilateral edge network (BiN), Subcortical network (SCN), Cognitive control network (CCN), and Cerebellum network. To quantify the connectivity between all ROIs, an entropy-based partitioning approach is employed [\(21\)](#page-6-17), whereby the mutual information between two brain regions R1 and R2 is derived from the entropy $H(R1)$ and $H(R2)$, and joint entropy H(R1, R2). Based on a threshold value, a connectivity matrix of size 116×116 is constructed, and ANCOVA analysis is used to compare the connectivity matrices of control and disease groups at a significance level of p-value of 0.05 and false discovery rate (FDR) correction. The proposed approach has demonstrated the potential to provide a more comprehensive understanding of brain connectivity in disease states.

3. Results

The experiments in this section serve as a validation and assessment of the interpretability of the proposed model. We first introduce the dataset and experimental platform utilized in the experiments (see Section 3.1). Next, the dynamics of the factor matrix learned by HB-DFL are analyzed through deep symbolic regression (see Section 3.2). Finally, the effectiveness of HB-DFL is evaluated by analyzing the functional connectivity of fMRI data reconstructed using the reference tensor generated by HB-DFL (see Section 3.3). In the following section, we will provide a detailed description of the autistic fMRI data required for these experiments. The experimental procedures were conducted on a single desktop system comprising an Intel i7 CPU clocked at 3.33GHz, an Nvidia RTX 2080Ti GPU, and 64GB RAM running on a 64-bit Windows 7 operating system. This system served as the primary computational platform for all experimental activities, ensuring consistent and standardized testing conditions.

3.1. Dataset

The current investigation validates and assesses the proposed approach by utilizing the ASD fMRI dataset publicly available from the Kennedy Krieger Institute (KKI) site of the ABIDE II dataset.^{[1](#page-3-0)} DPABI [\(22\)](#page-6-18) software is utilized to preprocess both the original fMRI data and the fMRI data produced by HB-DFL. To be more specific, for all fMRI data, fMRIPrep generates brain voxels following skull stripping while disregarding magnetic field inhomogeneity correction. Later on, bbregister (FreeSurfer) is employed to align T1 images to T1W reference images. Before applying spatiotemporal filters, FSL 5.0.9 is utilized to evaluate head motion parameters, and ICA-AROMA is applied to remove motion or noise components. Finally, the images are resampled into two spaces, including the fsaverage5 space and the standard MNI152NLin2009cAsym space.

3.2. Dynamic interpretation of factor matrices

This study employs HB-DFL to perform multidimensional tensor factorization on fMRI frames, aiming to investigate the underlying mechanisms of the brain. The resulting factor matrices are randomized and divided into training (80%) and testing (20%) sets. The Deep Symbolic Regression method is used to learn the formula of the factor matrices in the training set, and the performance is evaluated based on the testing set. [Table 1](#page-4-0) presents the mathematical formulas and verification accuracy of different factor matrices learned from various groups. The study's findings show that most of the mathematical formulas derived from the data are straightforward and have an efficacy rate exceeding 90%, providing evidence of the effectiveness of the proposed method. Overall, this approach using HB-DFL allows for the dynamic interpretation of autism fMRI data with multidimensional attributes and enables the investigation of the underlying mechanisms of the brain, contributing to the advancement of research in this area. The results presented in this study underscore the essentiality of this approach for extracting meaningful information from fMRI data.

¹ http://fcon_1000.projects.nitrc.org/indi/abide/abide_II.html

Mode	Group	Equation	Accuracy (%)'
Coronal plane	Autism	$0.14\cos(\sin(\log(-0.165\cos(\frac{X_1}{0.05}))) - 0.11X_2)$	91.3
	Normal	$\sqrt{\left \left(\frac{relu(sin(0.624 - \frac{1}{X_2}))}{0.568} \right) + X_1 - 5.29 \right }$	90.61
Sagittal plane	Autism	$cos(\sqrt{ 27.709 + relu(sin(X_1 + \frac{1}{cos(\sqrt{ 4.71X_2 })})))$	93.41
	Normal	$cos(e^{0.017log(X_1X_2-4.2X_2)})$	92.15
Axis plane	Autism	0.431y ² -0.055 cos($\frac{1}{X_1e^{cos(\frac{1}{X_1})}} + X_2$)	94.25
	Normal	$sin(cos(17.65\sqrt{ log X_1 }))sin(X_2))$	91.74

TABLE 1 Mathematical formulas and verification accuracy of different factor matrices learned from different groups (Autism group and Normal control group) in KKI fMRI datasets, where y' represents the first-order differential.

The original tensor $\underline{\mathbf{X}} = \underline{\mathcal{G}} \times_1 f(\mathbf{A}) \times_2 f(\mathbf{S}) \times_3 f(\mathbf{C})$, which is based on the HB-DFL multidimensional tensor factorization theory [\(20\)](#page-6-16), can be understood as a composite function comprising factor matrices from each dimension and the core tensor G . By obtaining the dynamics of the factor matrices $f(A)$, $f(S)$, and $f(C)$ when G is known, the composite function can be utilized for the dynamic interpretation of Autism fMRI data with multidimensional attributes. The direct extraction of mathematical formulas from fMRI frames yields a maximum verification accuracy of only 70% with highly complex formulas, highlighting the essentiality of factor decomposition via the HB-DFL utilized in this study.

3.3. Brain network of reference tensor

The utilization of HB-DFL methodology in this study allowed for the analysis of brain functional connectivity networks using a substantial number of reference tensors reconstructed from the KKI dataset. The resulting brain network demonstrated significant functional connectivity and was further analyzed. [Figure 3A](#page-4-1) presents the functional connectivity of subnetworks and the differential connectivity of corresponding brain regions, which revealed noteworthy findings. Additionally, connectivity matrices of subnetworks [\(Figure 3B\)](#page-4-1) and cortical brain regions [\(Figure 3C\)](#page-4-1) provided insight into significant statistical differences in functional

connectivity between different groups of regions of interest (ROIs). These results provide valuable information on the functional connectivity of the brain, and contribute to the advancement of research in this area.

The findings of the present study replicate those of three prior investigations. Insufficient activation was observed in the DMN [\(23\)](#page-6-19) and VN [\(24\)](#page-6-20) of individuals with ASD, in terms of functional connectivity within the network. Enhanced functional connectivity was observed in individuals with ASD between the DMN and VN [\(25\)](#page-6-21), DMN and AN [\(26\)](#page-6-22), and DMN and SMN [\(27\)](#page-6-23), in terms of functional connectivity between networks. Additionally, enhanced functionality was observed between the angular gyrus and insula [\(28\)](#page-6-24) at the functional brain region level. These results illustrate the effectiveness of HB-DFL in compressive sensing, thereby maximizing information extraction.

4. Discussions and conclusions

Autism, a neurodevelopmental disorder characterized by social interaction and communication difficulties, as well as repetitive and stereotyped behaviors, is often studied using fMRI to investigate brain activity patterns. However, the complexity of fMRI data presents significant challenges, such as high dimensionality, noise, and redundancy, which make data interpretation and analysis difficult. To address these challenges, this paper proposes a model based on HB-DFL and deep symbolic regression techniques, which reduces the dimensionality of fMRI data and explains its dynamics. By representing high-dimensional data as a combination of lowdimensional latent factors, the HB-DFL can provide interpretable and understandable factors, each corresponding to a specific pattern of activity in a group of neurons, which can help us understand changes in brain activity in individuals with autism.

The factor matrix, the core component of the HB-DFL, contains the low-dimensional representation of brain activity for each participant and can be interpreted as specific brain networks or functional patterns. The mathematical formula of the factor matrix can thus help us identify and explain changes and abnormalities in brain activity in individuals with autism. However, as a case study, the validation of a single dataset is insufficient to demonstrate the generalizability of the method. To address this, the proposed method will be applied to a wider range of datasets, and crossvalidation between different datasets will be performed in the future to demonstrate its robustness.

In conclusion, the proposed model based on deep non-linear decomposition provides a valuable approach to understanding patterns of brain activity in individuals with autism. The interpretability of the model through the mathematical formula of the factor matrix can help us better understand the pathophysiology of autism, and may contribute to the development of future treatment and intervention strategies.

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 author.

Author contributions

HK contributed to the conception of the study, the reagents, materials, and analysis tools, conceived and designed the experiments, and analyzed the data. BY and BC performed the experiments. All authors contributed to the article and approved the submitted version.

Funding

This work was partially supported by the Hubei Provincial Natural Science Foundation Project under Grant 2022CFB524, the Teaching and Research Reform Project of Hubei Provincial Department of Education (2022438), the Key Project of the Scientific Research Program of Hubei Provincial Department of Education (D20214503), the Talent Introduction Project of Hubei Polytechnic University (21xjz16R), the School Major Teaching and Research Project under Grant 2022A05, the Scientific Research Funding Project for Young Teachers of Hubei Normal University (HS2020QN038, KCSZY202148, and 2021035), and the National Natural Science Foundation of China under Grant 62273136.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

electrocardiograms using a deep neural network. Nat Med. (2019) 25:65–9. doi: [10.1038/s41591-018-0268-3](https://doi.org/10.1038/s41591-018-0268-3)

3. Ke H, Chen D, Shi B, Zhang J, Liu X, Zhang X, et al. Improving brain e-health services via high-performance EEG classification with grouping Bayesian optimization. IEEE Trans. Serv. Comput. (2020) 13:696–708. doi: [10.1109/TSC.2019.2962673](https://doi.org/10.1109/TSC.2019.2962673)

^{1.} Bi H, Liu J, Kato N. Deep learning-based privacy preservation and data analytics for IoT enabled healthcare. IEEE Trans. Indus. Inform. (2022) 18:4798–807. doi: [10.1109/TII.2021.3117285](https://doi.org/10.1109/TII.2021.3117285)

^{2.} Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, et al. Cardiologist-level arrhythmia detection and classification in ambulatory

4. Xia Y, Wu L, Wang Z, Zheng X, Jin J. Cluster-enabled cooperative scheduling based on reinforcement learning for high-mobility vehicular networks. IEEE Trans. Vehicular Technol. (2020) 69:12664–78. doi: [10.1109/TVT.2020.3029561](https://doi.org/10.1109/TVT.2020.3029561)

5. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV: IEEE (2016). p. 770–8.

6. Akagündüz E, Aslan M, Sengur A, Wang H, Ince MC. Silhouette orientation volumes for efficient fall detection in depth videos. IEEE J Biomed Health Inform. (2017) 21:756–63. doi: [10.1109/JBHI.2016.2570300](https://doi.org/10.1109/JBHI.2016.2570300)

7. Bai Y, Chen W, Chen J, Guo W. Deep learning methods for solving linear inverse problems: research directions and paradigms. Signal Process. (2020) 177:107729. doi: [10.1016/j.sigpro.2020.107729](https://doi.org/10.1016/j.sigpro.2020.107729)

8. Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T, editors. Computer Vision – ECCV 2014. Cham: Springer International Publishing (2014). p. 818–833.

9. Zhang F, Wei Y, Liu J, Wang Y, Xi W, Pan Y. Identification of Autism spectrum disorder based on a novel feature selection method and variational autoencoder. Comput Biol Med. (2022) 148:105854. doi: [10.1016/j.compbiomed.2022.105854](https://doi.org/10.1016/j.compbiomed.2022.105854)

10. Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?": explaining the predictions of any classifier," in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16 Association for Computing Machinery (2016). p. 1135–1144.

11. Haufe S, Meinecke F, Gorgen K, Dahne S, Haynes JD, Blankertz B, et al. On the interpretation of weight vectors of linear models in multivariate neuroimaging. Neuroimage. (2014) 87:96–110. doi: [10.1016/j.neuroimage.2013.10.067](https://doi.org/10.1016/j.neuroimage.2013.10.067)

12. Li X, Dvornek NC, Zhuang J, Ventola P, Duncan JS. Brain biomarker interpretation in ASD using deep learning and fMRI. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, editors. Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. Cham: Springer International Publishing (2018). p. 206–14.

13. Hu J, Cao L, Li T, Liao B, Dong S, Li P. Interpretable learning approaches in resting-state functional connectivity analysis: the case of autism spectrum disorder. Comput Math Methods Med[. \(2020\) 2020:1394830. doi: 10.1155/2020/13](https://doi.org/10.1155/2020/1394830) 94830

14. Prasad PKC, Khare Y, Dadi K, Vinod PK, Surampudi BR. Deep learning approach for classification and interpretation of autism spectrum disorder. In: 2022 International Joint Conference on Neural Networks (IJCNN). Padua (2022). p. 1–8.

15. Zhao M, Yan W, Luo N, Zhi D, Fu Z, Du Y, et al. An attention-based hybrid deep learning framework integrating brain connectivity and activity of resting-state functional MRI data. Med Image Anal. [\(2022\) 78:102413. doi: 10.1016/j.media.2022.](https://doi.org/10.1016/j.media.2022.102413) 102413

16. Lusch B, Kutz JN, Brunton SL. Deep learning for universal linear embeddings of nonlinear dynamics. Nat Commun. (2018) 9:4950. doi: [10.1038/s41467-018-07210-0](https://doi.org/10.1038/s41467-018-07210-0)

17. Kubalík J, Derner E, Babuška R. Multi-objective symbolic regression for physics-aware dynamic modeling. Expert Syst Appl. (2021) 182:115210. doi: [10.1016/j.eswa.2021.115210](https://doi.org/10.1016/j.eswa.2021.115210)

18. Kim S, Lu PY, Mukherjee S, Gilbert M, Jing L, Čeperić V, et al. Integration of neural network-based symbolic regression in deep learning for scientific discovery. IEEE Trans Neural Netw Learn Syst. (2021) 32:4166–77. doi: [10.1109/TNNLS.2020.3017010](https://doi.org/10.1109/TNNLS.2020.3017010)

19. Ke H, Wang F, Ma H, He Z. ADHD identification and its interpretation of functional connectivity using deep self-attention factorization. Knowledge Based Syst. (2022) 250:109082. doi: [10.1016/j.knosys.2022.109082](https://doi.org/10.1016/j.knosys.2022.109082)

20. Ke H, Chen D, Yao Q, Tang Y, Wu J, Monaghan J, et al. Deep factor learning for accurate brain neuroimaging data analysis on discrimination for structural MRI and functional MRI. IEEE/ACM Trans Comput Biol Bioinform. (2023) 1–14. doi: [10.1109/TCBB.2023.3252577](https://doi.org/10.1109/TCBB.2023.3252577)

21. Wang F, Ke H. Global epileptic seizure identification with affinity propagation clustering partition mutual information using cross-layer fully connected neural network. Front Hum Neurosci. (2018) 12:396. doi: [10.3389/fnhum.2018.00396](https://doi.org/10.3389/fnhum.2018.00396)

22. Yan CG, Wang XD, Zuo XN, Zang YF. DPABI: data processing & analysis for (resting-state) brain imaging. Neuroinformatics. (2016) 14:339–51. doi: [10.1007/s12021-016-9299-4](https://doi.org/10.1007/s12021-016-9299-4)

23. Yan CG, Chen X, Li L, Castellanos FX, Bai TJ, Bo QJ, et al. Reduced default mode network functional connectivity in patients with recurrent major depressive disorder. Proc Natl Acad Sci USA. (2019) 116:9078–83. doi: [10.1073/pnas.1900390116](https://doi.org/10.1073/pnas.1900390116)

24. Lombardo MV, Eyler L, Moore A, Datko M, Carter Barnes C, Cha D, et al. Default mode-visual network hypoconnectivity in an autism subtype with pronounced social visual engagement difficulties. eLife. (2019) 8:e47427. doi: [10.7554/eLife.47427](https://doi.org/10.7554/eLife.47427)

25. Duan X, Chen H, He C, Long Z, Guo X, Zhou Y, et al. Resting-state functional under-connectivity within and between large-scale cortical networks across three low-frequency bands in adolescents with autism. Prog Neuropsychopharmacol Biol Psychiatry. (2017) 79:434–41. doi: [10.1016/j.pnpbp.2017.07.027](https://doi.org/10.1016/j.pnpbp.2017.07.027)

26. Takeda Y, Itahashi T, aki Sato M, Yamashita O. Estimating repetitive spatiotemporal patterns from many subjects' resting-state fMRIs. Neuroimage. (2019) 203:116182. doi: [10.1016/j.neuroimage.2019.116182](https://doi.org/10.1016/j.neuroimage.2019.116182)

27. Nunes AS, Peatfield N, Vakorin V, Doesburg SM. Idiosyncratic organization of cortical networks in autism spectrum disorder. Neuroimage. (2019) 190:182–90. doi: [10.1016/j.neuroimage.2018.01.02](https://doi.org/10.1016/j.neuroimage.2018.01.02)

28. Salmi J, Roine U, Glerean E, Lahnakoski J, Nieminen-von Wendt T, Tani P, et al. The brains of high functioning autistic individuals do not synchronize with those of others. Neuroimage Clin. (2013) 3:489–97. doi: [10.1016/j.nicl.2013.10.011](https://doi.org/10.1016/j.nicl.2013.10.011)