NAMANGAN DAVLAT UNIVERSITETI
ILMIY AXBOROTNOMASI, [2026-8]
ISSN:2181-1458
ISSN:2181-0427

01.00.00-FIZIKA-MATEMATIKA
journal.namdu.uz
PHYSICS AND MATHEMATICS

TOPOLOGICAL DATA ANALYSIS IN MENTAL-HEALTH RESEARCH: SYSTEMATIC REVIEW

Sarvinoz Anvarova Xafizjon kizi[1], Munira Rakhmatova Sherzodbek kizi[2], Sejuti Rahman[3]
[1]New Uzbekistan University, Computer science department, master's student; [2]New Uzbekistan University, Computer science department, master's student; [3]New Uzbekistan University, Computer science department, scientific supervisor.
[3]ORCID: 0000-0001-6226-2434; [1]e-mail: sarvinozanvarova26@gmail.com; [2]e-mail: munirarahmatova2004@gmail.com; [3]e-mail: s.rahman@newuu.uz

Abstract. *Mental disorders involve complex and overlapping changes that cannot be readily described by one variable or an equation. The PRISMA-based review process is focused on diseases, databases, and analysis objectives to which the TDA approach was applied. Papers eligible in this study included schizophrenia, depression, severe mental disorder, autism, ADHD, obsessive-compulsive disorder, and substance abuse disorder. Among the data sources used were the ADHD-200 and ABIDE which were the most reused open datasets, COBRE, UK Biobank, ENIGMA-OCD, and several local cohorts. Persistent homology was useful for multi-level analysis independent from one threshold while Mapper was used for identification of subgroups and outcomes. Features extracted by TDA have shown promise in distinguishing individuals with mental illness from control groups; however, these results are insufficient for clinical diagnosis. Therefore, topological features need to be validated and examined alongside statistical, machine learning, and clinical techniques.*

Keywords: *t*opological data analysis; persistent homology; Mapper; mental illness; psychiatry; brain networks; datasets; PRISMA.

RUHIY SALOMATLIK TADQIQOTLARIDA MA'LUMOTLARNING TOPOLOGIK ANALIZI: TIZIMLI TAHLIL

Annotatsiya. *Ruhiy buzilishlar bitta o' zgaruvchi yoki tenglama bilan tavsiflash qiyin bo' lgan murakkab va o' zaro bog' liq o' zgarishlarni qamrab oladi. PRISMA asosidagi ushbu sharhda TDA qo' llanilgan kasalliklar, ma'lumotlar to' plamlari va tahliliy maqsadlar o' rganildi. Sharhga shizofreniya, depressiya, og' ir ruhiy buzilishlar, autizm, ADHD, obses-siv-kompulsiv buzilish va modda iste'moli bilan bog' liq buzilishlar bo' yicha tadqiqotlar kiritildi. ADHD-200 va ABIDE eng ko' p ishlatilgan ochiq ma'lumotlar to' plamlari bo' lib, COBRE, UK Biobank, ENIGMA-OCD va mahalliy kohortlardan ham foydalanilgan. Persistent gomologiya bir chegara qiymatiga bog' liq bo' lmagan ko' p masshtabli tahlilni, Mapper esa kichik guruhlar va natijalarni aniqlashni ta'minladi. TDA xususiyatlari ruhiy buzilishi bo' lgan shaxslarni nazorat guruhidan ajratishda natijaga erishgan bo' lsa-da, klinik tashxis uchun yetarli emas. Shu sababli ular statistik, mashinaviy o' rganish va klinik usullar bilan birgalikda tekshirilishi lozim.*

Kalit so' zlar: *topologik ma'lumotlar tahlili; persistent gomologiya; Mapper; ruhiy kasallik; psixiatriya; miya tarmoqlari; ma'lumotlar bazalari; PRISMA.*

ТОПОЛОГИЧЕСКИЙ АНАЛИЗ ДАННЫХ В ИССЛЕДОВАНИЯХ ПСИХИЧЕСКОГО ЗДОРОВЬЯ: СИСТЕМАТИЧЕСКИЙ ОБЗОР

Аннотация. *Психические расстройства характеризуются сложными и взаимосвязанными изменениями, которые трудно описать одной переменной или уравнением. В систематическом обзоре, проведённом в соответствии с методом PRISMA, рассмотрены расстройства, наборы данных и аналитические задачи, для которых применялся топологический анализ данных (TDA). В обзор вошли исследования шизофрении, депрессии, тяжёлых психических расстройств, аутизма, СДВГ, обсессивно-компульсивного расстройства и расстройств, связанных с употреблением психоактивных веществ. Наиболее часто использовались открытые данные как ADHD-200 и ABIDE, а также COBRE, UK Biobank, ENIGMA-OCD и несколько локальных когорт. Персистентная гомология обеспечивала многомасштабный анализ без выбора единственного порога, а Mapper применялся для выявления подгрупп и исходов. Топологические признаки показали потенциал для различения пациентов и контрольных групп, однако пока недостаточны для клинической диагностики. Поэтому их необходимо валидировать и анализировать совместно со статистическими, клиническими методами и методами машинного обучения.*

Ключевые слова: *топологический анализ данных; персистентная гомология; Mapper; психические расстройства; психиатрия; мозговые сети; наборы данных; PRISMA.*

INTRODUCTION

Clinical syndromes of mental illnesses display great heterogeneity concerning symptoms, courses, comorbidities, and response to treatment. In areas of psychosis, mood disorders, neurodevelopmental disorders, compulsive disorders, and substance-use disorders, studies face two related difficulties in that, on one hand, average differences between cases and controls do not necessarily capture the presence of biologically different subgroups, and, on the other hand, the characteristics of each one could prove to be site-, preprocessing-, or threshold-dependent. Thus, reviews of biomarkers describe dispersed and overlapping abnormalities, not one reproducible lesion [1–3].

Questions guiding the research include: RQ1, what topological data analysis techniques and data representations have been used in the study of mental health? RQ2, for which psychiatric illnesses and datasets has topological data analysis been applied? RQ3, which biological, behavioral, or phenotypic differences have been found? RQ4, how have topological features been used in comparing groups, finding statistical relationships, making classifications, predicting outcomes, or identifying subtypes? RQ5, what are the limitations that hinder translational use, and what standards should apply to transdiagnostic studies?

LITERATURE REVIEW

Previous work and research gap. Usual Graph analysing techniques involve converting connectivity matrix to a graph after the application of a certain threshold and computation of descriptors such as degree, centrality, clustering, or modularity. Although these descriptors provide useful information, they depend on the chosen threshold. TDA, on the other hand, evaluates the structures at different thresholds. Persistent homology captures the creation and destruction of components, loops, and higher dimensional voids while Mapper describes the shape of high dimensional data points as a cluster graph [4–6]. Both approaches maintain the information regarding the global structure of the data that might be lost during edge by edge analysis.

Current reviews either cover TDA for medical imaging and network neuroscience applications [4–6] or discuss conventional biomarkers for psychiatry [1–3]. However, none of them presents a disorder-specific assessment of TDA techniques, data sets, and accessibility. Such an overview is provided below without equating accuracy with practicality.

Mathematical foundations of the reviewed methods

Let G=(V,E,w) be a weighted undirected brain graph, where vertices are regions or sensors and w_{ij} is a functional, structural or spectral association. Persistent homology requires an ordered family of complexes. For correlations r_{ij}, a common dissimilarity is $d_{ij} = 1 - r_{ij}$; alternatives such as $d_{ij} = \sqrt{2\,(1 - r_{ij})}$, $d_{ij} = 1 - |r_{ij}|$ or separate positive and negative filtrations encode different scientific assumptions. The clique (Vietoris–Rips) complex at scale ε contains a simplex whenever all of its pairwise distances are at most ε:

$$K_\varepsilon = \{\sigma \subseteq V : d_{ij} \le \varepsilon \text{ for every } i, j \in \sigma\}, \quad K_{\varepsilon_1} \subseteq K_{\varepsilon_2} \quad \text{when} \quad \varepsilon_1 \le \varepsilon_2.$$

For a simplicial complex K, the boundary operator $\partial_k : C_k(K) \longrightarrow C_{k-1}(K)$ and satisfies $\partial_{k-1} \circ \partial_k = 0$. Cycles are $Z_k(K) = \ker(\partial_k)$, boundaries are $B_k(K) = \mathrm{im}(\partial_{k+1})$, and the quotient removes cycles that bound a higher-dimensional simplex. Thus H_0 counts connected components, H_1 independent loops and H_2 cavities:

$$H_k(K) = \frac{\ker(\partial_k)}{\mathrm{im}(\partial_{k+1})}, \quad \beta_k(K) = \dim H_k(K).$$

As illustrated in Figure 1, the birth and death of H_1 loop happens at t = 0.6 and t = 0.9 respectively, thereby resulting in a lifetime of 0.3.

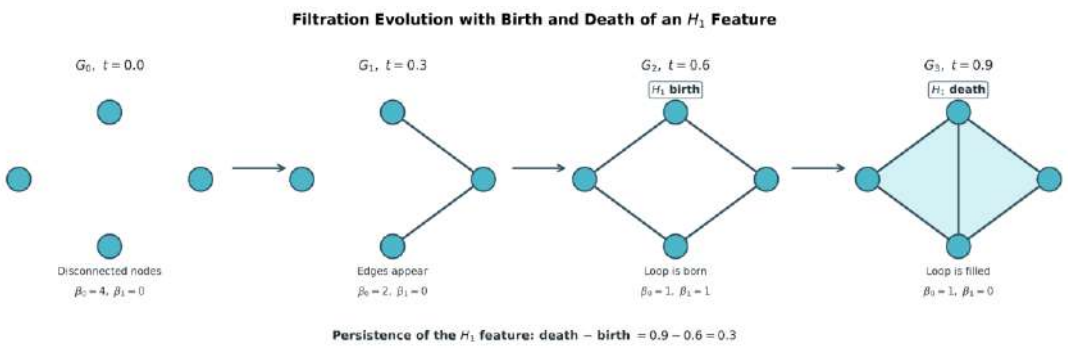


Figure 1. Example of a graph filtration in which an H1 feature is born at 0.6 and filled at 0.9, giving persistence 0.3.

Tracking these groups through the filtration gives intervals $[b_i,d_i)$, summarized by a persistence diagram D_k ={(b_i,d_i)}. Lifetime $\ell_i = d_i - b_i$ measures persistence, while the Betti curve counts features alive at scale t, which is described by an equation as

$$\beta_k(t) = \sum_i \mathbf{1}_{[b_i,d_i)}(t).$$

Persistence landscapes and images represent exactly the same information about the intervals for the purposes of inference and prediction [5-8, 35, 37]. Figure 2 shows a weighted functional brain graph representation and its persistence diagrams: the persistence diagram for H_1 and H_0 Betti curve.

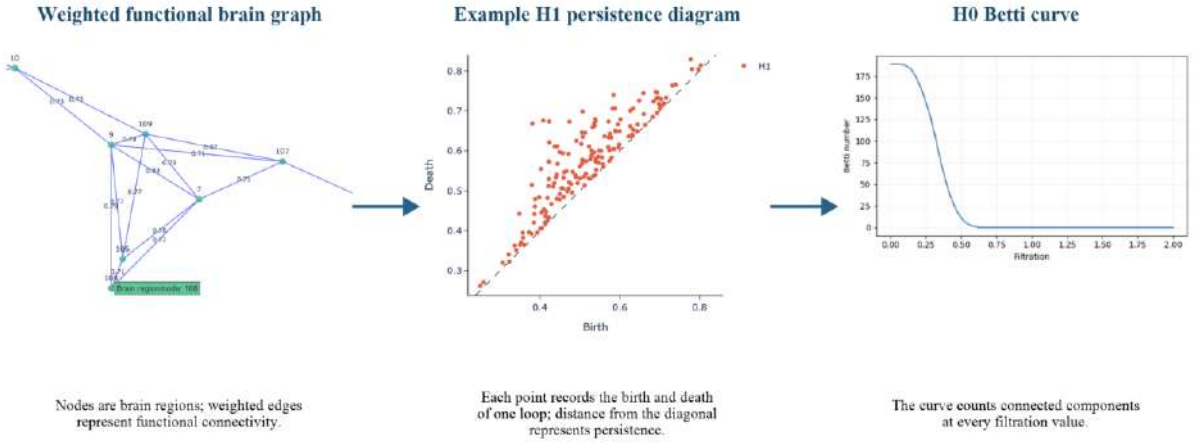


Figure 2. Weighted functional brain graph with persistence-based summaries: an H1 persistence diagram and an H0 Betti curve.

Let *D* and *E* be persistence diagrams, and let $\Delta = \{(a,a) : a \in \mathbb{R}\}$ denote the diagonal, counted with infinite multiplicity. Let Γ(*D*,*E*) be the set of bijections $\Gamma(D,E) = \{\gamma : D \cup \Delta \to E \cup \Delta \mid \gamma \text{ is a bijection}\}$. Thus, a persistence feature may be matched either to a feature in the other diagram or to the diagonal. The bottleneck distance minimizes the largest matching cost:

$$d_B(D,E) = \inf_{\gamma \in \Gamma(D,E)} \sup_{x \in D \cup \Delta} \|x - \gamma(x)\|_\infty$$

In contrast, the p -Wasserstein distance aggregates all matching costs:

$$W_p(D,E) = \left(\inf_{\gamma \in \Gamma(D,E)} \sum_{x \in D \cup \Delta} \|x - \gamma(x)\|_\infty^p \right)^{1/p}.$$

Here, $\|\cdot\|_\infty$ denotes the maximum norm. The stability theorem bounds changes in persistence diagrams by changes in the filtration function:

$$d_B\left(D_k(f), D_k(g)\right) \le \|f - g\|_\infty .$$

However, this theorem does not guarantee clinical robustness to motion, parcellation, or correlation-sign handling because these choices can alter the filtration function itself [36]. This does not imply clinical robustness against movement, parcellation, or sign changes because f can be modified.

Mapper relies on observations X as opposed to anatomical nodes. A lens $f : X \longrightarrow \mathbb{R}^m$ is covered by overlapping sets $\mathcal{U} = \{U_\alpha\}_{\alpha \in A}$ where the clusters of each pullback $f^{-1}(U_\alpha)$ become vertices, while common observations form edges. In this case, we have the nerve of the clustered cover, meaning that the lens, cover resolution, overlap, and clustering are now mathematical parameters, so the mapper graph is $\text{Mapper}(X,f,\mathcal{U}) = \text{Nerve}\left(\text{Clust}\left(f^{-1}(\mathcal{U})\right)\right)$. Instead of association, spectral TDA incorporates frequency-dependent coherence. With Hodge theory, the boundary matrix B_k and combinatorial Laplacian $L_k = B_k^\top B_k + B_{k+1} B_{k+1}^\top$ are used to decompose the gradient, curl, and harmonic parts of simplicial signals, having overall equation as $C_k = \text{im}(B_k^\top) \oplus \ker(L_k) \oplus \text{im}(B_{k+1})$[28]. The overall workflow in terms of data representation and analysis is depicted in Figure 3.

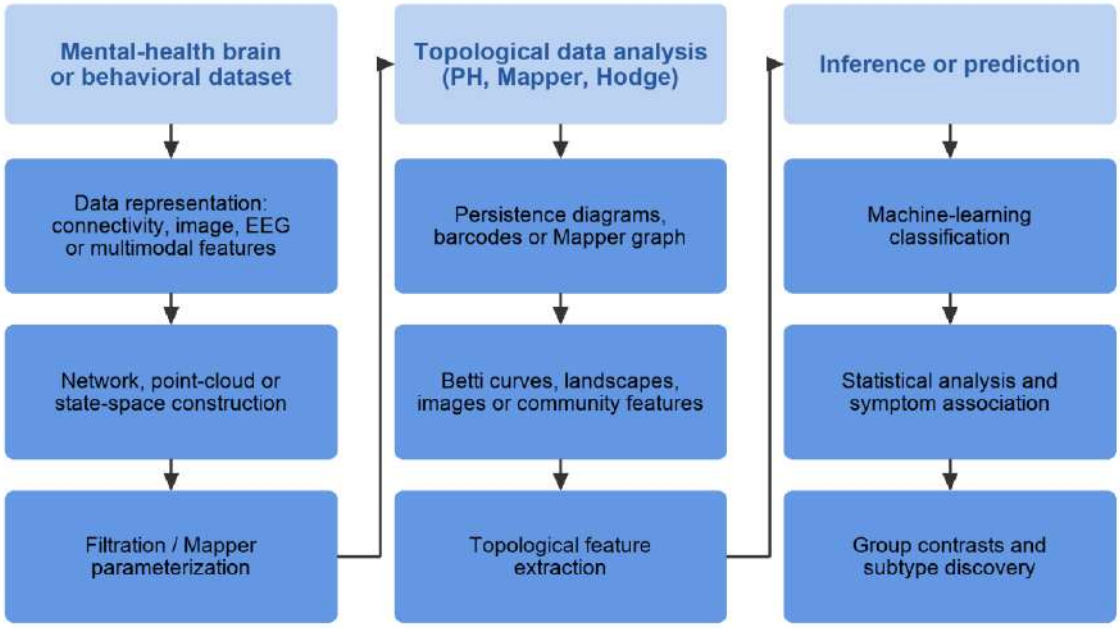


Figure 3.

General workflow for topological analysis of mental-health data.

Table 1. Mathematical comparison of TDA strategies

Method	Mathematical input	Construction	Invariant / summary	Distance or test	Stability	Main parameter
Network PH	Weighted graph $G=(V,E,w)$	Clique/VR filtration K_ε	H_k , β_k , D_k	Bottleneck or Wasserstein	Stable for fixed filtration [36]	Weight-to-distance rule; ε
Cubical PH	Voxel field $f:\mathbb{Z}^3 \to \mathbb{R}$	Sublevel cubical complex	H_0, H_1, H_2	Diagram distance; ML test	Stable to bounded field perturbation	Voxel size; smoothing; level
Mapper	Observations X and lens f	Nerve of clustered pullback cover	Mapper graph	Graph/SAFE measures; group tests	Parameter-sensitive; test resampling	Lens; cover; overlap; clustering
Spectral TDA	Frequency-dependent network	Coherence filtration	Spectral landscapes	Functional/global test	Depends on spectral estimation	Frequency resolution; bandwidth
Hodge TDA	Simplicial signals and B_k	$L_k = B_k{}^T B_k + B_{k+1} B_{k+1}^T$	Harmonic/curl/gradient parts	GLM, permutation, FDR	Complex and skeleton dependent	Cycle basis; skeleton
PH + ML	Diagram or curve vectors	Image/landscape/entropy map	Finite feature vector	CV accuracy; calibration	Must be nested in resampling	Vectorization; classifier

MATERIALS AND METHODS

Review design and search strategy. The review followed PRISMA 2020 [9]. PubMed and arXiv were searched on 7 August 2026 without date or language filters. The PubMed query was ("topological data analysis" OR "persistent homology" OR Mapper OR "Hodge Laplacian" OR "persistence landscape" OR "Betti curve") AND (psychiatr* OR "mental illness" OR "mental health" OR "mental disorder" OR schizophrenia OR psychosis OR depression OR bipolar OR autism OR ADHD OR anxiety OR PTSD OR OCD OR "substance use" OR "substance abuse" OR addiction). The same concepts

were translated to arXiv all-field syntax. Reference lists of two broad TDA reviews [4,5] were checked backward. Exact queries, exports and record-level decisions are retained in the supplementary audit workbook.

Eligibility and data extraction. Eligible reports examined human diagnosed or clinically defined psychiatric cohorts; used patient-level neuroimaging, electrophysiological, genetic, environmental, behavioral or clinical data; applied persistent homology, Mapper, Hodge-based topology or another explicit multiscale topological construction; and reported interpretable group, symptom, subgroup, outcome or predictive results. Reviews, theoretical or non-human work, neurological studies without a psychiatric cohort, conventional graph analyses without TDA and reports with insufficient methods were excluded. Preprints were retained only for distinct methods and labelled as such.

Disorder, sample, dataset, modality, representation, TDA method, summary, objective, validation and conclusion were extracted. Titles/abstracts and full reports were assessed against the same criteria, duplicate and companion reports were linked, and exclusions were logged. Heterogeneity prevented meta-analysis, so study design and direction of findings were synthesized qualitatively. Figure 4 summarizes selection.

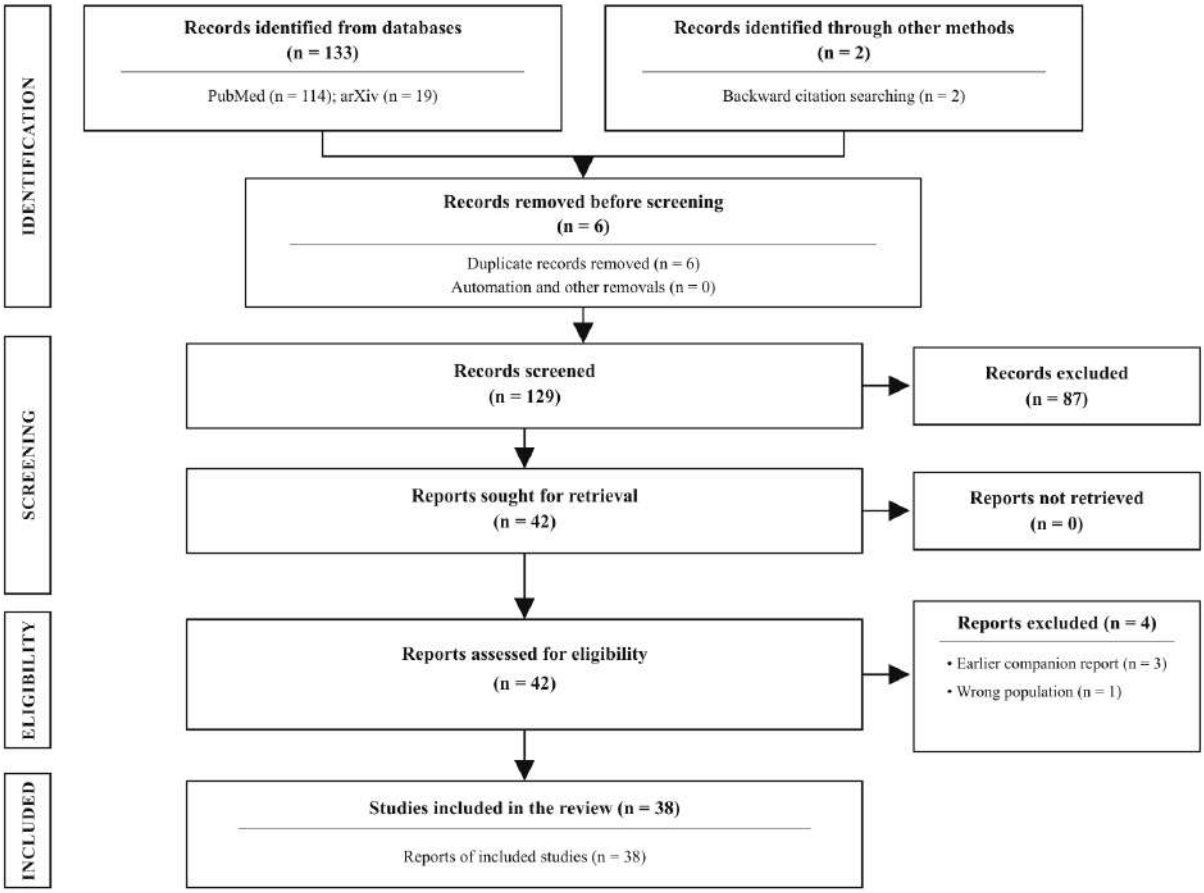


Figure 4. PRISMA flow diagram for study identification and selection.

RESULTS

Study selection and field composition. The searches identified 135 records: 114 from PubMed, 19 from arXiv and 2 through backward citation searching. After removal of 6 duplicate records, 129 titles and abstracts were screened and 87 were excluded. All 42 reports sought were retrieved; four were excluded at full text (three earlier companion reports and one report without an eligible psychiatric cohort), leaving 38 included reports. The evidence remained concentrated in ASD, ADHD and schizophrenia/psychosis, with fewer studies of depression, OCD, serious mental illness and substance use.

Table 2. Mathematical classification of every reviewed study

Method family and studies	Complex and filtration / lens	Homology and topological summary	Distance / test and software	Stability / sensitivity
Mapper [10,11,24–26]	Mapper nerve; clinical, FC, PCA or multimodal lens with clustered cover	N/A; Mapper graph, branches, communities, modularity, SAFE/EL	ANOVA, correlation, regression, outcome and replication tests; software NR	Two-site, independent and multisite checks; no external validation in one study
Graph PH [12,15–16,18–19,27,40,47,49–50,55]	Weighted, functional or dynamic graph; FC, link-density, rank, time-window or magnitude filtration	H_0–H_1; Betti curves, PD, landscapes, kernels, robustness and magnitude homology	Permutation/group tests, classification and kernel PLS; custom, shared or unreported software	Site, preprocessing, state, modality and CV checks
Point-cloud / VR PH [14,20–21,23,41,44,52–54,56]	VR, adaptive VR, dependence, delay-embedding, clinical, movement or permutation complex	H_0–H_2; PD, PI, Betti curves, landscapes, entropy, kernels and persistence descriptors	Bottleneck/Wasserstein, permutation and ML tests; JavaPlex, sklearn-tda, R, Python or custom code	Noise, simulation, representation, classifier, temporal and behavioral checks
Cubical / surface PH [17,51]	Cubical or cortical-surface complex; 1−ReHo or cortical-field sublevels	H_0–H_2; barcode vectors, PD and min–max pairs	Group comparison, SVM and ensemble validation; custom or unreported software	ABIDE I/II and autism-control comparison
Spectral PH [13]	Spectral clique; frequency-dependent coherence filtration	H_1; spectral persistence landscape	Global functional test; software NR	Frequency-band sensitivity
Higher-order PH [22,28,39,42–43,45,48]	Order, simplicial, Hodge, multimodal, interactomic, hypergraph or signed complex	H_1, higher or multidimensional; generators, harmonic coefficients, spectra and signed/hypergraph features	GLM/FDR, group, prediction, ablation and drug-ranking tests; custom or public code	Cross-data, site, medication, modality and cohort checks
Gene-network PH [38,46]	Gene co-expression complex; expression or weighted-network filtration	H_0–H_1; PD and cycle persistence	Group, enrichment and multi-cohort tests; R/Python	Gene-expression evidence; independent-cohort validation

Note: PH—persistent homology; PD—persistence diagram; PI—persistence image; VR—Vietoris–Rips; FC—functional connectivity; PCA—principal component analysis; ML—machine learning; SVM—support vector machine; PLS—partial least squares; GLM—general linear model; FDR—false discovery rate; CV—cross-validation; NR—not reported.

RQ1: differences between topological methods

Persistent homology remained dominant because filtrations test whether differences persist across connection strengths rather than at one threshold. The expanded evidence also included cortical-surface, transcriptomic, behavioral-movement, multimodal PET/MRI and higher-order simplicial constructions [38–50]. Main design choices were the distance transformation, sign handling, complex, homology dimension and vectorization. Clique filtrations suit connectivity matrices but become costly in higher dimensions; cubical or surface filtrations answer spatial-field rather than connectome questions.

Mapper primarily addressed heterogeneity and outcome stratification, whereas newer hybrids combined persistent summaries with kernels, hypergraphs, transformers or permutation tests [49–56]. Mapper branches remain sensitive to lens, cover and clustering choices. Spectral TDA preserves frequency-specific structure, and Hodge methods extend analysis from pairwise edges to signals on higher-order simplices.

RQ2: datasets and modalities

Three evidence tiers emerged: reusable benchmarks (especially ADHD-200 and ABIDE); large or controlled resources such as COBRE, UK Biobank and ENIGMA-OCD; and local imaging, EEG, transcriptomic, behavioral or inpatient cohorts. The expanded studies add multimodal PET/MRI, blood and brain transcriptomics, stereotyped-movement video and further ABIDE/ADHD-200 analyses [38–56]. Repository size must not be confused with the quality-controlled sample actually analysed.

HBN, ABCD, OpenNeuro and the NIMH Data Archive provide future transdiagnostic validation opportunities [32–34]. The principal gap is a harmonized cross-disorder TDA benchmark, not a lack of psychiatric data.

Table 3. Mental-health datasets identified in TDA studies and validation resources

Dataset	Population	Scale	Modalities	Access	TDA status
ADHD-200	ADHD / controls	~800 contributed participants; 8 contributing centers	rs-fMRI, T1, phenotype	Open via INDI; preprocessed derivatives	Primary ADHD TDA resource
ABIDE I	ASD / controls	1,112 released; 17 sites	rs-fMRI, T1, phenotype	Open via INDI	Primary ASD TDA resource
ABIDE II	ASD / controls	1,114 released; 19 sites	rs-fMRI, T1, phenotype	Open via INDI	Used for PH and ReHo topology
COBRE	Schizophrenia / controls	Public cohort; MesoSCOUT used 44+44	DWI, T1; rs-fMRI also available	Open via COINS/NITRC terms	Used for structural-connectome PH
UK Biobank MDD subsets	Major depression	3,052 in 2025 study; larger staged subsets in 2026	Genetic, environmental, imaging, clinical	Controlled application	Used for Mapper/community stratification
GSRD and HSR cohorts	Major depression	Multisite replication cohorts	Clinical, environmental; selected imaging	Collaborative/ controlled	Partial external replication of MDD topology
ENIGMA-OCD	OCD / controls	2,052 in identified analysis	Multisite rs-fMRI	Consortium; not participant-level public	Used for Hodge-Laplacian TDA
Huilongguan EEG	First-episode schizophrenia / controls	198 total	Resting EEG	Local/not identified as public	Used for multimodal persistent features

Dataset	Population	Scale	Modalities	Access	TDA status
Working-memory EEG cohort	Schizophrenia / controls	Local clinical study	Task EEG	Local	Used for VR-PH and stability analysis
Task/semantic fMRI cohorts	Schizophrenia / controls	Small local cohorts	Task fMRI	Local	Used for network and representation PH
SMI inpatient cohort	Mixed serious mental illness	235 adults	rs-fMRI and clinical outcome	Local	Mapper treatment-response stratification
ISAD adolescent cohort	Inhaled substance-use disorder / controls	Local adolescent cohort	rs-fMRI	Local	Persistent-homology group contrasts
Public pediatric EEG	ADHD / controls	Single EEG benchmark	Resting EEG	Public	Spectral TDA
HBN / ABCD	Transdiagnostic youth / population	Thousands; longitudinal	MRI, EEG, behavior, clinical	Open or controlled	Future cross-disorder validation; not included evidence

RQ3: neurobiological findings

ADHD and ASD studies most consistently reported altered multiscale integration or cycle structure rather than one missing connection. Evidence now spans functional connectomes, cortical surfaces, gene co-expression, transcriptomics, multimodal PET/MRI, EEG and motor trajectories [10–18,38–56]. Direction depended on representation, state and filtration, so neither disorder is uniformly more or less connected.

Schizophrenia studies used task and resting fMRI, EEG, semantic representations and white-matter connectomes, converging on altered higher-order organization at different spatial or frequency scales [19–23]. Depression studies used Mapper-like multimodal topology and longitudinal prediction [24,25,44], while serious-mental-illness topology related prefrontal–striatal organization to treatment improvement [26].

Beyond neurodevelopmental disorders, persistent or higher-order topology was applied to inhalant use, anti-addiction transcriptomics and OCD [27,28,42,46]. Across diagnoses, findings concerned integration or higher-order organization, but localization and clinical meaning remained modality- and task-dependent. Few studies compared several diagnoses in one harmonized cohort, so case-control separation does not establish diagnostic specificity.

RQ4: prediction, stratification and clinical association

TDA features entered prediction as persistence images, landscapes, Betti curves, entropy, lifetime statistics and Hodge-derived summaries. Recent hybrids added kernel regression, graph kernels, hypergraph fusion and transformer models [43,45,49–56]. Performance was often promising, but small samples, reused repositories and incomplete external validation prevent diagnostic interpretation.

The most distinctive contribution of Mapper was not binary classification but stratification. Its graphs linked topology to symptoms, comorbidity, treatment improvement or multimodal outcome profiles in ADHD, depression and serious mental illness [10,11,24–26]. Nevertheless, exploratory topological subgroups require stability analysis, independent replication and longitudinal association with outcomes before they can guide treatment.

RQ5: limitations, bias and reporting priorities

Four risks recur. First, data reuse creates an illusion of independent replication: multiple results originate from overlapping ADHD-200 or ABIDE participants, while local psychiatric cohorts may be impossible to reproduce. Second, preprocessing decisions, including motion control, parcellation, EEG referencing, global-signal regression, correlation-to-distance conversion and handling of negative edges, alter the filtration itself. Third, tuning the topological representation and classifier before cross-validation can leak information. Fourth, demographic, medication, illness-stage and comorbidity imbalance limit generalization across psychiatric populations.

Future studies should report subjects/sites, code, filtration definitions, parameters and vectorization. Validation should separate sites or repositories and test preprocessing, parcellation, sign handling and dynamic versus static connectivity. Multiple comparisons, calibration, uncertainty and clinical utility should accompany accuracy.

Table 4. Minimum standards for future clinical TDA studies

Domain	Recommended practice	Reason
Reproducibility	Share code, versions, subject/site counts, exclusions and random seeds	Allows exact reconstruction and detects overlapping samples
Filtration robustness	Report distance transform, sign treatment and threshold range; test alternatives	Shows whether conclusions are genuinely multiscale
Validation	Nested training; leave-site-out and independent-dataset testing	Separates disease signal from site or preprocessing signal
Clinical relevance	Relate features to symptoms, function and longitudinal outcomes	Moves beyond case-control prediction
Equity	Report sex, age, IQ, medication and comorbidity; test subgroup performance	Reduces demographic and spectrum bias
Transdiagnostic design	Analyze ASD, ADHD, co-occurrence and controls jointly	Tests specificity and shared neurodevelopmental dimensions

DISCUSSION

TDA contributes at three levels: filtrations replace a single threshold with multiscale analysis; persistent components and loops characterize biological integration and segregation; and Mapper or other topological feature spaces represent continua, outcome groups and candidate subtypes beyond average diagnostic patterns.

However, TDA cannot correct motion artefacts, site imbalance or inadequate phenotyping. Persistent features are not inherently disease-specific, and Mapper branches are not validated subtypes. Findings are most credible when stable across preprocessing choices, clinically associated and externally validated. TDA should therefore complement conventional connectivity and machine learning.

A transdiagnostic benchmark should compare conventional, graph, persistent-homology, Mapper, Hodge and hybrid features on identical participants and cross-site splits, evaluating accuracy, calibration, test–retest stability, symptoms and treatment outcomes.

This review was limited by inconsistent indexing, broad Mapper false positives, methodological heterogeneity and possible omission of non-indexed or non-English reports. Seven included reports were preprints rather than peer-reviewed evidence [13,22,28,42,54–56]. The record-level audit improves reproducibility but does not replace independent replication or formal risk-of-bias assessment.

CONCLUSION

TDA characterizes psychiatric data through multiscale organization: persistent homology captures components and cycles, Mapper represents continua and candidate subgroups, and spectral or Hodge extensions preserve frequency-specific and higher-order structure. Across 38 included reports, evidence spans ASD, ADHD, schizophrenia/psychosis, depression, serious mental illness, OCD and substance use, but remains uneven and weakly externally validated. TDA should therefore complement statistical, machine-learning and clinical assessment; the next step is a preregistered transdiagnostic benchmark across independent datasets.

AUTHOR STATEMENTS

Author contributions: Sarvinoz Anvarova was responsible for the topological data analysis component, literature search, data extraction, visualization and manuscript preparation. Munira Rakhmatova was responsible for the machine-learning component and contributed to screening, data extraction, validation and manuscript revision. Sejuti Rahman supervised the study, contributed to its conceptual and methodological development and critically reviewed the manuscript. All authors read and approved the final manuscript.

Funding: This research received no specific funding from public, commercial, or non-profit organizations.

Conflict of interest: The authors declare that they have no conflicts of interest related to this study.

FOYDALANILGAN ADABIYOTLAR

1. Insel T., Cuthbert B., Garvey M., Heinssen R., Pine D.S., Quinn K., et al. Research Domain Criteria (RDoC): toward a new classification framework for research on mental disorders. American Journal of Psychiatry. 2010;167(7):748-751. doi:10.1176/appi.ajp.2010.09091379.

2. Uddin L.Q., Dajani D.R., Voorhies W., Bednarz H., Kana R.K. Progress and roadblocks in the search for brain-based biomarkers of autism and attention-deficit/hyperactivity disorder. Translational Psychiatry. 2017;7(8):e1218. doi:10.1038/tp.2017.164.

3. Eitel F., Schulz M.-A., Seiler M., Walter H., Ritter K. Promises and pitfalls of deep neural networks in neuroimaging-based psychiatric research. Experimental Neurology. 2021;339:113608. doi:10.1016/j.expneurol.2021.113608.

4. Singh Y., Farrelly C.M., Hathaway Q.A., Leiner T., Jagtap J., Carlsson G.E., et al. Topological data analysis in medical imaging: current state of the art. Insights into Imaging. 2023;14(1):58. doi:10.1186/s13244-023-01413-w.

5. Sizemore A.E., Phillips-Cremins J.E., Ghrist R., Bassett D.S. The importance of the whole: topological data analysis for the network neuroscientist. Network Neuroscience. 2019;3(3):656-673. doi:10.1162/netn_a_00073.

6. Carlsson G. Topology and data. Bulletin of the American Mathematical Society. 2009;46(2):255-308. doi:10.1090/S0273-0979-09-01249-X.

7. Edelsbrunner H., Harer J. Computational Topology: An Introduction. Providence (RI): American Mathematical Society; 2010. 241 p.

8. Ghrist R. Barcodes: the persistent topology of data. Bulletin of the American Mathematical Society. 2008;45(1):61-75. doi:10.1090/S0273-0979-07-01191-3.

9. Page M.J., McKenzie J.E., Bossuyt P.M., Boutron I., Hoffmann T.C., Mulrow C.D., et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. doi:10.1136/bmj.n71.

10. Kyeong S., Park S., Cheon K.-A., Kim J.-J., Song D.-H., Kim E. A new approach to investigate the association between brain functional connectivity and disease characteristics of attention-deficit/hyperactivity disorder: topological neuroimaging data analysis. PLOS ONE. 2015;10(9):e0137296. doi:10.1371/journal.pone.0137296.

11. Kyeong S., Kim J.-J., Kim E. Novel subgroups of attention-deficit/hyperactivity disorder identified by topological data analysis and their functional network modular organizations. PLOS ONE. 2017;12(8):e0182603. doi:10.1371/journal.pone.0182603.

12. Gracia-Tabuenca Z., Diaz-Patino J.C., Arelio I., Alcauter S. Topological data analysis reveals robust alterations in the whole-brain and frontal lobe functional connectomes in attention-deficit/hyperactivity disorder. eNeuro. 2020;7(3):ENEURO.0543-19.2020. doi:10.1523/ENEURO.0543-19.2020.

13. El-Yaagoubi A.B., Jiao S., Chung M.K., Ombao H. Spectral topological data analysis of brain signals. arXiv:2401.05343 [q-bio.NC]. 2024, revised June 2026. Preprint. doi:10.48550/arXiv.2401.05343.

14. Rathore A., Palande S., Anderson J.S., Zielinski B.A., Fletcher P.T., Wang B. Autism classification using topological features and deep learning: a cautionary tale. In: Medical Image Computing and Computer Assisted Intervention - MICCAI 2019. Lecture Notes in Computer Science, vol. 11766. Cham: Springer; 2019. p. 736-744. doi:10.1007/978-3-030-32248-9_82.

15. Liang D., Xia S., Zhang X., Zhang W. Analysis of brain functional connectivity neural circuits in children with autism based on persistent homology. Frontiers in Human Neuroscience. 2021;15:745671. doi:10.3389/fnhum.2021.745671.

16. Talesh Jafadideh A., Mohammadzadeh Asl B. Topological analysis of brain dynamics in autism based on graph and persistent homology. Computers in Biology and Medicine. 2022;150:106202. doi:10.1016/j.compbiomed.2022.106202.

17. Zhang X., Gao Y., Zhang Y., Li F., Li H., Lei F. Identification of autism spectrum disorder using topological data analysis. Journal of Imaging Informatics in Medicine. 2024;37(3):1023-1037. doi:10.1007/s10278-024-01002-3.

18. Mondal M., Yadav Y., Jost J., Samal A. Node persistence from topological data analysis reveals changes in brain functional connectivity. Patterns. 2026;7:101427. doi:10.1016/j.patter.2025.101427.

19. Stolz B.J., Emerson T., Nahkuri S., Porter M.A., Harrington H.A. Topological data analysis of task-based fMRI data from experiments on schizophrenia. Journal of Physics: Complexity. 2021;2(3):035006. doi:10.1088/2632-072X/abb4c6.

20. Guo G., Zhao Y., Liu C., Fu Y., Xi X., Jin L., et al. Method for persistent topological features extraction of schizophrenia patients' electroencephalography signal based on persistent homology. Frontiers in Computational Neuroscience. 2022;16:1024205. doi:10.3389/fncom.2022.1024205.

21. Kang Y., Zhao J., Zhao Y., Zhao Z., Dong Y., Zhang M., et al. High-order brain network feature extraction and classification method of first-episode schizophrenia: an EEG study. Frontiers in Human Neuroscience. 2024;18:1452197. doi:10.3389/fnhum.2024.1452197.

22. Dmitruk E., Metzner C., Steuber V., Kadir S.N. MesoSCOUT: a novel tool for revealing mesoscale organization in the white-matter connectome. bioRxiv. 2025:2025.06.19.660631. Preprint. doi:10.1101/2025.06.19.660631.

23. Hayashi R., Kaji S., Matsumoto Y., Nishida S., Nishimoto S., Takahashi H. Homogenization of word relationships in schizophrenia: topological analysis of cortical semantic representations. Psychiatry and Clinical Neurosciences. 2024;78(11):687-695. doi:10.1111/pcn.13727.

24. Tassi E., Pigoni A., Turtulici N., Colombo F., Fortaner-Uya L., Bianchi A.M., et al. Gene-environment-brain topology reveals clinical subtypes of depression in UK Biobank. Scientific Reports. 2025;15(1):35538. doi:10.1038/s41598-025-19624-0.

25. Tassi E., Pigoni A., Colombo F., Fortaner-Uya L., Colombo C., Bianchi A.M., et al.; GSRD network. Topological data analysis communities reveal gene-environment-brain subtypes of major depression in UK Biobank and multi-site cohorts. Biological Psychiatry. Published online 10 July 2026. doi:10.1016/j.biopsych.2026.06.030.

26. Madan A., Fowler J.C., Patriquin M.A., Salas R., Baldwin P.R., Velasquez K.M., et al. A novel approach to identifying a neuroimaging biomarker for patients with serious mental illness. Journal of Neuropsychiatry and Clinical Neurosciences. 2017;29(3):275-283. doi:10.1176/appi.neuropsych.16090174.

27. Mijangos M., Pacheco L., Bravetti A., Gonzalez-Garcia N., Padilla P., Velasco-Segura R. Persistent homology reveals robustness loss in inhaled substance abuse rs-fMRI networks. PLOS ONE. 2024;19(9):e0310165. doi:10.1371/journal.pone.0310165.

28. Ruan H., Chung M.K., Bruin W.B., Dzinalija N., Abe Y., Alonso P., et al.; ENIGMA OCD Working Group. Disrupted higher-order topology in OCD brain networks revealed by Hodge Laplacian - an ENIGMA study. bioRxiv. 2026:2026.03.04.709586. Preprint. doi:10.64898/2026.03.04.709586.

29. Bellec P., Chu C., Chouinard-Decorte F., Benhajali Y., Margulies D.S., Craddock R.C. The Neuro Bureau ADHD-200 Preprocessed repository. NeuroImage. 2017;144(Pt B):275-286. doi:10.1016/j.neuroimage.2016.06.034.

30. Di Martino A., Yan C.-G., Li Q., Denio E., Castellanos F.X., Alaerts K., et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Molecular Psychiatry. 2014;19(6):659-667. doi:10.1038/mp.2013.78.

31. Di Martino A., O' Connor D., Chen B., Alaerts K., Anderson J.S., Assaf M., et al. Enhancing studies of the connectome in autism using the autism brain imaging data exchange II. Scientific Data. 2017;4:170010. doi:10.1038/sdata.2017.10.

32. Alexander L.M., Escalera J., Ai L., Andreotti C., Febre K., Mangone A., et al. An open resource for transdiagnostic research in pediatric mental health and learning disorders. Scientific Data. 2017;4:170181. doi:10.1038/sdata.2017.181.

33. Casey B.J., Cannonier T., Conley M.I., Cohen A.O., Barch D.M., Heitzeg M.M., et al. The Adolescent Brain Cognitive Development (ABCD) study: imaging acquisition across 21 sites. Developmental Cognitive Neuroscience. 2018;32:43-54. doi:10.1016/j.dcn.2018.03.001.

34. Markiewicz C.J., Gorgolewski K.J., Feingold F., Blair R., Halchenko Y.O., Miller E., et al. The OpenNeuro resource for sharing of neuroscience data. eLife. 2021;10:e71774. doi:10.7554/eLife.71774.

35. Caputi L., Pidnebesna A., Hlinka J. Promises and pitfalls of topological data analysis for brain connectivity analysis. NeuroImage. 2021;238:118245. doi:10.1016/j.neuroimage.2021.118245.

36. Cohen-Steiner D., Edelsbrunner H., Harer J. Stability of persistence diagrams. Discrete & Computational Geometry. 2007;37(1):103-120. doi:10.1007/s00454-006-1276-5.

37. Bubenik P. Statistical topological data analysis using persistence landscapes. Journal of Machine Learning Research. 2015;16(1):77-102.

38. Balcı M.A., Akgüller Ö. Persistent homology of blood gene co-expression networks reveals reduced cycle structure in autism spectrum disorder: a multi-cohort analysis. Mammalian Genome. 2026;37(1):88. doi:10.1007/s00335-026-10259-8.

39. Zhao D., Zhou Z., Yan G., Yu D., Qi X. HOI-brain: a novel multi-channel transformers framework for brain disorder diagnosis by accurately extracting signed higher-order interactions from fMRI data. Medical Image Analysis. 2026;111:104009. doi:10.1016/j.media.2026.104009.

40. Kachura A., Chernyshev V., Kachan O., Levchenko E. Blurred magnitude homology of functional connectome for ASD diagnosis. Frontiers in Psychiatry. 2025;16:1677282. doi:10.3389/fpsyt.2025.1677282.

41. Mbaye A.A., Perea J.A., Tralie C.J., Goodwin M.S. Automated quantification of stereotypical motor movements in autism using persistent homology. bioRxiv. 2025:2025.09.03.674008. Preprint. doi:10.1101/2025.09.03.674008.

42. Bhaskar D., Zhang Y., Moore J., Gao F., Rieck B., Wolf G., et al. Neurospectrum: a geometric and topological deep learning framework for uncovering spatiotemporal signatures in neural activity. bioRxiv. 2025:2023.03.22.533807. Preprint. doi:10.1101/2023.03.22.533807.

43. Zhu Z., Li J., Yao L. BrainCHEF: cross-level hypergraph enhanced fusion model for brain networks. Computers in Biology and Medicine. 2025;195:110446. doi:10.1016/j.compbiomed.2025.110446.

44. Carr E., Rietschel M., Mors O., Henigsberg N., Aitchison K.J., Maier W., et al. Optimizing the prediction of depression remission: a longitudinal machine learning approach. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics. 2025;198(3):e33014. doi:10.1002/ajmg.b.33014.

45. Du H., Wei G.-W., Hou T. Multiscale topology in interactomic network: from transcriptome to antiaddiction drug repurposing. Briefings in Bioinformatics. 2024;25(2):bbae054. doi:10.1093/bib/bbae054.

46. Shnier D., Voineagu M.A., Voineagu I. Persistent homology analysis of brain transcriptome data in autism. Journal of the Royal Society Interface. 2019;16(158):20190531. doi:10.1098/rsif.2019.0531.

47. Palande S., Jose V., Zielinski B., Anderson J., Fletcher P.T., Wang B. Revisiting abnormalities in brain network architecture underlying autism using topology-inspired statistical inference. Brain Connectivity. 2019;9(1):13-21. doi:10.1089/brain.2018.0604.

48. Lee H., Kang H., Chung M.K., Lim S., Kim B.-N., Lee D.S. Integrated multimodal network approach to PET and MRI based on multidimensional persistent homology. Human Brain Mapping. 2017;38(3):1387-1402. doi:10.1002/hbm.23461.

49. Wong E., Palande S., Wang B., Zielinski B., Anderson J., Fletcher P.T. Kernel partial least squares regression for relating functional brain network topology to clinical measures of behavior. Proceedings of the IEEE International Symposium on Biomedical Imaging. 2016;2016:1303-1306. doi:10.1109/ISBI.2016.7493506.

50. Lee H., Kang H., Chung M.K., Kim B.-N., Lee D.S. Persistent brain network homology from the perspective of dendrogram. IEEE Transactions on Medical Imaging. 2012;31(12):2267-2277. doi:10.1109/TMI.2012.2219590.

51. Chung M.K., Bubenik P., Kim P.T. Persistence diagrams of cortical surface data. Information Processing in Medical Imaging. 2009;21:386-397. doi:10.1007/978-3-642-02498-6_32.

52. Robinson A., Turner K. Hypothesis testing for topological data analysis. Journal of Applied and Computational Topology. 2017;1(2):241-261. doi:10.1007/s41468-017-0008-7.

53. Anirudh R., Thiagarajan J.J., Kim I., Polonik W. Autism spectrum disorder classification using graph kernels on multidimensional time series. In: 2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI). 2017.

54. Cai T., Zhao G., Zang J., Zong C., Zhang Z., Xue C. Topological feature search method for multichannel EEG: application in ADHD classification. arXiv:2404.06676 [Preprint]. 2024. doi:10.48550/arXiv.2404.06676.

55. Chen Y., Qin D., Kuruoglu E.E. Topological and graph theoretical analysis of dynamic functional connectivity for autism spectrum disorder. arXiv:2410.16874 [Preprint]. 2024. doi:10.48550/arXiv.2410.16874.

56. Dakurah S. Topologically invariant permutation test. arXiv:2511.06153 [Preprint]. 2025. doi:10.48550/arXiv.2511.06153.