Journal of Imaging Informatics in Medicine (2024) 37:1023–1037
https://doi.org/10.1007/s10278-024-01002-3

Identification of Autism Spectrum Disorder Using Topological Data Analysis

Xudong Zhang[1] · Yaru Gao[1] · Yunge Zhang[2] · Fengling Li[1] · Huanjie Li[2] · Fengchun Lei[1]

Received: 3 August 2023 / Revised: 30 October 2023 / Accepted: 21 November 2023 / Published online: 13 February 2024

Abstract

Autism spectrum disorder (ASD) is a pervasive brain development disease. Recently, the incidence rate of ASD has increased year by year and posed a great threat to the lives and families of individuals with ASD. Therefore, the study of ASD has become very important. A suitable feature representation that preserves the data intrinsic information and also reduces data complexity is very vital to the performance of established models. Topological data analysis (TDA) is an emerging and powerful mathematical tool for characterizing shapes and describing intrinsic information in complex data. In TDA, persistence barcodes or diagrams are usually regarded as visual representations of topological features of data. In this paper, the Regional Homogeneity (ReHo) data of subjects obtained from Autism Brain Imaging Data Exchange (ABIDE) database were used to extract features by using TDA. The average accuracy of cross validation on ABIDE I database was 95.6% that was higher than any other existing methods (the highest accuracy among existing methods was 93.59%). The average accuracy for sampling with the same resolutions with the ABIDE I on the ABIDE II database was 96.5% that was also higher than any other existing methods (the highest accuracy among existing methods was 75.17%).

Keywords Autism spectrum disorder · Machine learning · Topological data analysis · ABIDE · Persistent homology

Introduction

Autism is the representative disease of pervasive brain-based developmental disorder. The core symptom of typical autism is the so-called triad syndrome, which mainly reflects in the social communication skills, language ability and ritualized rigid behavior. Although autism has been around for thousands of years, it was not identified by Kanner and Asperger until 1943 and 1944, respectively [1, 2]. Autism spectrum disorder (ASD) is a broad definition of the core symptoms of typical autism. It includes both typical and atypical autism, as well as suspected autism, autism borderline, autistic tendencies and developmental delays. It has been widely considered as a neurological disorder caused by abnormalities that found in brain coordinated functioning regions [2].

Most individuals with autism begin to exhibit abnormal development before the age of 3, and are obvious before the age of 5. Generally speaking, the younger the age of onset, the more severe the symptoms. A recent statistic from the Centers for Disease Control and Prevention has showed that 1 in 68 children had autism in the United States [3]. The National Institutes of Health conservatively estimated the prevalence of autism at 5 to 6 per 1000 people in the United States. Overall, the rate of autism in males is 3 to 4 times higher than that in females, but the symptoms in females are more severe than those in males.

The causes of ASD have been an unsolved problem in the medicine field. It is generally believed that various developmental disorders exhibited by people with ASD are mainly caused by brain biology. Many researches have analyzed the causes of brain biological changes from the aspects of ecology, neuropsychology and medical biology [4–6].

Children with autism often have abnormal electroencephalogram, suggesting that children may have abnormal brain structure or function [7]. Using magnetic resonance imaging (MRI), the researchers found that the abnormal white and gray matter hyperplasia was the most obvious in the frontal lobe [7]. In addition, the volumes of the amygdala, left hippocampus and caudate nucleus were larger than those in the control group, but the cerebellar vermis and corpus callosum knee decreased significantly. Functional magnetic

✉ Fengling Li
fenglingli@dlut.edu.cn

[1] School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China

[2] School of Biomedical Engineering, Dalian University of Technology, Dalian 116024, China

resonance imaging (fMRI) has been used to detect the cerebral blood supply of autistic children in early stage. And the cerebral insufficiency of blood supply mainly appeared in the frontal lobe [8, 9], temporal lobe, cerebellum and thalamus. Nowadays, fMRI has become a very important and effective method for studying human brain function due to its non-invasive and high spatial and temporal resolution. fMRI is a new neuroimaging method that uses magnetic resonance to measure the hemodynamic changes caused by neuronal activity. Almost all fMRI detect reactive areas in the brain by Blood Oxygen Level Dependent (BOLD) contrast mechanism that first was proposed by Ogawa et al. in 1990 [10]. Rs-fMRI, which maps brain function by looking at brain signals during resting state, is widely used to study psychiatric disorders including ASD, bipolar disorder, schizophrenia and attention deficit hyperactivity disorder. It does not require complex task design and has good operability. It can avoid the incomparability of experimental results caused by different task designs and execution conditions of the subjects in task-based research.

It has pointed out that ASD subjects activate more relevant regions in the right hemisphere of the brain in language processing tasks [11]. Thus, it hypothesized that this opposite lateralization of language areas might be the reasons for the difficulties in language processing of this population [11]. It has also been reported that ASD subjects represent decreased frontal lobe activation along with increased left temporal lobe activation when the verbal stimuli was added [12].

A recent field in ASD research is to distinguish ASD subjects from typical controls based on resting-state fMRI (rs-fMRI) data combined with machine learning or deep learning techniques. It has proposed a novel element-wise layer for deep neural networks and obtained the highest accuracy of 68.7% [13]. Reiter et al. have showed the classification of ASD depends on sample heterogeneity. It has been shown to produce high accuracy for grouping ASD samples into more homogeneous subgroups based on gender and severity ranges. The highest accuracy in four validation groups was 73.75% [14]. The Spatial Feature based detection method has been proposed to extract connectivity features, and provided biologically interpretable results by highlighting the major differences in the BOLD signals between the typical subjects and ASD subjects [15]. This method achieved the classification accuracy of 77.3%. The graph convolutional networks have been proposed to extract features of control and ASD groups based on functional connectivity graph. They achieved the accuracy of 70% to distinguish healthy individuals and ASD subjects [16]. Almuqhim et al. developed a deep learning model called ASD-SAENet, Eslami et al. proposed a framework called ASD-DiagNetfor, and Heinsfeld et al. applied deep learning algorithms based on the patterns of brain activation for classifying subjects with ASD from typical controls [3, 17, 18]. Mastafa et al. designed ASD diagnostic features based on brain networks. They used 264 region based partitioning schemes to construct brain networks from fMRI, and applied feature selection algorithms to obtain 64 discriminant features. The classification accuracy of linear discriminant analysis reached 77.7% [19]. The classification model constructed based on the characteristics of transgenic macaque reached 82.14% in the distinction accuracy of between autistic individuals and healthy people [20]. Using brain dynamic networks and feature extraction, machine learning classifiers reached the accuracy of 88.8% when considering the temporal dynamics of data [21, 22]. A model based on SVM-RFE and stacked sparse auto-encoder has been proposed to identify ASD subjects and healthy controls. This model reached the average accuracy of 93.59% [23].

A large of data has been produced across various disciplines due to the significant advancements in experimental tools and techniques. A main objective of data analysis is to enable researchers to gain relevant insights into the data, including comprehending its overall organization and distribution patterns. Topological data analysis (TDA) refers to statistical methods which make use of topological methods and provide comprehension to the "shape" of data. These techniques can be used in understanding global features of data that are not easily obtained using other tools [24–26]. Topology and geometry are very natural tools to apply in this area, since geometry is regarded as the study of distance functions, and what we often work with are distance functions on large finite sets of data [25]. Topology is a study of special geometric properties of an object or space, which remains invariant after continuously changing its shape. Similar structural properties may exist inside complex data, and they are called the "shape" of data. Topology can describe the intrinsic information of data. The use of TDA is limited by the difficulty of combining the main tools of algebraic topology, machine learning algorithm, and the persistence diagrams or persistence barcodes with statistics [26]. Compared with common methods such as principal component analysis and cluster analysis, TDA can not only capture topological information of data effectively, but also be good at discovering some small categories that cannot be found by traditional methods.

In this paper, our main contributions were:

We established a new model based on persistent homology in TDA for classifying ASD individuals and healthy people on the ABIDE database. Persistent homology captured the topological structure of data from different dimensions, such as the number of connected components, loops and cavities. Our model achieved the state-of-the-art accuracy on both two subdatabases compared with other models.

The good performance of some evaluation indexes indicated that there were structural differences in ReHo between the ASD individuals and healthy people. The different dimensional and multi-scale (under different filtration values) connectivity differences of ReHo between the two groups of people were captured by persistent homology. ReHo data is effective for studying ASD.

The rest of this paper is organized as follows: In the "Materials and Methods" section, we introduce Autism Brain Imaging Data Exchange (ABIDE) database information, datasets preprocessing, related definitions about homology group and persistent homology and the method of feature extraction. In the "Experimental Results and Comparison with Existing Methods" section, we present our experimental results on two databases and compare our results with those of some other models used on the two datasets. The "Discussions" and "Conclusions" sections are devoted to discussions and conclusions, respectively.

Materials and Methods

ABIDE Database and Sampling of Subjects

The ABIDE database is a publicly shared database for ASD research, which contains two subdatabases, ABIDE I and ABIDE II. The ABIDE I is a collaboration of 17 international imaging sites that have combined and shared imaging data from 573 typical control subjects and 539 participants with ASD. There is the rs-fMRI imaging information of 1112 subjects and an extensive array of phenotypic information. All data is anonymized in accordance with HIPAA guidelines, with analyses performed in accordance with preapproved procedures by the University of Utah Institutional Review Board. All images have been obtained with informed consent according to procedures established by human subject research boards at each participating institution. Specific details are available https://fcon_1000.projects.nitrc.org/indi/abide/abide_I.html.

We downloaded the ABIDE I dataset by the github script https://github.com/preprocessed-connectomes-project/abide. There were three folders and seven files. The Phenotypic_V10b_preprocessed1.csv file recorded the relevant information of the data, such as disease, the data number medications and genders of all subjects. The Download_abide_prepro-c_guided.txt gave the specific downloading methods and alternative parameters. The download_abide_preproc.py script allowed any user to download outputs from the ABIDE preprocessed data release. We could open the command prompt window and then run the related commands written in the file download_abide_preproc_guide.txt. At a minimum, the script needed a specific derivative, pipeline and strategy to search. We chose them as Regional Homogeneity (ReHo) [27], Configurable Pipeline for the Analysis of Connectomes (C-PAC) and nofilt_noglobal. We also got the database by pytorch nilearn downloading. The choice of quality_checked affected the number of data, with qualit_checked=True downloading 884 data and qualit_checked=False downloading 1035 data. The ReHo data of 884 subjects from ABIDE I database could be obtained in the above two methods. Some information about the dataset is shown in Table 1.

The ABIDE II involves 19 international sites that donated 1114 rs-fMRI imaging data from 593 typical controls and 521 participants with ASD (age range: 5–64 years). This database has been openly released to the scientific community on June 2016. In accordance with HIPAA guidelines and 1000 Functional Connectomes Project/INDI protocols, all data was anonymous, with no protected health information included. An extensive array of phenotypic information about subjects of ABIDE II is shown in Table 2. We can download the ABIDE II data from https://fcon_1000.projects.nitrc.org/indi/abide/abide_II.html. The ReHo data of 795 subjects from ABIDE II database can also be obtained.

Data Preprocessing

In order to obtain reliable results, it is often necessary to preprocess the collected raw data. The functional preprocessing of neuroimaging data from ABIDE I has been performed by the Preprocessed Connectomes Project (https://preprocessedconnectomesproject.org/abide/index.html). We selected the derivative, pipeline and strategy parameter as ReHo, C-PAC and nofilt_noglobal mentioned in the "ABIDE Database and Sampling of Subjects" section. The data was motion corrected and slice time corrected. Nuisance signal removal was performed using 24 motion parameters, CompCor with 5 components [28], low-frequency drifts (linear and quadratic trends). The voxel intensity was normalized.

For ABIDE II database, we preprocessed the original data using FSL FEAT, including removing the first six volumes, motion correction and spatial normalization to standard MNI space. The ReHo could be generated from the preprocessed rs-fMRI data with DPABI [29]. And then, we performed spatial smoothing (with Full Width at Half Maximum (FWHM) of 6 mm) after calculating ReHo.

Topological Data Analysis

TDA is an analysis method that combines topology and data analysis, which is used to study the topological properties in big data. In recent years, persistent homology was an analytical approach in TDA which has been considered as a practical method to represent topological features of objects [30, 31]. For more details about the TDA, persistent homology and its representation, we can refer to [24, 25, 32].

Table 1 Phenotypic information of 884 subjects of ABIDE I (M: Male, F: Female)

	ASD		TC	
Site	Avg-Age (SD)	Sample Size	Avg-Age (SD)	Sample Size
Caltech	27.44 (10.03)	19 (M:15,F:4)	28.02 (10.58)	18 (M:14, F:4)
CMU	30.33 (6.94)	3 (M:3, F:0)	25.5 (4.5)	2 (M:1, F:1)
KKI	9.56 (1.34)	12 (M:9, F:3)	10.08 (1.10)	27 (M:20, F:7)
Leuven	17.98 (4.98)	27 (M:25, F:2)	18.21 (4.98)	34 (M:29, F:5)
MaxMun	30.44 (13.59)	18 (M:15, F:3)	25.92 (8.14)	24 (M:23, F:1)
NYU	14.92 (7.04)	73 (M:72,F:26)	15.67 (6.14)	98 (M:64, F:9)
OHSU	11.43 (2.09)	12 (M:12, F:0)	10.37 (1.05)	11 (M:11, F:0)
Olin	16.79 (3.63)	14 (M:11, F:3)	17.55 (3.03)	11 (M:9, F:2)
Pitt	19.35 (7.35)	22 (M:18, F:4)	19.13 (6.19)	23 (M:20, F:3)
SBL	35.29 (10.37)	14 (M:14, F:0)	34.42 (5.78)	12 (M:12, F:0)
SDSU	15.05 (1.60)	12 (M:12, F:0)	14.32 (1.84)	21 (M:15, F:6)
Stanford	10.15 (1.60)	17 (M:13, F:4)	9.89 (1.58)	19 (M:15, F:4)
Trinity	17.01 (3.04)	21 (M:21, F:0)	17.48 (3.58)	23 (M:23, F:0)
UCLA	13.34 (2.56)	36 (M:34, F:2)	13.18 (1.76)	39 (M:33, F:6)
UM	13.85 (2.29)	48 (M:39, F:9)	15.03 (3.64)	65 (M:49, F:16)
USM	24.60 (8.46)	38 (M:38, F:0)	22.33 (7.70)	23 (M:23, F:0)
Yale	13.01 (3.03)	22 (M:15, F:7)	12.76 (2.78)	26 (M:19, F:7)

Homology and Persistent Homology

In order to calculate the homology and persistent homology of the data, we introduce the related concept of complexes. Cubical complexes can provide a good topological representation for image data. A cubical complex is a topological space obtained by gluing some elementary cubes. For some $l \in \mathbb{Z}$, a closed interval $I = [l, l+1]$ or $I = [l, l]$ is an elementary interval, and the elementary interval $I = [l, l]$ is degenerate. For $i \in [1, 2, \cdots, d]$, I_i is an elementary interval, an elementary cube is a finite product of elementary intervals, that is, $Q = I_1 \times I_2 \times \cdots \times I_d \subseteq R^d$, for example, see Fig. 1. The dimension of Q is the number of its non-degenerate intervals. A boundary of an elementary cube $Q = I_1 \times I_2 \times \cdots \times I_d$ is a chain obtained in the following way:

$$\partial Q = (\partial I_1 \times I_2 \times \cdots \times I_d) + (I_1 \times \partial I_2 \times \cdots \times I_d) \\ + \cdots + (I_1 \times I_2 \times \cdots \times \partial I_d).$$

Homology groups characterize the type and the numbers of "holes" of the given topological space and provide a fundamental description about its structure. In order to calculate homology groups of complexes which are topological representations of objects, it needs a series of chain groups and linear maps between two adjacent dimensional chain groups which are called boundary maps. A k-chain of the cubical complex $\mathscr{C}$ is the sum of its some k-dimensional cubes over the field $\mathbb{Z}_2$. All k-chains of $\mathscr{C}$ form the k-th chain group of $\mathscr{C}$ denoted by $C_k(\mathscr{C})$, which is a free Abelian group. For $k \geq 1$, the kernel of the boundary map $\partial_k : C_k \rightarrow C_{k-1}$ is called the cycle group and denoted by $Z_k(\mathscr{C})$. The image of the boundary map $\partial_{k+1} : C_{k+1} \rightarrow C_k$ is called the boundary group and denoted by $B_k(\mathscr{C})$. It is easy to see $\partial_{k+1} \circ \partial_k = 0$ and $B_k \subseteq Z_k \subseteq C_k$, the k-th homology group of $\mathscr{C}$ can be calculated by $H_k(\mathscr{C}) = Z_k(\mathscr{C}) / B_k(\mathscr{C})$.

Persistent homology presents topological information from different scales by employing a family of nested complexes through a filtration process. The filtration requires that if a new cubical complex is added, each of its proper faces should be added before that. A nested subsequence of complexes $\emptyset = K_0 \subset K_1 \subset \cdots K_n = K$ is a filtration of a complex K, for all $i \geq n, K_i = K$. At filtration time l, the p-persistent k-th homology group can be written as $H_k^{l,p} = Z_k^l / (Z_k^l \cap B_k^{l+p})$. The p-persistent k-th Betti number $\beta_k^{l,p}$ of K^l is the rank of $H_k^{l,p}$. Thus, we can calculate and evaluate the intrinsic topological properties of spaces or objects.

Visualization of Persistent Homology

It is noticed that some topological features "live" longer in these complexes, whereas others "live" shorter or "die" quicker with filtration value changes. Their persistent time provides a relative geometric measurement of the associated topological properties [32–35]. We refer to the living longer features as topological invariants and the shorter ones as noise.

We can use pairs of birth time and death time to represent the results from persistent homology. Specifically, for every topological invariant, birth time and death time are the filtration values at which the generators are born and vanished,

Table 2 Phenotypic information of 795 subjects of ABIDE II (M: Male, F: Female)

	ASD		TC	
Site	Avg-Age (SD)	Sample Size	Avg-Age (SD)	Sample Size
BNI	37.45 (15.81)	29 (M:29,F:0)	40.32 (14.56)	28 (M:28, F:0)
EMC	8.74 (1.12)	14 (M:12, F:2)	8.02 (0.72)	13 (M:10, F:3)
ETH	21.57 (3.68)	7 (M:7, F:0)	23.94 (4.60)	22 (M:20, F:7)
GU	17.98 (4.98)	27 (M:25, F:2)	10.60 (1.72)	41 (M:22, F:0)
IU	25.28 (9.51)	18 (M:14, F:4)	24.00 (4.77)	19 (M:14, F:5)
IP	15.99 (5.18)	13 (M:8,F:5)	25.53 (11.18)	20 (M:7, F:13)
KUL	23.76 (5.00)	25 (M:25, F:0)	0 (0)	0 (M:0, F:0)
KKI	10.72 (1.51)	25 (M:16, F:9)	10.30 (1.19)	123 (M:73, F:50)
NYU	9.13 (5.26)	61 (M:55, F:6)	9.49 (3.38)	28 (M:26, F:2)
ONRC	21.60 (3.77)	15 (M:14, F:1)	24.19 (3.88)	26 (M:16, F:10)
OHSU	11.91 (2.21)	33 (M:27, F:6)	10.41 (1.66)	51 (M:25, F:26)
TCD	15.73 (3.58)	10 (M:10, F:0)	16.66 (2.64)	16 (M:16, F:0)
SDSU	13.23 (3.07)	30 (M:24, F:6)	13.25 (2.98)	25 (M:23, F:2)
SU	10.78 (1.18)	14 (M:13, F:1)	11.06 (1.05)	17 (M:15, F:2)
UCLA	12.02 (2.14)	12 (M:11, F:1)	10.07 (2.27)	12 (M:8, F:4)
USM	21.90 (7.32)	9 (M:7, F:2)	26.17 (6.74)	12 (M:10, F:2)

respectively. And persistent time of every invariant represents the length of the "lifespan" interval (death time minus birth time) [32]. For j-th generator of the k-th persistent homology, birth and death time always come in pairs and can be denoted as a_j^k and b_j^k. We use l_j^k to represent the persistence of j-th generator of the k-th persistent homology.

We usually use persistence barcodes or persistence diagram to visualize the topological persistence. Each l_j^k is considered as a bar in persistence barcodes and also treated as a 2-dimensional point with coordinate (a_j^k, b_j^k) in persistence diagram. The persistence diagram consists of these points (finite multiplicity) and all points in the main diagonal which are considered as infinite multiplicity in $\bar{\mathbb{R}}^2$ where $\bar{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$. The total multiplicity of points that are not on the diagonal is the size of the persistence diagram. If there are multiple bars with the same endpoint and start point simultaneously, the point is counted with multiplicity in the persistence diagram. For example, persistence barcodes and its corresponding persistence diagram generated by constructing a filtration are shown in Fig. 2.

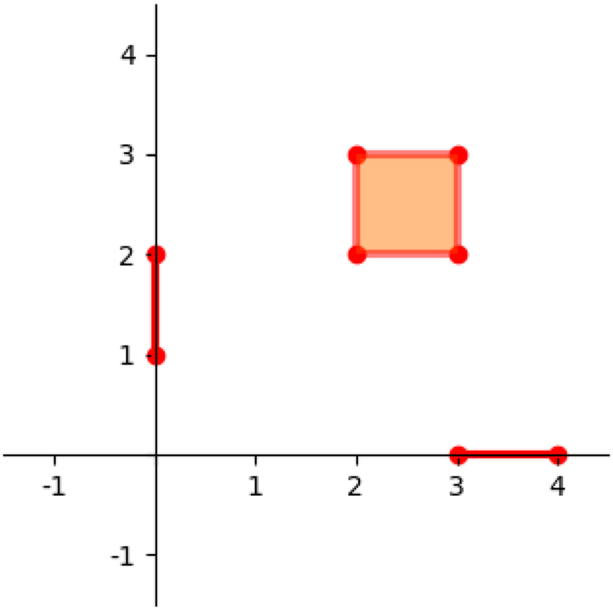


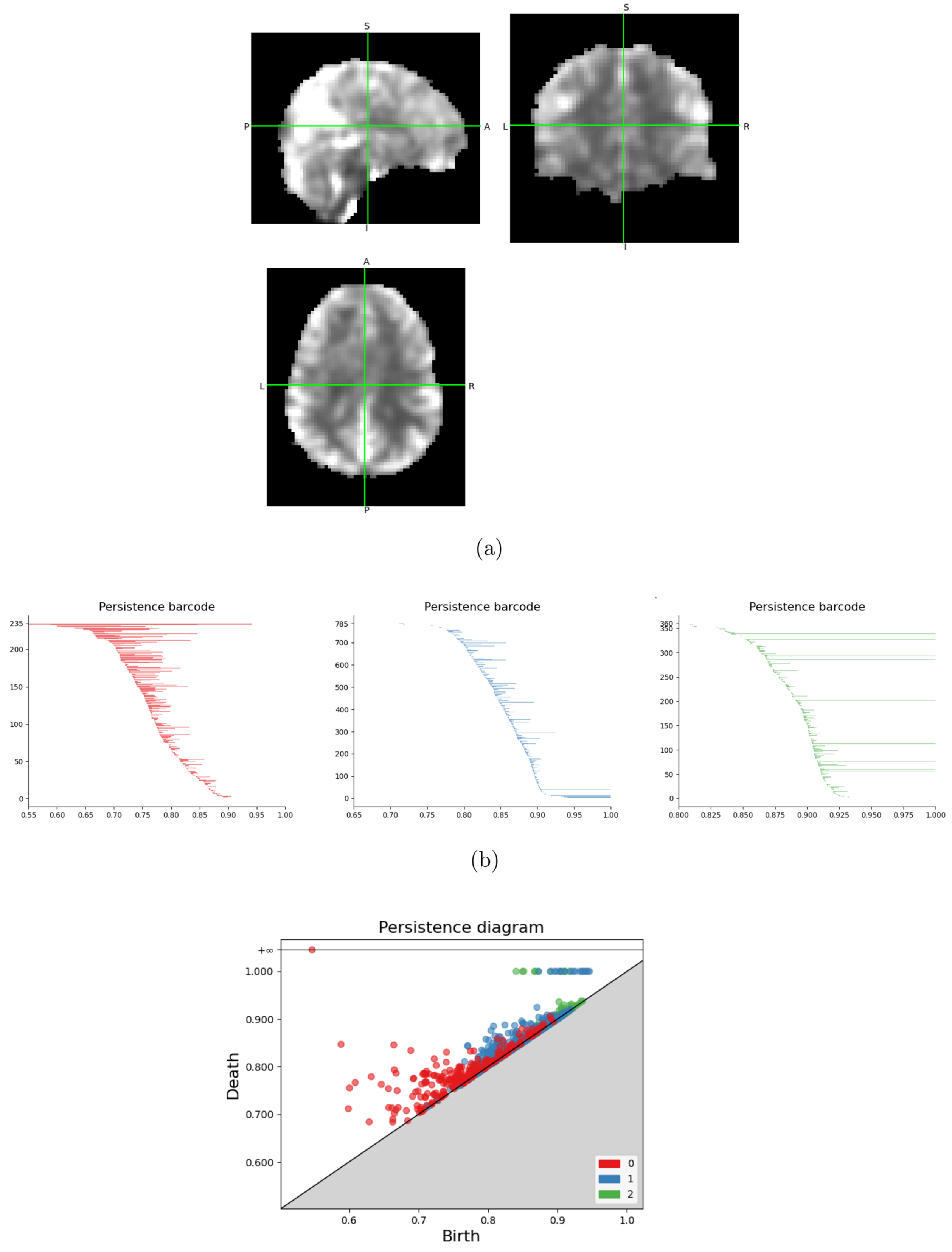
Fig. 1 The 1-dimensional elementary cubes $[3,4] \times [0,0]$ and $[0,0] \times [1,2] \subset \mathbb{R}^2$, 2-dimensional elementary cube $[2,3] \times [2,3] \subset \mathbb{R}^2$

Persistent homology provides a very important and promising way of structure representation, and is applied to various fields. Recently, the machine learning models based on persistent homology have been used in various research fields, including noise data [36], shape analysis [37], computational biology [38], image analysis [39] and drug design [40]. There are various persistent homology softwares, for example, GUDHI [41], JavaPlex [42], R-TDA package [43], Ripser [44], PHAT [45] and CubicalRipser [46].

Extracting Features

One of the reliable and frequently used fMRI indexes for measuring local connectivity is ReHo. The ReHo index measures the consistency of time series between voxels and their neighboring voxels, and is designed to represent local synchronization of spontaneous neural activity on a centimeter scale. The size of the scale depends on the size of voxels and the number of neighboring voxels included in the calculation.

In this paper, we used persistent homology to extract the topological information of ReHo data of each subject. The sampling resolution sizes of the three dimensions for each subject selected on the ABIDE I and ABIDE II databases were 61, 73 and 61, respectively. The ReHo value calculated using C-PAC represents the correlation (Kendall coefficient)


S
P
A
I
S
L
R
I
A
L
R
P
Persistence barcode
235
200
150
100
50
0
0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90 0.95 1.00
Persistence barcode
785
700
600
500
400
300
200
100
0
0.65 0.70 0.75 0.80 0.85 0.90 0.95 1.00
Persistence barcode
360
350
300
250
200
150
100
50
0
0.800 0.825 0.850 0.875 0.900 0.925 0.950 0.975 1.000
Persistence diagram
+∞
1.000
0.900
0.800
0.700
0.600
Death
0.6 0.7 0.8 0.9 1.0
Birth
0
1
2

(a)

(b)

(c)

◀**Fig. 2** The illustration of the persistence barcodes and persistence diagram generated by an fMRI ReHo data under a suitable filtration. **a** An fMRI ReHo data of a subject. **b** The 0-, 1- and 2-dimensional persistence barcodes generated by the ReHo data in **a**. **c** The corresponding persistence diagram

of time series between a given voxel and the surrounding 26 voxels [27]. Thus, we obtained the 3-dimensional ReHo matrix for each subject. From the calculation method of ReHo, it describes the functional connection relationship between a given node and its neighbor nodes, thereby quantifying the degree of connection between the node and its neighbors in the brain image [27]. It can be understood as a network centrality index to represent the importance of nodes in the human brain connectome in their local functional interaction.

For individuals with ASD, local connections are disrupted in different ways. It results in differences in the distribution and topology of ReHo values across the entire brain region between individuals with ASD and healthy people. A region with a higher ReHo value indicates higher consistency with the time series of surrounding regions. We are concerned about the connectivity of human brain regions between ASD subjects and healthy people in regions with relatively high local consistency. Therefore, it is necessary to characterize the topological structure of the time series of adjacent voxels in areas greater than or equal to the given correlation value.

We converted every original ReHo value to 1 minus original ReHo value. Given a threshold, voxels with ReHo values greater than or equal to the threshold before transformation were the same as voxels with the voxel values less than or equal to 1 minus the threshold after transformation. Using all the voxels of each subject from ABIDE I and ABIDE II as the regions of interest (ROIs), we used the persistent homology to calculate the desired topological information. Cubical complexes were used to represent voxel data. Given a filtration value of ϵ, if the voxel value of a 3-dimensional cube of a subject was less than or equal to ϵ, the cube was labeled as A, otherwise, it was labeled as B. In this way, we obtained the 0-, 1- and 2-dimensional topological generators of all cubes labeled A under this filtration value. As the filtration value ϵ increased, the number of cubes labeled as A also gradually increased. We could obtain birth and death information of the 0-, 1- and 2-dimensional topological generators throughout the entire process. Thus, we obtained the 0-, 1- and 2-dimensional persistence barcodes or persistence diagrams of the ROIs of every subject.

Due to the rule that if a higher dimensional cube is added in the filtering process, each proper face of it should be added before this. There are always some 0-dimensional features that are generated earlier than the 1- and 2-dimensional features. There are always some 1-dimensional features that are generated earlier than the 2-dimensional features. The threshold range for each dimension we chose was to ensure that the majority of persistence barcodes generated from data from different sites were within this range as much as possible. For the ABIDE I database, the 0-dimensional filtration values were increased by 0.01 each time from 0.6 to 0.95, for a total of 36 different thresholds. For every filtration value of these 36 thresholds, we counted the number of 0-dimensional bars whose persistence intervals contained the filtration value. We used the number obtained at each threshold as a feature value corresponding to this threshold. Following this method, we obtained 36 0-dimensional features for every subject. Similarly, the numbers of 1-dimensional bars that crossed given filtration values were counted, respectively (Filtration values were increased by 0.01, ranging from 0.65 to 0.97, resulting in a total of 33 thresholds). We obtained 33 1-dimensional features for every subject. We also counted the number of 2-dimensional bars in which every given filtration value was located (Filtration values were increased by 0.01, ranging from 0.8 to 0.97, resulting in a total of 18 thresholds). We obtained 18 2-dimensional features for every subject. Then, we combined all 0-, 1- and 2-dimensional features into an 87-dimensional vector for every subject. Figure 3 shows the process of converting from persistence barcodes to a feature vector.

The data we selected came from 17 international sites, with an age distribution ranging from 6 to 64 years old, as well as differences in genders and DSM-IV-TR diagnostic criteria. ReHo values could be affected and the transformed voxel values could be also affected by these variables. We performed zero-mean operation for each feature of 87 features of all subjects, and removed the influence of covariates, such as international sites, genders and diagnostic criteria. And then we added 87 mean values back to the corresponding features of each subject, respectively. Finally, we performed a linear regression between each feature of all subjects and their age, respectively. The residuals before and after regression of each feature of all subjects with age were used as the inputs to machine learning classifiers to classify ASD subjects and typical controls.

For the ABIDE II database, we counted the numbers of 0-dimensional bars whose persistence intervals contained filtration values (The filtration values were increased by 0.01, ranging from 0.55 to 0.8, resulting in a total of 26 filtration values). We obtained 26 0-dimensional features for every subject. The numbers of 1-dimensional bars that crossed given filtration values were counted, respectively (Filtration values were increased by 0.01, ranging from 0.6 to 0.85, resulting in a total of 26 thresholds). We obtained 26 1-dimensional features for every subject. We also counted the number of 2-dimensional bars in which each filtration value was located (Filtration values were

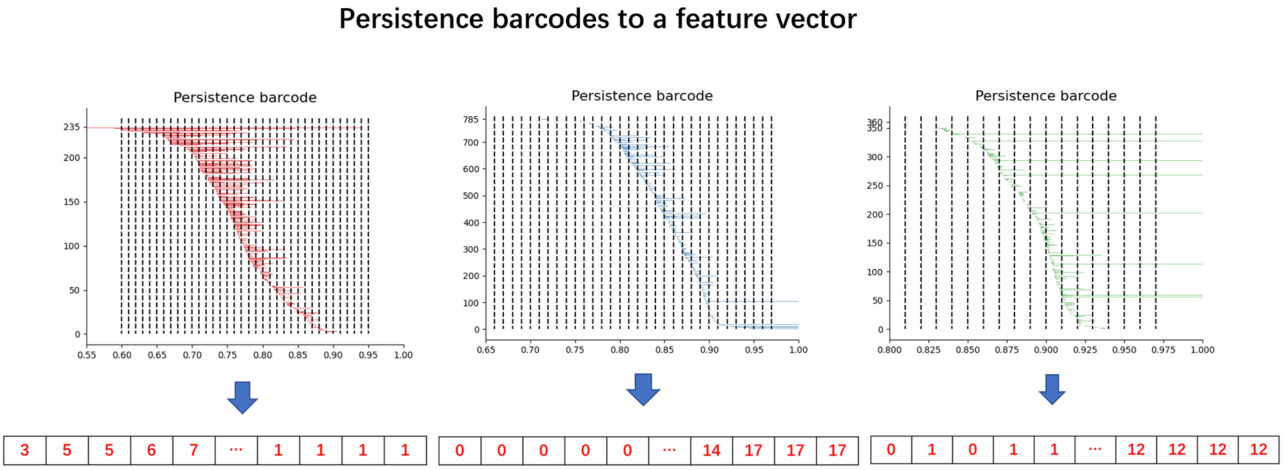


Fig. 3 Persistence barcodes to a feature vector. For given 87 filtration values, we counted separately the numbers of 0-, 1- and 2-dimensional bars whose persistence intervals contained these filtration values and then took these numbers to form an 87-dimensional feature vector

increased by 0.01, ranging from 0.7 to 0.9, resulting in a total of 21 thresholds). We obtained 21 2-dimensional features for every subject. Then, we combined all 0-, 1- and 2-dimensional features into a 73-dimensional vector for every subject. We selected data from 16 international sites, the age distribution from 5 to 64 years and the genders of subjects could affect the ReHo values and thus also affect the transformed voxel values. The following operations were the same as those in subjects from ABIDE I. The residuals of each feature of all subjects before and after regression with age were used as features of machine learning classifiers.

We extracted features for learning classifiers and then evaluated models and verified how good (or bad) they were. For a binary classification model of supervised learning, accuracy is a very important evaluation index. It is the proportion of the number of labels of the classified objects that are the same as their original labels under the extracted features to the total number of labels. For binary classification problems, the combination of accuracy and other indexes can evaluate the performance of the model more precisely.

The receiver operating characteristic (ROC) curve is also known as the sensitivity curve. The ROC curve is a coordinate graph composed of false positive rate (FPR) as horizontal axis and true positive rate (TPR) as vertical axis, and the curve drawn by subjects under specific stimulus conditions due to different results obtained by different judgment criteria. The area under ROC curve (AUC) can be used to evaluate the performance of the binary problem machine learning algorithms. Sensitivity and specificity indicate what proportion of positive and negative cases are correctly classified, respectively.

Experimental Results and Comparison with Existing Methods

We note that the accuracy, specificity, sensitivity and F1 score mentioned from the "Experimental Results and Comparison with Existing Methods" to "Conclusions" sections refer to the average under 10 times 10-fold cross validation.

Experimental Results

The resulting accuracy, specificity, sensitivity and F1 score of performing 10 times 10-fold cross validation on the entire ABIDE I database under the Support Vector Machine (SVM) [47], Multilayer Perceptron (MLP) [48], Random Forest (RF) [49] and Gradient Boosting Decision Tree (GBDT) [50] algorithms are shown in Table 3. This result indicated that TDA could be a very effective feature extraction tool for studying the ReHo data of fMRI. Figure 4 shows the two ROC curves with the worst performance of 10-fold cross validation using the GBDT algorithm on the ABIDE I database and the AUC corresponding to each fold.

Table 3 The resulting of performing on the entire ABIDE I database under SVM, MLP, RF and GBDT algorithms

algorithms	avg-acc	avg-spec	avg-sen	avg-F1 score
SVM	84.2%	90.4%	76.8%	79.7%
MLP	90.7%	90.7%	90.6%	89.5%
RF	92.5%	97.7%	86.1%	90.0%
GBDT	95.6%	96.7%	94.3%	95.0%

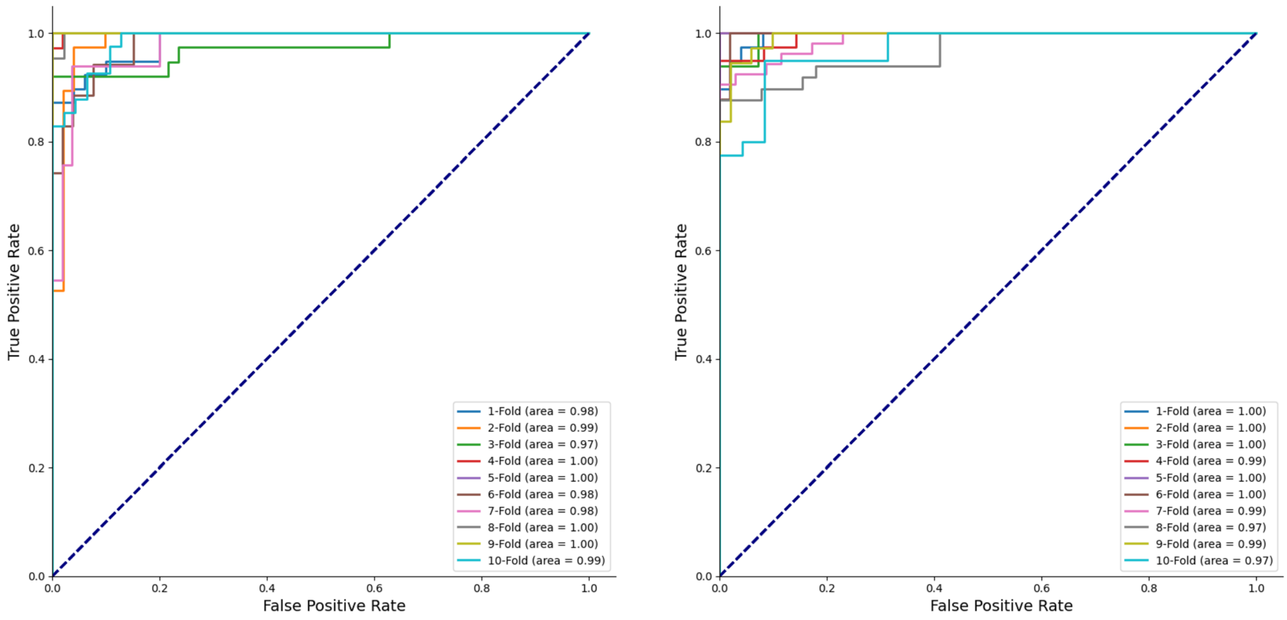


Fig. 4 The ROC curves of the worst 2 times 10-fold cross validation and the corresponding AUC of each fold by using the GBDT algorithm on ABIDE I

For the ABIDE I database, we selected the top 30 most important features for the GBDT classifier, with contributions ranging from 70% to 80%. We conducted t-tests on ASD subjects and typical controls using these features that were often in the top 30 under 10 times 10-fold cross validation. We chose these with significant differences (p-value ≤ 0.05) from important features that often appeared. We used these important and significantly different features as inputs for the GBDT classifier. The classification accuracy, specificity, sensitivity and F1 score of the model constructed with these features were 92.3%, 94.3%, 89.8% and 91.1%, respectively.

The accuracy, specificity, sensitivity and F1 score obtained through 10 times 10-fold cross validation on the ABIDE II database using SVM, MLP, RF and GBDT algorithms are shown in Table 4. Figure 5 shows the two ROC curves with the worst performance of 10-fold cross validation using the GBDT algorithm on this database and the AUC corresponding to each fold.

Table 4 The resulting of performing on the entire ABIDE II database under SVM, MLP, RF and GBDT algorithms

algorithms	avg-acc	avg-spec	avg-sen	avg-F1 score
SVM	86.1%	99.2%	38.2%	52.3%
MLP	94.4%	95.0%	92.6%	86.5%
RF	95.2%	99.7%	79.2%	84.8%
GBDT	96.5%	97.6%	92.6%	92.4%

For the ABIDE II database, we also selected the top 30 most important features for the GBDT classifier, with contributions ranging from 60% to 70% in this database. We also conducted t-tests on ASD subjects and typical controls from the top 30 features that frequently appeared in cross validation. We also chose these with significant differences (p-value ≤ 0.05) from important features that often appeared. And we used these important and significantly different features of subjects as inputs for the GBDT classifier. The classification accuracy, specificity, sensitivity and F1 score of the model constructed with these features were 96.4%, 97.6%, 92.5% and 92.0%, respectively.

Comparisons with Existing Methods

To the best of our knowledge, our accuracy based on persistent homology on both subdatabases was the most accurate among existing results. The comparisons of all 26 models in the ABIDE I were shown in Table 5. The accuracy results between all 26 models are shown in Fig. 6. The accuracy of other 25 models ranged from 60% to 93.59%, and most models had classification accuracy below 80%. The classification accuracy of our proposed model reached 95.6%, which was about 2% higher than the highest one of the other 25 models. The comparisons of all 12 models in the ABIDE II are shown in Table 6. The accuracy results between all 12 models are shown in Fig. 7. The accuracy of our proposed model reached 96.5%, which was much higher than that of the other 11 models. This was a very huge improvement on this database. These results indicated that features extracted

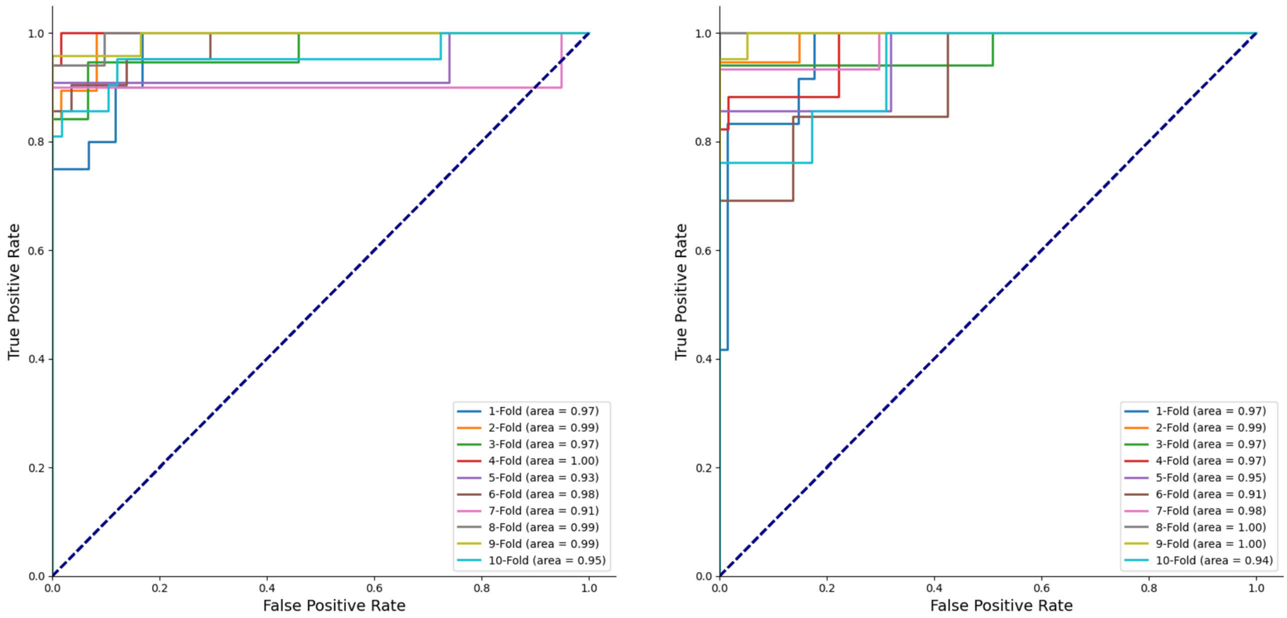


Fig. 5 The ROC curves of the worst 2 times 10-fold cross validation and the corresponding AUC of each fold by using the GBDT algorithm on ABIDE II

using the persistent homology could better represent the data from two databases.

Compared to previous methods, we neither directly utilized time series nor ReHo data, but rather used persistent homology to obtain the topological features of ReHo values for all voxels in the whole brain. We extracted features, removed covariates and then send them into machine learning classifiers. Thus, our model had fewer hyperparameters than traditional deep learning models. This avoided the effect on the classification results due to the selection of more parameters. It took lots of time to train deep learning models (typically several hours, a day or even a few days). Our model chose cubical complexes to represent data and then established filtration complexes, which greatly accelerated the running time. The entire process to get the average accuracy of cross validation ran for about 20 min. We used a 12th Generation Intel Core i9 processor with 14 cores running at 2.50 GHz and 16 GB of RAM and a RTX 3060 Laptop GPU with 6 GB of RAM. This provided time facilitation for the diagnostic process.

Discussions

We established a new model that combined persistent homology in TDA with some machine learning models to classify ASD populations and healthy people. To the best of our knowledge, we obtained the highest accuracy on both commonly used databases. Other evaluation indexes performed well. All results indicated that there were structural differences in ReHo between two groups of people. The different dimensional and multi-scale connectivity differences of ReHo between the two groups of people were captured by persistent homology. ReHo is very effective for studying ASD. The main reasons of very high results were as follows: In our model, the data obtained after transforming ReHo data was still volume data, it could be well characterized by a series of nested cubical complexes. The connectivity information obtained by persistent homology were topological invariants that remained unchanged under continuous deformation could be used to distinguish different topological structures. We chose the proper filtration process to establish our model. The filtration values range for each dimension we chose was to ensure that the majority of persistence barcodes generated from data were within this range as much as possible. Compared to other models, we neither considered the correlation between any two ROIs nor directly used some ReHo values as features, but rather utilized the connectivity information of voxels with high homogeneity throughout the brain as features. These features not only reflected the intrinsic structural information of the complexes, but also maintained multi-scale properties (under different filtration values). These factors were not considered by other models. The method that regarded the correlation of regions on time series as features did not consider the impact of correlation between three or even more regions on ASD prediction. Our model considered the relationship among multiple voxels.

Table 5 The performance of our proposed model with other models in the entire ABIDE I database

Models	Pub Date	Sample Size	Data	Accuracy
MFC [51]	2013	964	4D fMRI	60%
SVM-169 [52]	2015	878	ReHo	63.03%
DRB [53]	2017	871	4D fMRI	66.8%
CP-DNN [13]	2018	1013	4D fMRI	68.7%
PBL-McRBFN-169 [52]	2015	878	ReHo	68.9%
SGCN [54]	2017	871	4D fMRI	69.5%
DL algorithm [3]	2018	1035	4D fMRI	70.0%
LSTM networks [55]	2017	1100	4D fMRI	70.1%
ASD-DiagNet [18]	2019	1035	4D fMRI	70.1%
GCN [56]	2018	871	4D fMRI	70.4%
ASD-SAENet [17]	2021	1035	4D fMRI	70.8%
AIM [57]	2018	871	4D fMRI	71.1%
3D-CNN CC400 [58]	2018	774	4D fMRI	71.7%
Ensemble 3D-CNN [59]	2018	774	4D fMRI	73.3%
SH-ML [14]	2021	656	4D fMRI	73.75%
CAFN [60]	2021	452	4D fMRI	75%
SFM [15]	2017	1035	4D fMRI	77.3%
EBN [19]	2019	871	4D fMRI	77.7%
MISO-DNN [61]	2021	1038	4D fMRI	78.07%
PTA-DNN [62]	2019	871	4D fMRI	79.2%
ECNN [63]	2022	1112	4D fMRI	80%
Monkey-Derived [20]	2020	336	4D fMRI	82.14%
SSAE [64]	2022	871	4D fMRI	87.2%
BDN [21]	2020	1112	4D fMRI	88.8%
SVM-RFE AE [23]	2019	1054	4D fMRI	93.59%
Our proposed model		884	ReHo	**95.6%**

Rigorous Interpretation of Our Work

The high classification results were reliable and the design of the whole work was also rigorous. Firstly, there are abnormalities in the structure of certain brain regions for ASD and healthy populations. The functional connections and degree of connectivity between different brain regions also differ from those of healthy individuals. ReHo describes the functional connection relationship between a given node and its nearest 26 neighbor nodes. In individuals with autism, local connections are disrupted in different ways, resulting in differences in the distribution and topology of ReHo values across the entire brain region between ASD and healthy populations. Secondly, homology is a tool for characterizing the structure of objects. The theories of homology and persistent homology have clear definitions and are highly rigorous and refined in their theoretical derivation. Persistent homology in TDA could capture multi-dimensional and multi-scale connectivity information of ReHo data of subjects. In recent years, there have been many applications of TDA, as mentioned in the "Topological Data Analysis" section, which achieved excellent results in many studies related to data structure. Thirdly, the data obtained after transforming ReHo data was still volume data, we chose cubical complexes to represent this data, which provided a good topological representation. Finally, the filtration values range for each dimension we chose was to ensure that the majority of persistence barcodes generated from data from different sites were within this range as much as possible. These reasons were the prerequisite and key to achieving high classification accuracy and various good performance indexes. The results obtained by persistent homology can be visualized through persistent barcodes or persistent diagram, through which we can easily obtain features under some filtration values. This helps us understand the structure of objects from different scales.

In this paper, the numbers of ASD subjects on both ABIDE I and II databases were in a suitable proportion to that of typical control subjects (408:476 and 342:453), and all data was randomly grouped under a 10-fold cross validation. Therefore, AUC could indeed reflect the strength of our model. We chose the commonly used 10-fold cross validation to ensure that our experiment did not overfit or underfit. This made the experimental results more credible.

Limitations

The original features extracted by our model were the number of bars at some certain thresholds, which were the Betti numbers corresponding to the complexes at these thresholds. They reflected the number of connected components, loops and cavities of the entire ReHo data under these filtration values. Although these features reflected some of the structure of the data and we have achieved high classification accuracy, the medical and biological significance represented by these features was not yet clear. The distributions, sizes and specific locations of connected components, loops and cavities in the brain could not be directly reflected through these features. It was not sufficient to directly locate abnormal brain regions in diagnostic patients. Starting from topological features, exploring the biological significance of topological features that differ between individuals with ASD and healthy people is our future work. This may contribute to medical diagnosis.

Future Directions

Although our proposed method has achieved good accuracy results in ASD detection, there is still room for further improvement. We can also optimize the extracted topological features by combining them with our prior knowledge of this disease. In addition, as a topological feature extraction,

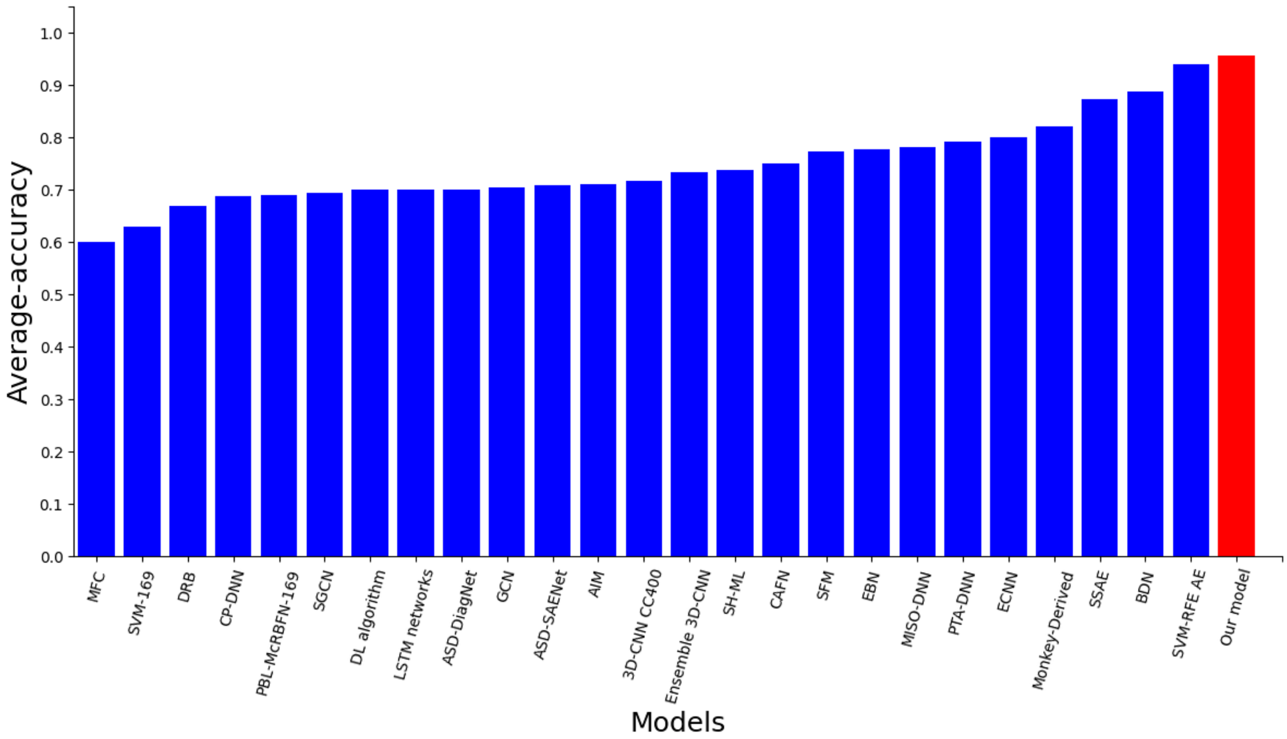


Fig. 6 Performance comparison of our model with other existing models on the ABIDE I database

there may be a better choice than using all Reho values of every subject as ROIs.

Our proposed model can also be used in ReHo data from rs-fMRI of other neurological diseases to classify, such as Alzheimer's disease, attention deficit hyperactivity disorder, and so on. Many articles in the medical field indicated that ASD is associated with a lot of brain function areas. Multiple brain regions as ROIs to extract features is also a viable direction.

In addition, there are some related generalizations of persistent homology that can also be used as improvements on the two databases. Recently, many new mathematical theories have been developed, such as hypergraph-based persistent embedded homology [67] and super-persistent homology theories [68], which allow for topological invariants to be applied to a wider range of problems. These theories can be used to analyze point cloud data as well as graph data, while overcoming the topological noise and constraint requirements for data in persistent homology theory. Applying these theories to fMRI data to characterize the structure of data is also a promising method for the ASD classification problem.

Table 6 The performance of our proposed model with other models in the entire ABIDE II database

Models	Pub Date	Sample Size	Data	Accuracy
3D-CNN EZ [57]	2018	393	4D fMRI	66.4%
3D-CNN DOS160 [57]	2018	393	4D fMRI	67.0%
SVM HO [57]	2018	393	4D fMRI	68.7%
3D-CNN AAL [57]	2018	393	4D fMRI	69.5%
MCNNE [65]	2019	343	4D fMRI	70.0%
BrainNet CC400 [57]	2018	393	4D fMRI	71.5%
3D-CNN MA-Ensemble [57]	2018	393	4D fMRI	71.7%
3D-CNN SP-Ensemble [57]	2018	393	4D fMRI	72.3%
3D-CNN CC200 [57]	2018	393	4D fMRI	72.8%
FNDA [66]	2021	352	4D fMRI	73.1%
Monkey-Derived [20]	2020	149	4D fMRI	75.17%
Our proposed model		795	ReHo	**96.5%**

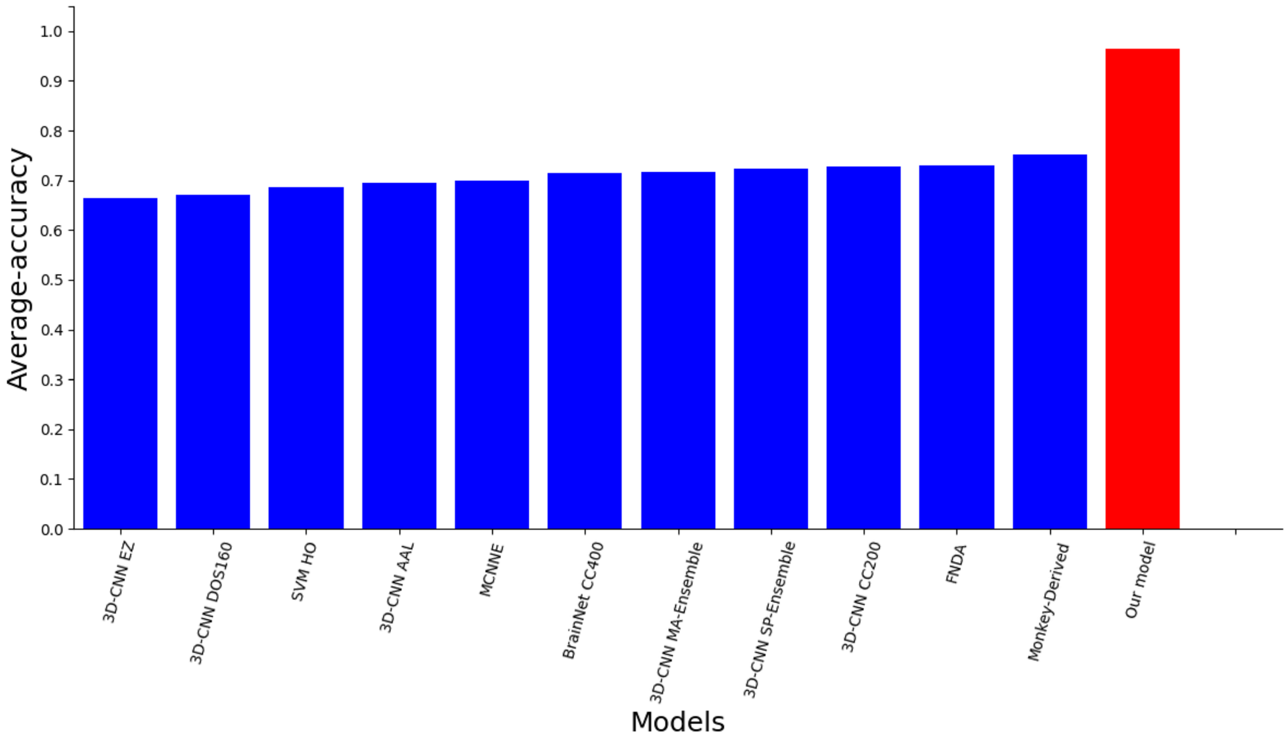


Fig. 7 Performance comparison of our model with other existing models on the ABIDE II database

Conclusions

In this paper, we proposed a new model for extracting topological feature from volume data using persistent homology in TDA. We used this model to classify individuals with ASD and healthy populations. The classification data was a publicly shared ABIDE database commonly used by ASD research, which includes two subdatabases, ABIDE I and ABIDE II. The different dimensional connectivity information (the number of connected components, loops and cavities) of the ReHo data of each subject from ABIDE I and ABIDE II were extracted by persistent homology as the features. We obtained the higher classification accuracy than other state-of-the-art results on both ABIDE I and ABIDE II, and other evaluation indexes performed well, such as specificity, sensitivity, AUC and F1 score. The results indicated that there were differences in the connectivity of different dimensions and scales of ReHo values. Persistent homology captured the spatial structure differences of ReHo between two groups of people, and thus distinguished ASD individuals and healthy people. Our model opened up a new perspective for studying ASD and related types of diseases.

Acknowledgements We thank the publicly shared fMRI data provided by the ABIDE database and the preprocessing for the ABIDE I database.

Author Contributions Xudong Zhang and Yaru Gao wrote articles, Xudong Zhang established relevant models and edited code. Fengling Li, Huanjie Li and Fengchun Lei supervised this study and provided guidance on the writing of this manuscript. All authors revised this manuscript. Yunge Zhang, Xudong Zhang and Huanjie Li obtained and preprocessed the data used in this study. Fengling Li, Huanjie Li, Yaru Gao and Yunge Zhang rechecked the code. All authors collected relevant literature. All authors have read and agreed to publish the final manuscript.

Funding This work was supported in part by a grant (No. 12071051) of National Natural Science Foundation of China (NSFC) and State Key Laboratory of Structural Analysis, Optimization and CAE Software for Industrial Equipment.

Code Availability The identification of ASD applies our established codes, which are publicly available in the Github repository: https://github.com/zxd1-1/zxd1-1.

Declarations

Ethics Approval This study uses data that are publicly available. All data have been openly released to the scientific community. In accordance with HIPAA guidelines and 1000 Functional Connectomes Project/INDI protocols, all datasets are anonymous, with no protected health information included. The Institute Research Ethics Committee has confirmed that no ethical approval is required.

Consent to Participate Informed consent was obtained from all individual participants included in the study.

Consent for Publication The manuscript contains no identifiable individual data or images that would require consent to publish from any participant.

Conflict of Interest The authors declare no competing interests.

References

1. T. Wahlberg, A. F. Rotatori, J. Deisinger, S. Burkhardt, Students with autism spectrum disorders, Advances in Special Education 15 (03) (2003) 195–232.
2. M. A. Just, T. A. Keller, V. L. Malave, R. K. Kana, S. Varma, Autism as a neural systems disorder: A theory of frontal-posterior underconnectivity, Neurosci Biobehav Rev 36 (4) (2012) 1292–1313.
3. A. S. Heinsfeld, A. R. Franco, R. C. Craddock, A. Buchweitz, F. Meneguzzi, Identification of autism spectrum disorder using deep learning and the abide dataset, NeuroImage: Clinical 17 (2018) 16–23.
4. C. Wong, E. L. Meaburn, A. Ronald, T. S. Price, J. Mill, Methylomic analysis of monozygotic twins discordant for autism spectrum disorder and related behavioural traits, Molecular Psychiatry.
5. K. Lyall, J. N. Constantino, M. G. Weisskopf, A. L. Roberts, A. Ascherio, S. L. Santangelo, Parental social responsiveness and risk of autism spectrum disorder in offspring, Jama Psychiatry 71 (8) (2014) 936–942.
6. Elmose, Mette, Happe, Francesca, Being aware of own performance: How accurately do children with autism spectrum disorder judge own memory performance?, Autism Research Official Journal of the International Society for Autism Research (2014).
7. R. J. Swatzyna, N. N. Boutros, A. C. Genovese, E. K. MacInerney, A. J. Roark, G. P. Kozlowski, Electroencephalogram (eeg) for children with autism spectrum disorder: Evidential considerations for routine screening, European Child & Adolescent Psychiatry 28 (2019) 615–624.
8. R. A. Carper, P. Moses, Z. D. Tigue, E. Courchesne, Cerebral lobes in autism: Early hyperplasia and abnormal age effects, Neuroimage 16 (4) (2002) 1038–1051.
9. S. R. Chandana, M. E. Behen, C. Juhász, O. Muzik, R. D. Rothermel, T. J. Mangner, P. K. Chakraborty, H. T. Chugani, D. C. Chugani, Significance of abnormalities in developmental trajectory and asymmetry of cortical serotonin synthesis in autism, International Journal of Developmental Neuroscience 23 (2-3) (2005) 171–182.
10. S. Ogawa, T. M. Lee, A. Tank, Brain magnetic resonance imaging with contrast dependent on blood oxygenation, Proceedings of the National Academy of Sciences of the United States of America 87 (24) (1990) 9868–9872.
11. N. M. Kleinhans, R. Müller, D. N. Cohen, E. Courchesne, Atypical functional lateralization of language in autism spectrum disorders, Brain Research 1221 (2008) 115–125.
12. G. J. Harris, C. F. Chabris, J. Clark, T. Urban, I. Aharon, S. Steele, L. Mcgrath, K. Condouris, H. Tager-Flusberg, Brain activation during semantic processing in autism spectrum disorders via functional magnetic resonance imaging, Brain & Cognition 61 (1) (2006) 54–68.
13. C. J. Brown, J. Kawahara, G. Hamarneh, Connectome priors in deep neural networks to predict autism, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), 2018.
14. M. A. Reiter, A. Jahedi, A. J. Fredo, I. Fishman, B. Bailey, R.-A. Müller, Performance of machine learning classification models of autism using resting-state fmri is contingent on sample heterogeneity, Neural Computing and Applications 33 (2021) 3299–3310.
15. V. Subbaraju, M. B. Suresh, S. Sundaram, S. Narasimhan, Identifying differences in brain activities and an accurate detection of autism spectrum disorder using resting state functional-magnetic resonance imaging : A spatial filtering approach, Medical Image Analysis 35 (2017) 375–389.
16. H. Felouat, S. Oukid-Khouas, Graph convolutional networks and functional connectivity for identification of autism spectrum disorder, in: 2020 Second International Conference on Embedded & Distributed Systems (EDiS), 2020.
17. F. Almuqhim, F. Saeed, Asd-saenet: A sparse autoencoder, and deep-neural network model for detecting autism spectrum disorder (asd) using fmri data, Frontiers in Computational Neuroscience 15 (2021).
18. T. Eslami, V. Mirjalili, A. Fong, A. Laird, F. Saeed, Asd-diagnet: A hybrid learning approach for detection of autism spectrum disorder using fmri data, Frontiers in Neuroinformatics (2019).
19. S. Mostafa, L. Tang, F.-X. Wu, Diagnosis of autism spectrum disorder based on eigenvalues of brain networks, Ieee Access 7 (2019) 128474–128486.
20. Y. Zhan, J. Wei, J. Liang, X. Xu, Z. Wang, Diagnostic classification for human autism and obsessive-compulsive disorder based on machine learning from a primate genetic model, American Journal of Psychiatry 178 (1) (2020) appi.ajp.2020.1.
21. H. Guo, W. Yin, S. Mostafa, F. X. Wu, Diagnosis of asd from rs-fmri images based on brain dynamic networks, in: Springer, Cham, 2020.
22. E. Canario, D. Chen, B. Biswal, A review of resting-state fmri and its use to examine psychiatric disorders, Psychoradiology (2021).
23. C. Wang, Z. Xiao, B. Wang, J. Wu, Identification of autism based on svm-rfe and stacked sparse auto-encoder, IEEE Access PP (99) (2019) 1–1.
24. R. Ghrist, Barcodes: The persistent topology of data, Bulletin of the American Mathematical Society 45 (1) (2008) 61–75.
25. G. Carlsson, Topology and data, Bulletin of the American Mathematical Society 46 (2) (2009) 255–308.
26. P. Bubenik, Statistical topological data analysis using persistence landscapes, Journal of Machine Learning Research 16 (1) (2015) 77–102.
27. Y. Zang, T. Jiang, Y. Lu, Y. He, L. Tian, Regional homogeneity approach to fmri data analysis, Neuroimage 22 (1) (2004) 394–400.
28. Y. Behzadi, K. Restom, J. Liau, T. T. Liu, A component based noise correction method (compcor) for bold and perfusion based fmri., Neuroimage 37 (1) (2007) 90–101.
29. C. G. Yan, X. D. Wang, X. N. Zuo, Y. F. Zang, Dpabi: Data processing & analysis for (resting-state) brain imaging, Neuroinformatics 14 (3) (2016) 339–351.
30. A. Zomorodian, G. Carlsson, Computing persistent homology, in: Twentieth Symposium on Computational Geometry, 2019.
31. Edelsbrunner, Letscher, Zomorodian, Topological persistence and simplification, Discrete & Computational Geometry 28 (4) (2002) 511–533.
32. C. S. Pun, S. X. Lee, K. Xia, Persistent-homology-based machine learning: a survey and a comparative study, Artificial Intelligence Review 55 (7) (2022) 5169–5213.
33. T. K. Dey, K. Li, S. Jian, D. Cohen-Steiner, Computing geometry-aware handle and tunnel loops in 3d models, ACM Transactions on Graphics 27 (3) (2008).
34. T. K. Dey, Y. Wang, Reeb graphs: Approximation and persistence, ACM (2011).
35. K. Mischaikow, V. Nanda, Morse theory for filtrations and efficient computation of persistent homology, Discrete & Computational Geometry 50 (2) (2013) 330–353.
36. P. Niyogi, S. Smale, S. Weinberger, A topological view of unsupervised learning from noisy data, Siam Journal on Computing (2011).

37. T. Bonis, M. Ovsjanikov, S. Oudot, Persistence-based pooling for shape pose recognition, Springer International Publishing (2016).
38. Z. Cang, L. Mu, K. Wu, K. Opron, K. Xia, G. W. Wei, A topological approach for protein classification, International Society for Optics and Photonics (2015).
39. T. Qaiser, Y. W. Tsang, D. Taniyama, N. Sakamoto, K. Nakane, D. Epstein, N. Rajpoot, Fast and accurate tumor segmentation of histology images using persistent homology and deep convolutional features (2018).
40. Z. Cang, G.-W. Wei, Analysis and prediction of protein folding energy changes upon mutation by element specific persistent homology, Bioinformatics 33 (22) (2017) 3549–3557.
41. J. D. Boissonnat, M. Glisse, C. Maria, M. Yvinec, Gudhi library.
42. A. Tausz, M. Vejdemo-Johansson, H. Adams, Javaplex: A research software package for persistent (co) homology, Software available at http://code.google.com/javaplex 2 (2011).
43. B. T. Fasy, J. Kim, F. Lecci, C. Maria, Introduction to the r package tda, arXiv preprint arXiv:1411.1830 (2014).
44. U. Bauer, Ripser: a lean c++ code for the computation of vietoris–rips persistence barcodes, Software available at https://github.com/Ripser/ripser 436 (2017).
45. U. Bauer, M. Kerber, J. Reininghaus, H. Wagner, Phat–persistent homology algorithms toolbox, Journal of symbolic computation 78 (2017) 76–90.
46. S. Kaji, T. Sudo, K. Ahara, Cubical ripser: Software for computing persistent homology of image and volume data, arXiv preprint arXiv:2005.12692 (2020).
47. V. Vapnik, The support vector method of function estimation, NATO ASI SERIES F COMPUTER AND SYSTEMS SCIENCES (1998).
48. K. Hornik, M. B. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators, Neural Networks (1989).
49. L. Breiman, Random forests, machine learning 45, Journal of Clinical Microbiology 2 (2001) 199–228.
50. J. Friedman, Greedy function approximation : A gradient boosting machine, Annals of Statistics 29 (2001).
51. J. A. Nielsen, B. A. Zielinski, F. P. Thomas, A. L. Alexander, L. Nicholas, E. D. Bigler, J. E. Lainhart, J. S. Anderson, Multisite functional connectivity mri classification of autism: Abide results, Frontiers in Human Neuroscience 7 (1) (2013) 599.
52. S. Vigneshwaran, B. Mahanand, S. Suresh, N. Sundararajan, Using regional homogeneity from functional mri for diagnosis of asd among males, in: 2015 International Joint Conference on Neural Networks (IJCNN), IEEE, 2015, pp. 1–8.
53. A. Abraham, M. P. Milham, A. Di Martino, R. C. Craddock, D. Samaras, B. Thirion, G. Varoquaux, Deriving reproducible biomarkers from multi-site resting-state data: An autism-based example, NeuroImage 147 (2017) 736–745.
54. S. Parisot, S. I. Ktena, E. Ferrante, M. Lee, R. G. Moreno, B. Glocker, D. Rueckert, Spectral graph convolutions for population-based disease prediction, in: Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III 20, Springer, 2017, pp. 177–185.
55. N. C. Dvornek, P. Ventola, K. A. Pelphrey, J. S. Duncan, Identifying autism from resting-state fmri using long short-term memory networks, in: Machine Learning in Medical Imaging: 8th International Workshop, MLMI 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, September 10, 2017, Proceedings 8, Springer, 2017, pp. 362–370.
56. P. Sarah, K. S. Ira, F. Enzo, L. Matthew, G. Ricardo, G. Ben, R. Daniel, Disease prediction using graph convolutional networks: Application to autism spectrum disorder and alzheimer's disease, Medical Image Analysis (2018) S1361841518303554–.
57. M. Khosla, K. Jamison, A. Kuceyeski, M. R. Sabuncu, Ensemble learning with 3d convolutional neural networks for connectome-based prediction, NeuroImage (2018).
58. E. Wong, J. S. Anderson, B. A. Zielinski, P. T. Fletcher, Riemannian regression and classification models of brain networks applied to autism, in: Connectomics in NeuroImaging: Second International Workshop, CNI 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 2, Springer, 2018, pp. 78–87.
59. M. Khosla, K. Jamison, A. Kuceyeski, M. R. Sabuncu, 3d convolutional neural networks for classification of functional connectomes, in: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4, Springer, 2018, pp. 137–145.
60. S. Itani, D. Thanou, Combining anatomical and functional networks for neuropathology identification: A case study on autism spectrum disorder, Medical image analysis 69 (2021) 101986.
61. T. M. Epalle, Y. Song, Z. Liu, H. Lu, Multi-atlas classification of autism spectrum disorder with hinge loss trained deep architectures: Abide i results, Applied soft computing 107 (2021) 107375.
62. S. Mostafa, W. Yin, F.-X. Wu, Autoencoder based methods for diagnosis of autism spectrum disorder, in: International Conference on Computational Advances in Bio and Medical Sciences, Springer, 2019, pp. 39–51.
63. R. Kashef, Ecnn: Enhanced convolutional neural network for efficient diagnosis of autism spectrum disorder, Cognitive Systems Research 71 (2022) 41–49.
64. W. Yin, L. Li, F.-X. Wu, A semi-supervised autoencoder for autism disease diagnosis, Neurocomputing 483 (2022) 140–147.
65. M. A. Aghdam, A. Sharifi, M. M. Pedram, Diagnosis of autism spectrum disorders in young children based on resting-state functional magnetic resonance imaging data using convolutional neural networks, Journal of Digital Imaging 32 (6) (2019) 899–918.
66. M. Pominova, E. Kondrateva, M. Sharaev, A. Bernstein, E. Burnaev, Fader networks for domain adaptation on fmri: Abide-ii study, in: International Conference on Machine Vision, 2021.
67. S. Bressan, J. Li, S. Ren, J. Wu, The embedded homology of hypergraphs and applications, Asian Journal of Mathematics (2016).
68. J. Grbić, J. Wu, K. Xia, G.-W. Wei, Aspects of topological approaches for data science, Foundations of data science (Springfield, Mo.) 4 (2) (2022) 165.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.