Chaos, Solitons and Fractals 208 (2026) 118234

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

journal homepage: www.elsevier.com/locate/chaos

Disentangling high-dimensional functional brain network structure in major depression via cortically inspired sparse encoding and contrastive learning

Muhammad Salman Kabir [a], Semen Kurkin [a,*], Rositsa Paunova [b,c], Drozdstoy Stoyanov [b,c], Alexander Hramov [a,d]

[a] Research Institute of Applied Artificial Intelligence and Digital Solution, Plekhanov Russian University of Economics, Russian Federation
[b] Department of Psychiatry and Research Institute, Medical University Plovdiv, Bulgaria
[c] Strategic Research and Innovation Program for the Development of MU-PLOVDIV–(SRIPD-MUP), European Union-NextGenerationEU, Bulgaria
[d] Laboratory for the Study of Tactile Communication, Pushkin State Russian Language Institute, Russian Federation

ARTICLE INFO

Keywords:
Major depressive disorder
rs-fMRI
Functional connectivity
Complex brain networks
Contrastive learning
Sparse distributed representations
High-dimensional systems

ABSTRACT

Functional brain networks are high-dimensional complex systems whose large-scale organization is disrupted in major depressive disorder (MDD), yet characterizing these disruptions remains challenging due to the extreme dimensionality of neuroimaging data and the substantial entanglement of connectivity patterns between patients and healthy individuals. Here we introduce an integrated computational framework that combines cortically inspired sparse encoding with contrastive manifold learning to disentangle the high-dimensional structure of functional brain networks in MDD from resting-state functional magnetic resonance imaging (rs-fMRI) data. The Spatial Pooler from Hierarchical Temporal Memory (HTM-SP) — a neocortex-derived algorithm that implements nonlinear competitive inhibition and Hebbian self-organization — serves as an unsupervised sparse encoder that reduces dimensionality while preserving the topological relationships within the network. Contrastive principal component analysis (cPCA) subsequently separates disorder-specific variance from shared neural architecture by identifying directions of maximal between-group divergence in the high-dimensional connectivity space. Evaluated on a cohort of 70 MDD patients and 70 matched healthy controls, the HTM-SP + cPCA pipeline achieved a classification accuracy of 86% and an F1-score of 87%, substantially outperforming conventional approaches. Critically, among the top 20 functional connections contributing most to the discriminative cPCA components, only 5 (25%) showed statistically significant group differences in traditional univariate testing, demonstrating that contrastive learning captures coordinated multivariate network alterations invisible to standard statistical approaches. The framework identified 129 discriminative functional connections, of which 34 exhibited significantly reduced connectivity in MDD, implicating distributed cross-network alterations spanning sensorimotor, dorsal attention, visual and frontal executive systems. The spatial distribution of these alterations — predominantly bridging separate functional modules rather than confined within individual subnetworks — suggests impaired integrative network organization consistent with systems-level models of depression as a disorder of large-scale network dysregulation. These results demonstrate that combining biologically grounded sparse

* Corresponding author.
E-mail address: kurkinsa@gmail.com (S. Kurkin).

https://doi.org/10.1016/j.chaos.2026.118234
Received 27 January 2026; Received in revised form 5 March 2026; Accepted 13 March 2026
Available online 19 March 2026

encoding with contrastive manifold separation provides both enhanced diagnostic sensitivity and interpretable insights into the emergent network-level reorganization underlying MDD.

1. Introduction

The past decade has witnessed a surge in the application of machine learning (ML) techniques across various healthcare domains. ML algorithms have demonstrated utility in diverse applications, including biomarker identification from medical images, accelerated drug discovery, automated disease diagnosis based on symptom profiles, risk factor assessment for neurological conditions and the identification of phenotypic subgroups to facilitate personalized medicine [1]. This progress is driven by the inherent capacity of ML algorithms to uncover latent structures and subtle relationships within complex, high-dimensional data. The ongoing exponential growth of healthcare data from diverse sources suggests a continued and accelerating adoption of ML methodologies within the healthcare sector.

One such example is the use of ML to diagnose major depressive disorder (MDD) [2]. MDD is a significant global health concern, that is projected to rank first by 2030 [3]. MDD has been linked to structural and functional brain alterations, immunological dysregulation, neurotransmitter imbalances, neurotrophic factor deficiencies and oxidative stress [4]. Despite extensive research demonstrating alterations in various biological systems, the underlying pathophysiology of MDD remains incompletely understood [4]. Given the complexity of MDD and the challenges in identifying reliable biomarkers, ML offers a promising approach for both screening and understanding the neural mechanisms of this disorder. Research in this area has evolved from conventional statistical approaches toward sophisticated deep learning models [5–15]. These studies employ diverse analytical techniques and classification algorithms that generally report promising performance metrics [16–19]. Researchers have utilized both custom-developed datasets and publicly available data repositories to train and validate these models, though standardization across studies remains challenging.

Although above mentioned studies reported impressively high diagnostic performance, most failed to address two critical challenges, researchers face in neuroimaging studies generally and MDD particularly: the curse of dimensionality [20–22] and the problem of functional connectivity entanglement [23–26]. The curse of dimensionality occurs when the number of features (variables) significantly exceeds the sample size, creating statistical and computational difficulties that can lead to overfitting, increased computational complexity and reduced model generalizability. This mathematical challenge is particularly prevalent in neuroimaging studies where thousands of voxels or connectivity measures are analyzed from relatively small patient cohorts [21]. Researchers employ two primary approaches to mitigate dimensionality issues. Feature engineering methods select relevant features while preserving their interpretability [27,28], which allows researchers to connect findings to underlying neural mechanisms [25,29,30]. In contrast, dimensionality reduction techniques mathematically transform the data into a lower-dimensional space, which improves computational efficiency but often sacrifices direct interpretability of the results [31].

The second critical challenge involves the entanglement of functional connectivity patterns. Due to the spectrum nature of neurological and psychiatric disorders, functional connectivity patterns in the treatment group frequently overlap substantially with those of healthy control subjects. This entanglement makes it difficult to clearly differentiate between pathological and normal brain function using conventional analytical approaches. Contrastive machine learning [32] offers a sophisticated framework specifically designed to address this entanglement challenge. Unlike traditional methods that analyze groups independently, contrastive ML approaches directly compare samples across different populations and systematically identify both shared characteristics and group-specific features. Contrastive principal component analysis (cPCA) [33] represents a particularly effective algorithm in this category. cPCA identifies linear combinations of features that maximize variance between groups rather than within groups, thereby highlighting the specific patterns that distinguish patients from controls. This approach enables detection of subtle yet clinically significant differences that conventional methods might overlook. Recent studies [23,25,34,35] demonstrate the effectiveness of contrastive ML in successfully disentangling functional connectivity patterns in neurological and psychiatric conditions.

In this study, we propose an integrated methodological approach that combines feature engineering and contrastive learning to address the above mentioned dual challenges of dimensionality curse and functional connectivity entanglement in neuroimaging-based MDD diagnostics. Our pipeline sequentially implements Spatial Pooler from Hierarchical Temporal Memory architecture (HTM-SP) and cPCA. The HTM-SP algorithm serves as our feature engineering algorithm and mitigates the curse of dimensionality while preserving critical topological relationships and interpretability of features. Following this feature engineering, we employ cPCA to disentangle the principal components derived from functional connectivity patterns. Our findings demonstrate that HTM-SP not only successfully addresses dimensionality curse but also provides meaningful insights into the neural mechanisms underlying MDD. Comparative analyses between our proposed HTM-SP+cPCA pipeline and conventional ML approaches demonstrate the superior diagnostic accuracy and interpretability of our integrated approach across multiple performance metrics.

2. Materials and methods

2.1. Participants and data collection

Resting-state functional magnetic resonance imaging (rs-fMRI) data were acquired from 70 patients diagnosed with major depressive disorder (MDD) and 70 healthy controls (HC). A schematic overview of the study design is provided in Fig. 1.

Table 1
Demographic and clinical characteristics of the considered groups.

	HC ($n = 70$)	MDD ($n = 70$)	Significance
Age (mean ± SD)	41.5 ± 11.03	41.04 ± 13.28	0.826[a]
Sex (M/F)	39/31	44/26	0.356[b]
Education (higher/secondary/primary)	35/35/0	33/34/3	0.401[b]
MADRS score (mean ± SD)	1.85 ± 2.12	29.51 ± 6.0	<0.001[a]

SD — Standard Deviation, MADRS — Montgomery–Åsberg Depression Rating Scale.
[a] Two-sample Kolmogorov–Smirnov nonparametric test.
[b] χ^2 — test.

All participants underwent a comprehensive clinical evaluation conducted by experienced psychiatrists. Diagnostic status was confirmed using the Mini International Neuropsychiatric Interview (MINI) [36]. For the MDD group, depressive symptom severity was quantified using the Montgomery–Åsberg Depression Rating Scale (MADRS) [37,38]. To ensure sample specificity, individuals with a history of comorbid psychiatric conditions, autoimmune or neurological disorders, significant head trauma, or MRI-incompatible metal implants were excluded from both groups.

Demographic characteristics, including mean age, sex distribution and education level, were comparable between the MDD and HC groups, with no statistically significant differences observed. In line with the clinical distinction between groups, MADRS scores were significantly higher in the MDD cohort compared to the HC group. A detailed summary of participant characteristics is presented in Table 1.

2.2. Magnetic resonance imaging acquisition and preprocessing

Imaging protocol. All scans were acquired on a 3T GE Discovery 750w MRI system. The acquisition protocol consisted of:

1. A high-resolution T1-weighted anatomical scan (Sagittal 3D T1) with the following parameters: slice thickness = 1 mm, matrix = 256 × 256, repetition time (TR) = 7.2 s, echo time (TE) = 2.3 s, flip angle = 12°, field of view (FOV) = 24.
2. A resting-state functional MRI (rs-fMRI) scan [39] with the following parameters: slice thickness = 3 mm, matrix = 64 × 64, TR = 2000 ms, TE = 30 ms, flip angle = 90°, 192 volumes.

Data preprocessing. Rs-fMRI data were preprocessed using SPM12 software [40] following a standardized pipeline [15]. The steps included:

- Realignment. Functional images for each participant were realigned to correct for head motion. This step used the following parameters: quality = 0.9, separation = 4, no smoothing, 2nd-degree B-spline interpolation, no wrapping, 12 × 12 basis function, regularization = 1 with medium factor, without Jacobian deformations, 5 iterations, average Taylor expansion point.
- Coregistration. The mean functional image was coregistered to the high-resolution anatomical scan using a Normalized Mutual Information algorithm.
- Normalization. The coregistered images were spatially normalized into standard Montreal Neurological Institute (MNI) space. Normalization parameters were: bias regularization = 0.0001, bias FWHM cutoff = 60 mm, affine regularization to the ICBM European brain template, warping regularization, no smoothing, sampling distance = 3.
- Motion correction. Subject-level motion parameters estimated during realignment were regressed out as a standard confound.

The final output of this preprocessing pipeline was a set of voxel-level blood oxygen level-dependent (BOLD) time series for subsequent analysis.

2.3. Construction of functional brain networks

Brain functional networks were reconstructed from the preprocessed BOLD signals using the following pipeline (see also Fig. 1B).

Network Nodes and Edge Definition. The entire brain volume was parcellated into 165 distinct regions (network nodes) based on the Automated Anatomical Labeling Atlas version 3 (AAL3) [41]. For each node, the average BOLD time series, $x_i(t)$, was extracted. After detrending, pairwise functional connectivity between all nodes was quantified by calculating Pearson's correlation coefficients for the averaged time series of each region pair [14].

Thresholding and Adjacency Matrix. To retain only statistically robust connections, a significance threshold of $p < 0.05$ (uncorrected) was applied to each correlation coefficient. The result was a symmetric 165 × 165 adjacency matrix for each participant, representing the brain's functional network, where matrix elements (edges) correspond to significant inter-regional correlations.

Feature Vector Extraction. For subsequent group-level analysis, each individual's weighted adjacency matrix was vectorized by extracting its unique off-diagonal elements. This yielded a feature vector of dimension $N(N-1)/2 = 13530$ (where $N = 165$), capturing the full pattern of pairwise functional connections. In total, 70 feature vectors were generated for HC group and 70 for MDD group.

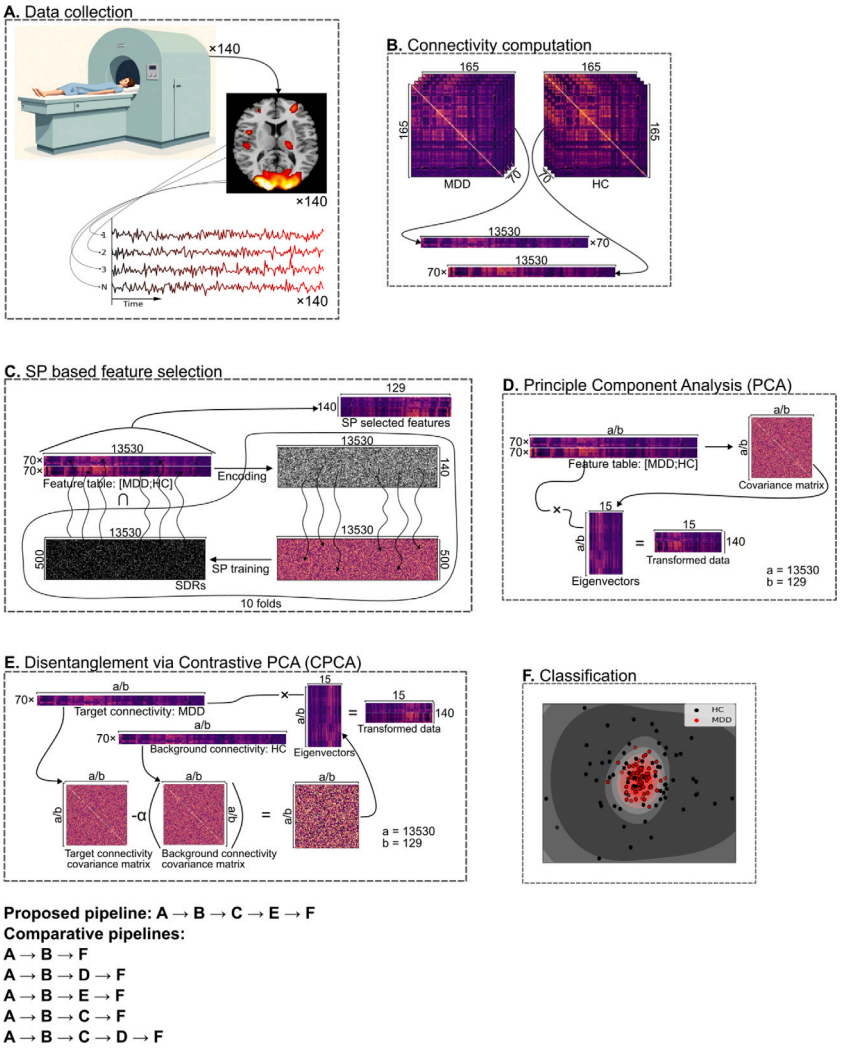


Fig. 1. Pipeline of fMRI data analysis. A. High resolution structural and resting-state functional MR scanning on 3T MRI system and extraction of BOLD signals, B. Subject-wise computation of correlation connectivity matrices and further reduction of symmetrical correlation matrices to correlation vectors containing 13 530 unique connections per vector, C. SP-based feature selection; correlation vectors are first passed through encoder to generate binary vectors and then through SP mechanism to select 129 features, D. Performing dimensional reduction using PCA, the input feature space may either be *"a"* (i.e, 13 530 features, without feature selection) or *"b"* (i.e, 129 features, selected features using SP), E. Disentangled principal components. The input feature space may either be *"a"* (i.e, 13 530 features, without feature selection) or *"b"* (i.e, 129 features, selected features using SP). Covariance matrices are computed and subtracted to get eigenvectors which represent the principal components specific to MDD. The resulting feature space consists of 15 features, F. Classification using Gaussian process classifier with an rbf kernel. The input to the classifier may either be 13 530 (original features), 129 (SP selected features), or 15 (features after dimensionality reduction).

2.4. Feature selection via the spatial pooler

The dimensionality of our dataset (13 530 features per participant across 140 subjects) presents a classic high-dimensionality challenge. To address this, we employed the Spatial Pooler (SP) [42] — a core component of Hierarchical Temporal Memory (HTM) — as an unsupervised feature selection algorithm. HTM is a biologically inspired machine learning framework that models the structural and functional properties of the mammalian neocortex, the brain region responsible for higher-order cognition [43]. Evidence suggests the neocortex employs a common set of algorithms across diverse cognitive tasks [44], a principle formalized by HTM. The HTM architecture consists of four interconnected components: an encoder, a spatial pooler, temporal memory and a classifier (Fig. 2) (Refer to Appendix A for a detailed description of HTM architecture).

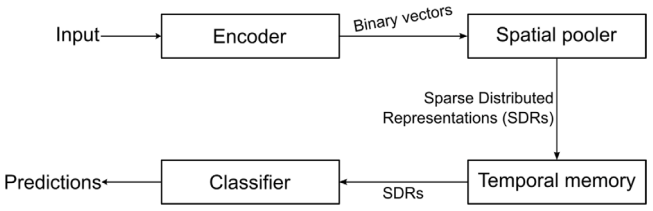


Fig. 2. HTM architecture. An HTM system comprises of an encoder, spatial pooler (SP), temporal memory and a classifier.

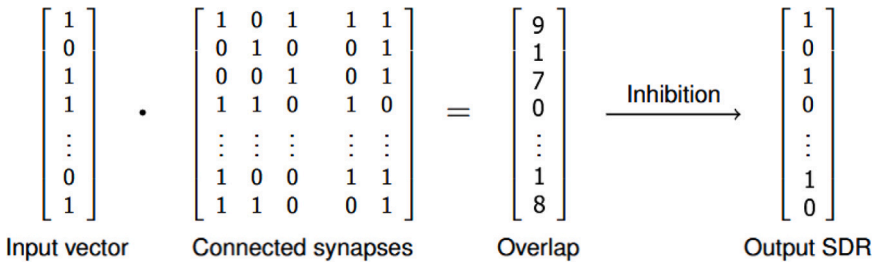


Fig. 3. SP mechanism. The encoded input vector is dot multiplied by the connected synapses to produce overlap values, which then undergo inhibition to generate the final SDR output.

A cornerstone of HTM's effectiveness is its use of sparse distributed representations (SDRs), which are observed in multiple cortical regions [45,46]. The SP specifically encodes input data streams into stable SDRs by incorporating neurobiological principles such as competitive Hebbian learning, homeostasis and activity-dependent structural plasticity [42].

Feature Selection Principle. Within this framework, the SP performs feature selection by transforming dense or sparse binary input vectors into SDRs. The SP contains columns, analogous to cortical mini-columns, each equipped with synapses whose connection strengths are defined by permanence values (initialized randomly). During processing, active input bits drive synaptic activity. Columns then compete through a local inhibition mechanism grounded in Hebbian learning [47,48]. Columns receiving the strongest input overlap become active, while others are suppressed, resulting in a sparse, stable pattern of active columns (the SDR). This process effectively selects a consistent subset of salient features from the high-dimensional input space.

Fig. 3 illustrates the transformation. A binary input vector is multiplied by a binary connectivity matrix, yielding a vector of overlap scores. An inhibition step then selects the columns with the highest overlaps (subject to a sparsity constraint) to form the output SDR. The indices of the active columns correspond directly to the selected feature subset.

In this study, the encoder binarized each continuous feature (x_i) into binary value (b_i) as,

$$b_i = \mathbb{I}(x_i > 0)$$

where $\mathbb{I}(\cdot)$ denotes the indicator function, which equals 1 if the condition holds and 0 otherwise. This encoding scheme was deliberately chosen to reflect the focus of the present study on positive functional correlations between brain regions, which represent coherent co-activation patterns and constitute a well-established and neurobiologically interpretable form of functional connectivity in resting-state fMRI research. It is worth noting that alternative encoding schemes are possible and may be more appropriate depending on the research objective. For instance, a symmetric threshold centered around zero could encode both strong positive and strong negative correlations as active bits, which may be preferable in studies where anti-correlated networks are of primary interest. The choice of encoding scheme therefore represents an important design decision that should be guided by the neurobiological hypotheses under investigation.

Validation and Implementation. Preliminary investigations confirm the SP's efficacy as a feature selector, performing comparably to conventional algorithms [49]. The specific hyperparameters used in this study are detailed in Table 2. We implemented a global inhibition scheme, set the potential radius to the full input size (13 530) and configured 100% potential synapse connectivity, meaning each column was initially connected to the entire input space. The output sparsity was set to 3%.

To ensure robust and stable feature selection, we employed a 10-fold cross-validation strategy [50]. In each fold, nine subsets were used to train the SP. After training, synapses with permanence values at or above the connection threshold (0.50) were identified; their corresponding input indices constituted the selected features for that fold. The final, consensus set of selected features was defined as the intersection of features selected across all ten folds (see Fig. 1C). This approach mitigates selection bias from any single data partition.

It is important to note that the folds used for SP training were identical to those used for classification in Section 2.6. That was, in each fold, the SP was trained on the same nine subsets used to train the classifier and both were evaluated on the same held-out subset. This ensured that no information from the test set influenced feature selection which prevents data leakage and preserves the integrity of the evaluation.

Table 2
Hyperparameters setting to train SP.

Hyperparameter	Value
Input size	13 530
Potential radius	13 530
Columns size	500
Boost strength	100
Output sparsity (%)	3
Percentage of potential synapses	100
Stimulus threshold	10
Increase in the permanence of active synapses	0.14
Decrease in the permanence of inactive synapses	0.02
Permanence threshold for a synapse to be consider connected (connected permanence)	0.50

2.4.1. HTM-SP and nonlinear sparse coding

It is worth to note that HTM-SP introduces properties that extend beyond simple linear feature selection and can be meaningfully interpreted within the framework of complex adaptive systems. Specifically, the competitive inhibition mechanism underlying the SP implements a winner-take-all dynamic, whereby columns compete for activation based on their overlap with the input and only those receiving the strongest drive become active. This is an inherently nonlinear operation. Furthermore, the Hebbian learning rule governing synaptic permanence values enables the SP to adaptively reorganize its connectivity structure in response to statistical regularities in the input data, without any global supervisory signal. The resulting SDRs emerge from purely local competition which is a hallmark of self-organizing complex adaptive systems.

2.5. Disentangling depression specific characteristics

Functional connectivity patterns in individuals with neuro-diversity exhibit a complex relationship with those observed in the neuro-typical population. This relationship can be characterized as an entanglement, where the neural patterns of affected individuals are not entirely distinct from, but rather intertwined with those of neurotypical. Multiple studies have identified this entanglement across various neurological [25,34] and psychiatric [23,35] conditions. This entanglement poses significant challenges to not only understand the neural mechanisms but also to distinguish neuro-diverse individuals from neuro-typical ones. This complexity requires more sophisticated approaches to analyzing and interpreting neuroimaging results in psychiatric research.

We worked under a similar assumption that functional connectivity of individuals with MDD exhibits entanglement with HC. We employed contrastive principal component analysis (cPCA) [33] to disentangle this entanglement (Fig. 1E). This method extends traditional PCA by explicitly accounting for differences between two groups. In our study, we first computed the covariance matrices for both the MDD and the HC functional connectivity vectors. The HC covariance matrix was then subtracted from the MDD covariance matrix to yield a disorder-specific covariance matrix as:

$$C = C_{MDD} - \alpha C_{HC}.$$

Contrastive PCA operates on this difference matrix, extracting eigenvectors that represent the principal components specific to the disorder. These components capture the directions of maximal variance that are unique to the MDD group when compared to the HC group. The value for $\alpha = 2.14$ was selected following a heuristic approach so that resulting subspaces form well-separated clusters.

In our analysis, we selected the top 15 principal components based on their explained variance (see Appendix B). This selection criterion ensures that we retain the most variance from original data. Selected principal components represent the most salient aspects of depression-specific functional connectivity and thus potentially offer insights into the neural substrates of MDD.

2.5.1. How cPCA achieves variance disentanglement

Geometrically, the functional connectivity patterns of MDD and HC subjects can be understood as point clouds residing on overlapping manifolds in a high-dimensional feature space. Standard PCA identifies directions of maximum total variance across the combined dataset, which will predominantly reflect shared neural architecture, i.e., the large common variance that is present in both groups, rather than disorder-specific structure. cPCA addresses this limitation by explicitly seeking directions along which the variance of the MDD is maximally enriched relative to the HC. Formally, this is achieved by computing the difference between the MDD and HC covariance matrices and extracting the leading eigenvectors of this difference matrix. Geometrically, this operation rotates the coordinate system of the feature space to align with the directions that are most distinctive to the MDD manifold. This effectively disentangles the disorder-specific subspace from the shared subspace. The resulting contrastive components therefore capture the differential geometry of the two manifolds and revealing structure that is invisible to standard variance-maximization approaches.

Table 3
Inputs to classifier and the corresponding pipelines (see also Fig. 1).

Input to classifier	Pipeline
Original connectivity	A-B-F
Entangled principal components of original connectivity	A-B-D-F
Disentangled principal components of original connectivity	A-B-E-F
Selected features using HTM-SP	A-B-C-F
Entangled principal components of selected features	A-B-C-D-F
Disentangled principal components of selected features	A-B-C-E-F

2.6. Classification

After deriving depression-specific principal components through cPCA, we employed Gaussian Process classifier (with an rbf kernel) [51] to distinguish MDD patients from HC (Fig. 1F). Also, we used SVM classifier for the control. To perform ablation studies, we trained our classifier on six different inputs (obtained in comparative pipelines). Table 3 details these inputs along with comparative pipelines. We encapsulated the whole classification process into a 10-fold cross-validation scheme. As described in Section 2.4, the SP feature selection and classification shared the same cross-validation folds which ensured a leakage-free evaluation pipeline. Accuracy, recall, precision and F1-score [52] were used to measure the classification performance.

2.7. Mapping contrastive components to original features

To understand which connections contributed most in driving contrastive principal components, we mapped contrastive components back to original features. The relationship between original features and contrastive principal components was characterized through the loading matrix V, where element v_{ij} represents the contribution of feature j to component i. In contrastive PCA, these loadings quantify how strongly each original feature participates in the variance patterns. Positive loadings indicate that a feature covaries positively with the component direction, while negative loadings indicate inverse relationships. However, the magnitude of the loading, rather than its sign, determines the feature's importance, as both strongly positive and strongly negative values indicate substantial contributions. To quantify overall feature importance, we computed:

$$I_j = \frac{1}{k} \sum_{i=1}^{k} |v_{ij}|$$

where I_j is the importance score for feature j, k is the number of selected components ($k = 15$) and the sum is taken over all contrastive components. This metric captures the average magnitude of each feature's contribution across the contrastive space, independent of directional effects. Features were ranked by their importance scores to identify their contribution to contrastive principal components.

3. Results

The proposed pipeline, integrating HTM-SP for feature engineering and cPCA for contrastive disentanglement, was evaluated against conventional classification approaches using rs-fMRI data from 70 MDD patients and 70 healthy controls.

3.1. Classification performance

Classification using the original high-dimensional functional connectivity vectors yielded near-chance performance, with an accuracy of approximately 50%, confirming the substantial entanglement between MDD and HC connectivity patterns (Fig. 4; see also Table C.5 in Appendix C for detailed values of classification performance).

When HTM-SP alone was applied for feature selection, classification accuracy remained modest (55%), indicating that dimensionality reduction alone was insufficient to disentangle group-specific signals. Applying cPCA directly to the original connectivity features improved the F1-score to 76%, highlighting the utility of contrastive learning in isolating disorder-relevant variance.

The synergistic combination of HTM-SP and cPCA produced the highest discriminative performance. Using the disentangled principal components derived from HTM-SP-selected features, a Gaussian process classifier achieved a mean accuracy of 86% and an F1-score of 87%. This represented a significant improvement (Friedman test, $p < 0.05$, post-hoc Wilcoxon signed-rank with Holm–Bonferroni correction) over all other tested configurations with the exception of A-B-E-F pipeline (Table 3, Fig. 4), underscoring the complementary roles of feature engineering and contrastive learning in addressing both dimensionality and entanglement.

Using an SVM classifier instead of a GPC in the final stage yields similar results (Table C.6 in Appendix C), which further confirms the previous conclusions.

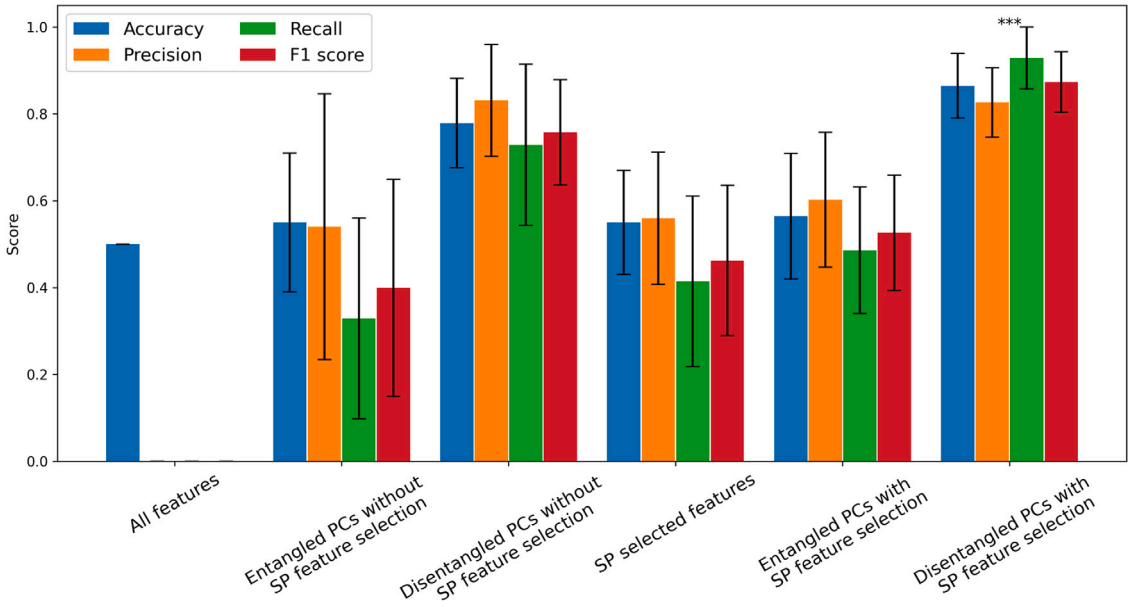


Fig. 4. 10-fold cross-validation scores comparison. Here, bars represent mean and whiskers represent standard deviations over the folds. In the first scenario *"All features"*, classifier classified all samples into one class, which resulted in 0 recall, undefined precision and 0 F1 score. *** indicates statistical significance.

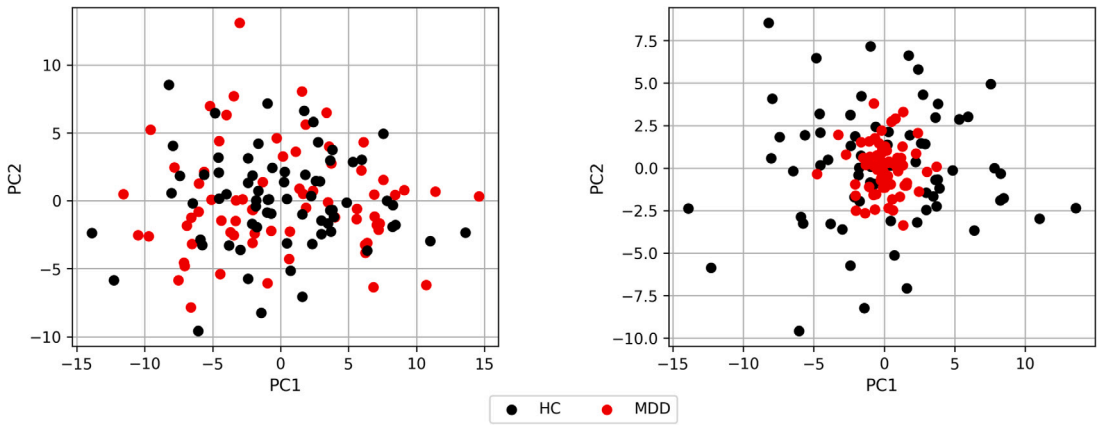


Fig. 5. Disentangled PCs. Scatter plots of top two principal components, that contributed most in MDD classification, before and after disentanglement; points correspond to the subjects.

3.2. Disentanglement of functional connectivity patterns

Visualization of the top two principal components before and after cPCA application clearly illustrates the effect of contrastive learning (Fig. 5). Prior to disentanglement, MDD and HC samples overlapped substantially in the principal component space. Following cPCA, the groups became more separable, with the derived components capturing variance specific to MDD. The first 15 contrastive principal components, selected based on explained variance, encapsulated the most salient depression-specific connectivity signatures.

3.3. Feature selection and neurobiological interpretation

HTM-SP consistently produced sparse distributed representations (SDRs) with a fixed output sparsity of 3%. From the original 13,530 possible connections, the consensus feature set across cross-validation folds comprised 129 stable and informative functional connections.

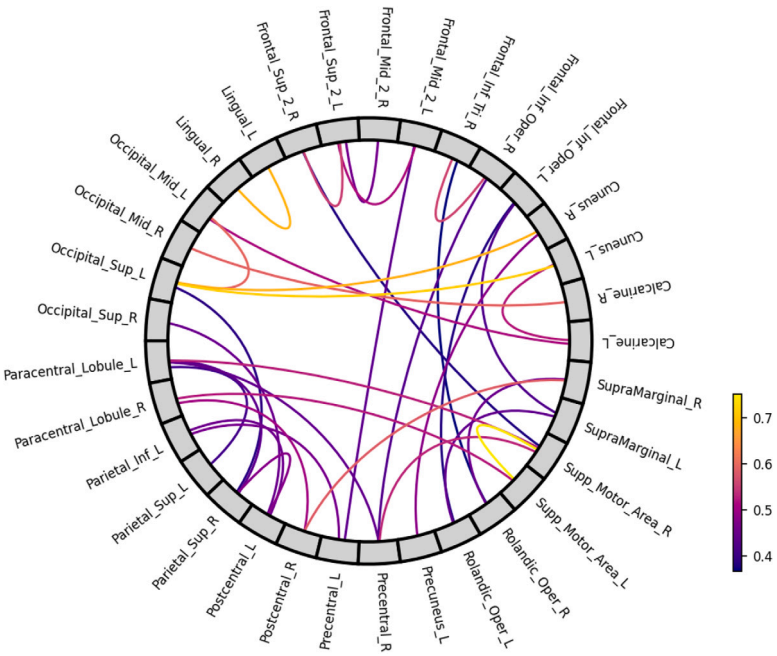


Fig. 6. Hypoconnected edges in MDD. We observe reduced connectivity in sensorimotor, dorsal attention, visual and frontal executive networks.

Among these, 34 connections demonstrated significantly lower correlation strength (Welch's t-test, FDR $q < 0.05$) in the MDD group compared to HC (Fig. 6, Appendix D). These hypoconnected edges involved distributed regions spanning multiple large-scale networks, including:

- Sensorimotor regions: Precentral L – Frontal Mid 2 L, Precentral R – Supp Motor Area R, Postcentral L – Parietal Inf L, etc.
- Dorsal attention and parietal circuits: Parietal Sup L – Paracentral Lobule L, Occipital Sup L – Parietal Sup R, Postcentral L – Parietal Sup R, etc.
- Visual cortex: Calcarine L – Cuneus L, Lingual L – Lingual R, Occipital Sup L – Occipital Mid L, etc.
- Frontal executive and motor planning nodes: Frontal Sup 2 L – Frontal Mid 2 L, Supp Motor Area L – Supp Motor Area R, etc.

The spatial distribution of these altered connections reinforces the conceptualization of MDD as a disorder of distributed network dysregulation, rather than focal pathology.

3.4. Features contribution to cPCA

Table 4 lists top 20 features that contributed largely to contrastive principal components. The cPCA loading heatmap (Fig. E.7 in Appendix E) revealed a sparse structure with most features showing weak loadings and specific feature-component combinations displaying strong contributions. This pattern indicates that each contrastive component is driven by a distinct subset of connectivity features. Components varied in their loading patterns: some (e.g., PC1) showed diffuse, weak contributions across many features, while others (e.g., PC2, PC3, PC7) exhibited concentrated, strong loadings in specific feature subsets. The presence of both positive (red) and negative (blue) loadings within components indicates bidirectional connectivity alterations in MDD, with some connections strengthened and others weakened relative to controls. Feature-wise analysis revealed that certain brain connections consistently contributed across multiple components, while others showed minimal discriminative value.

A striking observation from this analysis is the limited overlap between connections identified as highly discriminative by cPCA and those flagged as statistically significant by conventional univariate testing. Of the top 20 connections ranked by their contribution to contrastive principal components (Table 4), only 5 (25%) demonstrated statistically significant group differences using Welch's t-test with FDR correction. This finding underscores a fundamental limitation of traditional statistical approaches in neuroimaging: univariate tests evaluate each connection independently and require effect sizes large enough to survive multiple comparison correction, potentially missing connections that, while individually modest, contribute substantially to multivariate discrimination between groups. In contrast, cPCA directly optimizes for between-group separability, capturing the coordinated patterns of connectivity alterations that collectively distinguish MDD from healthy controls. This enhanced sensitivity to diagnostically relevant but statistically subtle features represents a key advantage of contrastive machine learning approaches for psychiatric neuroimaging.

Table 4
Top 20 connections contributing to cPCA. Connections showing statistically significant hypoconnectivity in MDD compared to HC (Welch's t-test, FDR $q < 0.05$) are marked in bold. Notably, only 5 of the 20 most discriminative connections (25%) were identified by traditional statistical testing, highlighting the enhanced sensitivity of contrastive learning in detecting disorder-relevant connectivity patterns.

Rank	Connection
1	Fusiform_L – Cerebellum_6_L
2	Precentral_R – Postcentral_R
3	Cingulate_Post_L – Cingulate_Post_R
4	**Frontal_Sup_2_R – Supp_Motor_Area_R**
5	Temporal_Pole_Sup_R – Temporal_Mid_R
6	**Cuneus_R – Precuneus_L**
7	Temporal_Mid_L – Cerebellum_6_R
8	**Frontal_Inf_Oper_L – Rolandic_Oper_L**
9	Thal_VPL_R – Thal_PuA_L
10	Lingual_L – Occipital_Mid_L
11	Occipital_Inf_R – Temporal_Inf_R
12	Cerebellum_4_5_L – Cerebellum_4_5_R
13	Insula_L – Putamen_L
14	Frontal_Sup_Medial_R – ACC_pre_L
15	**Calcarine_L – Cuneus_L**
16	Calcarine_L – Lingual_R
17	SupraMarginal_L – SupraMarginal_R
18	Temporal_Inf_L – Temporal_Inf_R
19	**Lingual_L – Lingual_R**
20	Caudate_L – Temporal_Mid_L

3.5. Robustness and stability

The use of 10-fold cross-validation for both feature selection (HTM-SP) and classification ensured that results were not biased by particular data partitions. The consensus feature set, defined as the intersection of features selected across all folds, demonstrated high stability and classification performance showed reduced variance across folds (Fig. 4), indicating the robustness of the proposed pipeline.

4. Discussion

This study introduced an integrated machine learning pipeline combining feature engineering via Spatial Pooler component of Hierarchical Temporal Memory (HTM-SP) architecture and contrastive learning via contrastive principal component analysis (cPCA) to address two persistent challenges in neuroimaging-based diagnostics of major depressive disorder (MDD): the curse of dimensionality and the entanglement of functional connectivity patterns between patients and healthy controls. Our results demonstrate that this hybrid approach not only significantly improves classification performance but also yields interpretable, neurobiologically meaningful insights into the network-level alterations associated with MDD.

4.1. Addressing dimensionality and entanglement simultaneously

Neuroimaging datasets are characteristically high-dimensional, often comprising thousands of features derived from relatively small cohorts. This imbalance can lead to overfitting, reduced generalizability and hindered interpretability — a challenge well-documented in the literature [20–22]. While conventional dimensionality reduction techniques such as PCA can improve computational efficiency, they often obscure the biological interpretability of retained features [31]. In contrast, HTM-SP operates as an unsupervised feature selector that preserves the topological and relational integrity of functional connections, thereby maintaining a link to underlying neural circuitry. Our findings confirm that HTM-SP effectively reduces feature space dimensionality while retaining discriminative power, a result consistent with prior investigations into its utility as a nonparametric feature selection tool [49].

However, feature selection alone proved insufficient to fully differentiate MDD from HC, yielding only modest gains in classification accuracy. This reflects a second, more subtle challenge: the substantial overlap in functional connectivity patterns between clinical and control groups, a phenomenon described as "entanglement" [23–25]. Such entanglement complicates the identification of disease-specific signatures, particularly in spectrum-like disorders such as MDD. By applying cPCA after HTM-SP-based feature selection, we explicitly modeled between-group differences, effectively disentangling disorder-relevant variance from shared neural architecture. The resultant improvement in classification performance — from approximately 50% accuracy using raw connectivity to 85%–90% using the proposed pipeline — underscores the complementary strengths of the two methods and highlights the necessity of addressing both dimensionality and entanglement in tandem.

4.2. Neurobiological interpretation of selected features

The proposed HTM-SP + cPCA framework enabled the identification of a compact and reproducible subset of functional connectivity features associated with MDD. From an initial set of 13,530 pairwise functional connections, the Spatial Pooler consistently selected 129 connections across cross-validation folds, suggesting that functional connectivity alterations in MDD are not uniformly distributed across the brain but are preferentially associated with specific functional subnetworks. The consistency of feature selection across folds indicates that these connectivity patterns are stable and unlikely to reflect sampling-related variability. Such stability is an important prerequisite for the identification of neuroimaging markers with potential clinical relevance, particularly in the context of psychiatric disorders characterized by heterogeneous and overlapping neural signatures [53,54].

Among the selected connections, 34 demonstrated significantly reduced functional connectivity in the MDD group relative to healthy controls. The predominance of hypoconnectivity is consistent with prior neuroimaging studies suggesting that MDD is associated with alterations in large-scale network integration and inter-regional communication, rather than isolated focal abnormalities [55,56]. Reduced functional connectivity at the network level has been associated with impairments in cognitive abilities, information processing efficiency and adaptive responses to environmental demands in individuals with MDD [57].

Within this subset, altered connectivity involving sensorimotor regions, including the precentral gyrus, postcentral gyrus and supplementary motor area, was observed. Dysconnectivity within these regions has been reported in previous studies and may be related to disturbances in motor planning and psychomotor function, which are commonly observed in depressive disorders [58,59]. Reduced coupling within sensorimotor circuits may therefore represent a neural correlate of psychomotor symptoms in MDD, although the present findings do not permit direct inference regarding causality or symptom specificity.

Disrupted functional connectivity within dorsal attention and parietal networks, including superior parietal, postcentral and paracentral regions, is consistent with prior reports of attentional control disturbances in MDD. Alterations in these networks have been associated with impairments in goal-directed attention, reduced cognitive abilities, and a tendency toward internally focused cognitive states in affected individuals [60,61]. Such network-level changes may contribute to difficulties in attentional disengagement from internally generated negative thoughts, a cognitive feature frequently observed in depression, although the present findings do not allow direct inference regarding symptom-specific mechanisms.

Hypoconnectivity within visual networks, involving calcarine, cuneus, lingual and occipital regions, suggests the presence of altered sensory network organization in MDD. Abnormal functional connectivity within visual and sensory processing regions has been reported in previous neuroimaging studies of depression and has been associated with reduced sensitivity to external stimuli, anhedonia and diminished affective responsiveness [62,63]. These findings indicate that alterations in sensory processing networks may represent an additional component of the broader network dysfunction observed in depressive disorders.

Altered connectivity involving frontal executive and medial prefrontal regions, including interactions among superior and middle frontal gyri and supplementary motor areas, is indicative of disturbances within networks supporting executive control and top-down regulatory processes. These regions constitute key components of the executive control network and play an essential role in cognitive control, emotion regulation and modulation of limbic and sensory systems. Reduced functional integration within these circuits has been consistently associated with rumination, impaired decision-making and difficulties in regulating negative affect in individuals with MDD [55,64].

Mapping contrastive principal components back to the original functional connectivity features demonstrated that depression-specific variance is distributed across multiple components, each characterized by distinct patterns of network involvement. The sparse and component-specific loading structure observed in the cPCA analysis suggests that no single connection or network exclusively accounts for the disorder-related variance. Instead, the findings indicate that MDD-related functional reorganization reflects coordinated alterations across multiple interacting brain networks. This multidimensional organization is consistent with contemporary models emphasizing the clinical and neurobiological heterogeneity of depression and the presence of multiple convergent neural pathways underlying the disorder [65,66].

Collectively, these findings indicate that the proposed HTM-SP + cPCA pipeline captures patterns of functional dysconnectivity that are both neurobiologically plausible and potentially relevant from a clinical perspective. By preserving interpretability at the level of individual connections and large-scale networks, the framework provides a balance between predictive performance and explanatory value, supporting its potential applicability in interpretable machine learning approaches for psychiatric neuroimaging.

4.3. System-level interpretation

Beyond individual edge-level findings, the spatial distribution of the selected connections suggests that depression-related alterations are not confined to a single subnetwork but instead reflect distributed dysregulation across large-scale brain systems. The observed reductions in connectivity involve sensorimotor, attentional, visual, and frontal executive areas, with numerous connections bridging separate functional systems. This pattern is more consistent with impaired cross-network integration than with isolated within-module disruption and aligns with contemporary systems-level models conceptualizing MDD as a disorder of large-scale network dysregulation rather than focal pathology [53,55]. From a psychiatric perspective, reduced coupling within sensorimotor and frontal circuits may relate to psychomotor slowing, while weakened interactions between attentional and executive systems may contribute to cognitive control deficits frequently observed in depression. Importantly, large-scale frameworks of psychopathology describe depression as involving altered balance among control, salience, and self-referential networks, reflecting compromised integrative coordination across distributed systems.

From a complex-systems perspective, these findings bear on several organizational principles. First, the observation that the majority of the 34 significantly hypoconnected edges bridge distinct functional modules, rather than being confined within a single module, suggests a weakening of inter-modular integration. In network-theoretic terms, such inter-modular connections serve as topological bridges that sustain global communication efficiency and support the coordination of functionally specialized subsystems. Their selective disruption may lead to increased modular segregation and decreased global efficiency, a configuration that has been associated with impaired cognitive flexibility and reduced adaptive capacity in complex network models. Second, several of the identified connections involve regions that occupy hub positions within the canonical functional connectome, notably supplementary motor area, superior frontal gyrus, and precuneus. Hubs are known to be disproportionately metabolically costly and vulnerable to pathological processes, and their disruption can produce cascading effects that propagate across multiple connected modules. The convergence of MDD-related hypoconnectivity on hub-adjacent regions raises the possibility that the observed network alterations reflect a partial hub disruption, potentially fragmenting the integrative backbone of the functional connectome.

Third, the distributed, multi-network character of the altered connections is consistent with the notion of emergent network reorganization. Rather than a simple loss of connectivity, the pattern suggests that MDD involves a qualitative shift in the functional architecture of the brain, where the balance between segregation and integration is reconfigured. Such a shift is characteristic of phase-like transitions in complex adaptive systems, where gradual parameter changes can produce discontinuous reorganization of macroscopic system behavior. Although the present analysis does not include explicit graph-theoretical measures such as modularity, global efficiency, or hub centrality, the spatial and topological profile of the identified alterations provides converging evidence for a systems-level interpretation. Future work incorporating quantitative network topology analysis, as well as dynamic functional connectivity approaches capable of capturing time-varying fluctuations in network organization, will be essential to clarify whether MDD-related alterations reflect reduced global efficiency, hub disruption, modular reconfiguration, or changes in the dynamical repertoire of brain network states.

4.4. Contrastive learning reveals discriminative patterns missed by traditional statistics

A particularly noteworthy finding of this study concerns the discrepancy between features identified as statistically significant by conventional univariate testing and those identified as highly discriminative by contrastive learning. Among the top 20 functional connections that contributed most strongly to the cPCA components differentiating MDD from HC, only 5 (25%) exhibited statistically significant group differences in traditional Welch's t-tests (Table 4). This observation has important methodological and clinical implications.

Traditional statistical approaches in neuroimaging typically evaluate each feature independently and require effect sizes sufficient to survive stringent multiple comparison corrections. While these methods control false positive rates, they may systematically overlook features that are individually subtle but collectively informative. In high-dimensional neuroimaging data, disease-related alterations are often distributed across many connections, each contributing modestly to the overall pathological pattern. Univariate tests, by design, cannot capture these coordinated multivariate effects.

Contrastive learning, exemplified here by cPCA, addresses this limitation by directly optimizing for between-group discrimination rather than within-feature significance. The resulting components capture directions of maximal variance difference, effectively aggregating information across multiple features. Connections that individually fail to reach statistical significance may nonetheless carry diagnostic information when considered jointly with other features in the contrastive framework.

This finding suggests that relying solely on traditional statistical testing may provide an incomplete picture of disease-related functional connectivity alterations. The contrastive approach offers a complementary perspective, potentially uncovering biomarkers and mechanistic insights that conventional methods overlook. For clinical applications, this enhanced sensitivity could improve early detection and differential diagnosis, particularly for spectrum disorders like MDD where pathological patterns overlap substantially with typical variation.

4.5. Implications for interpretable machine learning in psychiatry

The current emphasis in neuroimaging machine learning has often prioritized predictive accuracy at the expense of interpretability, resulting in "black-box" models with limited clinical translatability. Our work aligns with a growing call for methods that integrate domain knowledge and provide explainable outputs [1]. HTM-SP, inspired by neocortical organization and cPCA, designed to extract group-differential patterns, together offer a framework that is both data-driven and biologically constrained. This dual emphasis not only enhances diagnostic performance but also generates hypotheses regarding the neural mechanisms of MDD — a crucial step toward personalized and mechanistically informed interventions.

4.6. Limitations and future directions

Several limitations of this study warrant consideration. First, while the sample size is comparable to many neuroimaging studies, larger and multi-site cohorts would help validate the generalizability of the selected features and classification model. Second, the present analysis focused on resting-state fMRI; integrating multimodal data (e.g., task-based paradigms, structural MRI, EEG, or clinical covariates) could further enrich the feature set and improve differential diagnosis.

Future work should explore the applicability of this pipeline to other psychiatric and neurological conditions where dimensionality and entanglement pose similar challenges, such as schizophrenia or bipolar disorder. Moreover, extending the framework to dynamic functional connectivity or graph-theoretic metrics may capture temporal and topological nuances currently overlooked in static correlation analyses.

5. Conclusion

In summary, we have demonstrated that integrating cortically inspired sparse encoding with contrastive manifold learning provides an effective framework for disentangling the high-dimensional structure of functional brain networks in MDD. The HTM-SP + cPCA pipeline achieves superior diagnostic accuracy up to 86% while identifying interpretable, network-level alterations consistent with contemporary neurobiological models of depression. Importantly, our analysis revealed that only 25% of the most discriminative connections identified by cPCA showed statistically significant differences in traditional univariate testing, demonstrating that contrastive machine learning offers enhanced sensitivity to diagnostically relevant patterns that conventional statistical approaches fail to detect. From a complex-systems perspective, the identified alterations are concentrated at inter-modular connections bridging distinct functional subsystems and involve regions occupying hub-like positions within the functional connectome, suggesting that MDD is associated with impaired integrative network organization rather than focal disruption. These findings support the conceptualization of depression as an emergent property of large-scale network dysregulation and highlight the utility of combining biologically grounded dimensionality reduction with variance disentanglement techniques for probing the organizational principles of complex neural systems in health and disease.

CRediT authorship contribution statement

Muhammad Salman Kabir: Writing – original draft, Visualization, Software, Methodology, Investigation. **Semen Kurkin:** Writing – original draft, Visualization, Validation, Supervision, Methodology, Investigation, Data curation, Conceptualization. **Rositsa Paunova:** Writing – review & editing, Validation, Methodology, Investigation, Data curation. **Drozdstoy Stoyanov:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Formal analysis, Conceptualization. **Alexander Hramov:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Formal analysis, Conceptualization.

Institutional review board statement

The study was conducted in accordance with the Declaration of Helsinki, and approved by the Local Ethics Committee.

Informed consent statement

Informed consent was obtained from all subjects involved in the study.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The work was supported by Russian Science Foundation (Grant 23-71-30010).

Appendix A. Hierarchical temporal memory

The Hierarchical Temporal Memory architecture consists of four interconnected components: an encoder, a spatial pooler, temporal memory and a classifier.

A.0.1. Encoder

The encoder converts raw input data (which could be numbers, categories, dates, or other data types) into binary vectors. Each encoder is designed for specific data types and creates representations where semantically similar inputs produce overlapping binary patterns. For example, similar numbers would have more overlapping bits than distant numbers. This preserves the semantic relationships in the data.

A.0.2. Spatial pooler

The spatial pooler takes the binary vectors from the encoder and transforms them into Sparse Distributed Representations (SDRs). Its main functions are:

- Creating a fixed-size sparse representation (typically, 3% active bits);
- Learning which input patterns commonly occur together;
- Maintaining the semantic similarity from the input (similar inputs produce similar SDRs);
- Normalizing the input space so the temporal memory receives consistent representations.

It essentially performs a form of dimensionality reduction while preserving and enhancing the important structure in the data.

A.0.3. Temporal memory[1]

This is where HTM learns sequences and temporal patterns. The temporal memory:

- Learns which patterns typically follow other patterns in time;
- Forms predictions about what will come next based on the current input and context from previous inputs;
- Recognizes sequences even when they share common elements;
- Uses context from the past to disambiguate identical inputs that occur in different sequences.

It maintains a representation of temporal context, allowing it to make predictions and detect anomalies.

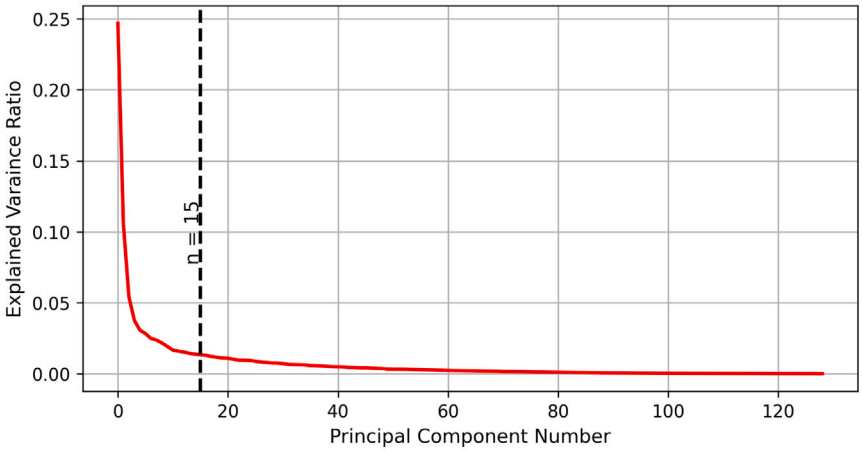
A.0.4. Classifier

The classifier takes the SDRs and maps them to actual predictions or outputs. It can be trained to:

- Predict future values;
- Classify the current input;
- Detect anomalies (when predictions do not match reality).

The classifier learns associations between the SDR patterns and the desired outputs.

Appendix B. Choice of 15 principal components



To reduce the dimensionality of our data while retaining most of the variance, we performed principal component analysis. The plot shows the eigenvalues of all principal components. We observed an 'elbow' at the 15th component, indicating that the first 15 principal components capture the majority of the variance. Therefore, we retained these 15 components for subsequent analyses.

Appendix C. Classification performance

See Tables C.5 and C.6.

Table C.5
10-fold cross-validation scores for different inputs to Gaussian process classifier.

Input to Gaussian process classifier	Accuracy	Precision	Recall	F1 score
All features	0.50	–	0	0
Entangled PCs without SP selection	0.55 ± 0.16	0.54 ± 0.31	0.33 ± 0.23	0.40 ± 0.25
Disentangled PCs without SP selection	0.78 ± 0.10	0.83 ± 0.13	0.73 ± 0.19	0.76 ± 0.12
SP selected features	0.55 ± 0.12	0.56 ± 0.15	0.41 ± 0.20	0.46 ± 0.17
Entangled PCs with SP selection	0.56 ± 0.14	0.60 ± 0.15	0.49 ± 0.16	0.53 ± 0.13
Disentangled PCs with SP selection	0.86 ± 0.07	0.83 ± 0.08	0.93 ± 0.07	0.87 ± 0.07

[1] This component is not used in this study.

Table C.6
10-fold cross-validation scores for different inputs to SVM classifier.

Input to SVM classifier	Accuracy	Precision	Recall	F1 score
All features	0.59 ± 0.14	0.55 ± 0.20	0.61 ± 0.27	0.57 ± 0.23
Entangled PCs without SP selection	0.64 ± 0.14	0.65 ± 0.24	0.54 ± 0.25	0.57 ± 0.23
Disentangled PCs without SP selection	0.79 ± 0.12	0.74 ± 0.13	0.93 ± 0.12	0.82 ± 0.11
SP selected features	0.65 ± 0.13	0.67 ± 0.30	0.51 ± 0.25	0.56 ± 0.25
Entangled PCs with SP selection	0.59 ± 0.14	0.61 ± 0.29	0.46 ± 0.21	0.51 ± 0.22
Disentangled PCs with SP selection	0.86 ± 0.09	0.86 ± 0.13	0.91 ± 0.11	0.87 ± 0.07

Appendix D. SP selected connections

See Table D.7.

Table D.7
The most informative connections obtained using SP-based feature selection. Significantly hypoconnected connections in MDD compared to HC are marked with bold.

Precentral_L – Precentral_R	**Occipital_Sup_L – Parietal_Sup_R**
Precentral_L – Frontal_Mid_2_L	Occipital_Sup_R – Occipital_Mid_R
Precentral_L – Postcentral_L	**Occipital_Sup_R – Postcentral_L**
Precentral_L – Postcentral_R	Occipital_Mid_R – Fusiform_L
Precentral_L – Parietal_Sup_L	Occipital_Inf_L – Fusiform_L
Precentral_L – Parietal_Inf_L	Occipital_Inf_L – Fusiform_R
Precentral_R – Frontal_Inf_Oper_R	Occipital_Inf_R – Fusiform_R
Precentral_R – Supp_Motor_Area_R	Occipital_Inf_R – Temporal_Inf_R
Precentral_R – Postcentral_R	Fusiform_L – Fusiform_R
Precentral_R – Paracentral_Lobule_L	Fusiform_L – Temporal_Inf_R
Frontal_Sup_2_L – Frontal_Sup_2_R	Fusiform_L – Cerebellum_6_L
Frontal_Sup_2_L – Frontal_Mid_2_L	Fusiform_L – Vermis_4_5
Frontal_Sup_2_L – Frontal_Mid_2_R	Fusiform_R – Temporal_Mid_R
Frontal_Sup_2_R – Frontal_Mid_2_L	Fusiform_R – Temporal_Inf_L
Frontal_Sup_2_R – Supp_Motor_Area_R	Fusiform_R – Temporal_Inf_R
Frontal_Sup_2_R – Frontal_Sup_Medial_R	Fusiform_R – Vermis_4_5
Frontal_Mid_2_L – Frontal_Mid_2_R	Postcentral_L – Postcentral_R
Frontal_Mid_2_L – Cuneus_R	**Postcentral_L – Parietal_Sup_R**
Frontal_Mid_2_L – Parietal_Sup_L	**Postcentral_L – Parietal_Inf_L**
Frontal_Mid_2_L – Parietal_Inf_L	Postcentral_L – SupraMarginal_L
Frontal_Inf_Oper_L – Rolandic_Oper_L	**Postcentral_L – SupraMarginal_R**
Frontal_Inf_Oper_L – Rolandic_Oper_R	Postcentral_R – SupraMarginal_R
Frontal_Inf_Oper_L – SupraMarginal_L	**Postcentral_R – Paracentral_Lobule_R**
Frontal_Inf_Oper_R – Frontal_Inf_Tri_R	Parietal_Sup_L – Parietal_Inf_L
Frontal_Inf_Oper_R – Rolandic_Oper_R	Parietal_Sup_L – SupraMarginal_L
Frontal_Inf_Tri_L – Frontal_Inf_Tri_R	**Parietal_Sup_L – Paracentral_Lobule_L**
Frontal_Inf_Tri_R – Rolandic_Oper_R	Parietal_Sup_R – Parietal_Inf_R
Rolandic_Oper_L – SupraMarginal_L	**Parietal_Sup_R – Paracentral_Lobule_L**
Rolandic_Oper_R – Insula_R	SupraMarginal_L – SupraMarginal_R
Rolandic_Oper_R – SupraMarginal_R	Precuneus_L – Precuneus_R
Rolandic_Oper_R – Heschl_R	Caudate_L – Caudate_R
Supp_Motor_Area_L – Supp_Motor_Area_R	Caudate_L – Temporal_Mid_L
Supp_Motor_Area_L – Paracentral_Lobule_L	Heschl_R – Temporal_Sup_R
Supp_Motor_Area_L – Paracentral_Lobule_R	Temporal_Pole_Sup_R – Temporal_Mid_R
Supp_Motor_Area_R – Paracentral_Lobule_L	Temporal_Mid_L – Temporal_Mid_R
Frontal_Sup_Medial_L – Frontal_Sup_Medial_R	Temporal_Mid_L – Temporal_Inf_L
Frontal_Sup_Medial_R – ACC_pre_L	Temporal_Mid_L – Cerebellum_6_R
Frontal_Med_Orb_L – Frontal_Med_Orb_R	Temporal_Mid_R – Temporal_Inf_L
Insula_L – Putamen_L	Temporal_Mid_R – Temporal_Inf_R
Insula_R – Heschl_R	Temporal_Mid_R – Vermis_4_5
Insula_R – Temporal_Sup_R	Temporal_Inf_L – Temporal_Inf_R
Cingulate_Mid_L – Cingulate_Mid_R	Cerebellum_Crus1_L – Cerebellum_6_R
Cingulate_Post_L – Cingulate_Post_R	Cerebellum_4_5_L – Cerebellum_4_5_R
Hippocampus_L – Hippocampus_R	Cerebellum_4_5_L – Vermis_4_5
Hippocampus_L – Temporal_Mid_L	Cerebellum_4_5_R – Cerebellum_6_L
Hippocampus_R – Fusiform_R	Cerebellum_6_L – Cerebellum_6_R
Hippocampus_R – Temporal_Mid_R	Cerebellum_6_L – Vermis_6
Hippocampus_R – Cerebellum_4_5_R	Cerebellum_6_R – Vermis_6
ParaHippocampal_R – Temporal_Mid_L	Thal_VL_L – Thal_VL_R
ParaHippocampal_R – Temporal_Mid_R	Thal_VPL_L – Thal_PuM_R
Calcarine_L – Calcarine_R	Thal_VPL_R – Thal_PuA_L

(*continued on next page*)

Table D.7 (*continued*).

Calcarine_L – Cuneus_L	Thal_PuI_R – Thal_PuM_R
Calcarine_L – Lingual_R	Thal_PuM_L – Thal_PuA_L
Calcarine_L – Occipital_Mid_L	ACC_pre_R – ACC_sup_L
Calcarine_L – Temporal_Inf_R	VTA_L – SN_pc_L
Calcarine_R – Cuneus_L	SN_pc_L – SN_pr_L
Calcarine_R – Cuneus_R	**Calcarine_R – Occipital_Mid_R**
Cuneus_L – Cuneus_R	**Cuneus_L – Occipital_Sup_L**
Cuneus_R – Occipital_Sup_L	Cuneus_R – Occipital_Sup_R
Cuneus_R – Angular_R	**Cuneus_R – Precuneus_L**
Lingual_L – Lingual_R	Lingual_L – Occipital_Mid_L
Lingual_L – Fusiform_R	Lingual_R – Occipital_Inf_L
Lingual_R – Fusiform_R	Lingual_R – Postcentral_L
Occipital_Sup_L – Occipital_Sup_R	**Occipital_Sup_L – Occipital_Mid_L**
Occipital_Sup_L – Occipital_Mid_R	

Appendix E. Mapping contrastive components to original features

See Fig. E.7.

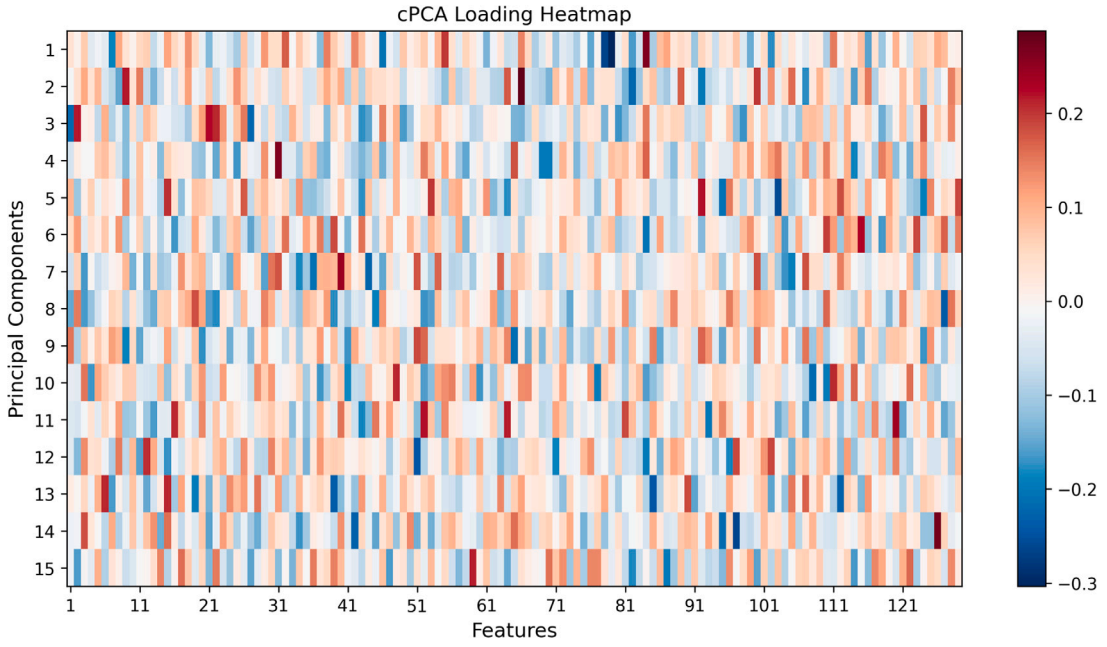


Fig. E.7. cPCA loadings. Heatmap showing the loading matrix from contrastive PCA analysis across 15 principal components and 129 brain connectivity connections (features). Color intensity indicates loading magnitude. The sparse pattern indicates component-specific feature contributions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Data availability

Data will be made available on request.

References

[1] Shehab M, Abualigah L, Shambour Q, Abu-Hashem MA, Shambour MKY, Alsalibi AI, Gandomi AH. Machine learning in medical applications: A review of state-of-the-art methods. Comput Biol Med 2022;145:105458. http://dx.doi.org/10.1016/j.compbiomed.2022.105458.

[2] Squires M, Tao X, Elangovan S, Gururajan R, Zhou X, Acharya UR, Li Y. Deep learning and machine learning in psychiatry: a survey of current progress in depression detection, diagnosis and treatment. Brain Inform 2023;10(1):10. http://dx.doi.org/10.1186/s40708-023-00188-6.

[3] Bains N, Abdijadid S. Major depressive disorder. Treasure Island (FL): StatPearls Publishing; 2025, http://www.ncbi.nlm.nih.gov/books/NBK559078/.

[4] Cui L, Li S, Wang S, Wu X, Liu Y, Yu W, Wang Y, Tang Y, Xia M, Li B. Major depressive disorder: hypothesis, mechanism, prevention and treatment. Signal Transduct Target Ther 2024;9(1):1–32. http://dx.doi.org/10.1038/s41392-024-01738-y.

[5] Chen Y, Zhao W, Yi S, Liu J. The diagnostic performance of machine learning based on resting-state functional magnetic resonance imaging data for major depressive disorders: a systematic review and meta-analysis. Front Neurosci 2023;17. http://dx.doi.org/10.3389/fnins.2023.1174080, https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2023.1174080/full.

[6] Chen Q, Bi Y, Yan W, Wu S, Xia T, Wang Y, Huang S, Zhou C, Xie S, Kuang S, Kong W, Lv Z. Abnormal voxel-mirrored homotopic connectivity in first-episode major depressive disorder using fmri: a machine learning approach. Front Psychiatry 2023;14. http://dx.doi.org/10.3389/fpsyt.2023.1241670, https://www.frontiersin.org/journals/psychiatry/articles/10.3389/fpsyt.2023.1241670/full.

[7] Gallo S, El-Gazzar A, Zhutovsky P, Thomas RM, Javaheripour N, Li M, Bartova L, Bathula D, Dannlowski U, Davey C, Frodl T, Gotlib I, Grimm S, Grotegerd D, Hahn T, Hamilton PJ, Harrison BJ, Jansen A, Kircher T, Meyer B, Nenadić I, Olbrich S, Paul E, Pezawas L, Sacchet MD, Sämann P, Wagner G, Walter H, Walter M, van Wingen G. Functional connectivity signatures of major depressive disorder: machine learning analysis of two multicenter neuroimaging studies. Mol Psychiatry 2023;28(7):3013–22. http://dx.doi.org/10.1038/s41380-023-01977-5.

[8] Lin C, Lee S-H, Huang C-M, Chen G-Y, Chang W, Liu H-L, Ng S-H, Lee TM-C, Wu S-C. Automatic diagnosis of late-life depression by 3d convolutional neural networks and cross-sample entropy analysis from resting-state fmri. Brain Imaging Behav 2023;17(1):125–35. http://dx.doi.org/10.1007/s11682-022-00748-0.

[9] Venkatapathy S, Votinov M, Wagels L, Kim S, Lee M, Habel U, Ra I-H, Jo H-G. Ensemble graph neural network model for classification of major depressive disorder using whole-brain functional connectivity. Front Psychiatry 2023;14. http://dx.doi.org/10.3389/fpsyt.2023.1125339, https://www.frontiersin.org/journals/psychiatry/articles/10.3389/fpsyt.2023.1125339/full.

[10] Wang J, Li T, Sun Q, Guo Y, Yu J, Yao Z, Hou N, Hu B. Automatic diagnosis of major depressive disorder using a high- and low-frequency feature fusion framework. Brain Sci 2023;13(1111):1590. http://dx.doi.org/10.3390/brainsci13111590.

[11] Zhu M, Quan Y, He X. The classification of brain network for major depressive disorder patients based on deep graph convolutional neural network. Front Hum Neurosci 2023;17. http://dx.doi.org/10.3389/fnhum.2023.1094592, https://www.frontiersin.org/articles/10.3389/fnhum.2023.1094592.

[12] Dai P, Zhou Y, Shi Y, Lu D, Chen Z, Zou B, Liu K, Liao S, Consortium TRmM. Classification of mdd using a transformer classifier with large-scale multisite resting-state fmri data. Hum Brain Mapp 2024;45(1):e26542. http://dx.doi.org/10.1002/hbm.26542.

[13] Noman F, Ting CM, Kang H, Phan RCW, Ombao H. Graph autoencoders for embedding learning in brain networks and major depressive disorder identification. IEEE J Biomed Health Informatics 2024;28(3):1644–55. http://dx.doi.org/10.1109/JBHI.2024.3351177.

[14] Stoyanov D, Khorev V, Paunova R, Kandilarova S, Simeonova D, Badarin A, Hramov A, Kurkin S. Resting-state functional connectivity impairment in patients with major depressive episode. Int J Environ Res Public Health 2022;19(21):14045.

[15] Pitsik EN, Maximenko VA, Kurkin SA, Sergeev AP, Stoyanov D, Paunova R, Kandilarova S, Simeonova D, Hramov AE. The topology of fmri-based networks defines the performance of a graph neural network for the classification of patients with major depressive disorder. Chaos Solitons Fractals 2023;167:113041.

[16] Pisarchik AN, Andreev AV, Kurkin SA, Stoyanov D, Badarin AA, Paunova R, Hramov AE. Topology switching during window thresholding fmri-based functional networks of patients with major depressive disorder: Consensus network approach. Chaos: An Interdiscip J Nonlinear Sci 2023;33(9).

[17] Khorev VS, Kurkin SA, Zlateva G, Paunova R, Kandilarova S, Maes M, Stoyanov D, Hramov AE. Disruptions in segregation mechanisms in fmri-based brain functional network predict the major depressive disorder condition. Chaos Solitons Fractals 2024;188:115566.

[18] Stoyanov D, Paunova R, Najar D, Zlateva G, Kandilarova S, Khorev V, Kurkin SA. Cross-validation of paranoid and depressive scales: Results from functional mri group independent component analysis. Ment Illn 2024;2024(1):7739939.

[19] Kurkin SA, Smirnov NM, Paunova R, Kandilarova S, Stoyanov D, Mayorova L, Hramov AE. Beyond pairwise interactions: higher-order q-analysis of fmri-based brain functional networks in patients with major depressive disorder. IEEE Access 2024.

[20] Berisha V, Krantsevich C, Hahn PR, Hahn S, Dasarathy G, Turaga P, Liss J. Digital medicine and the curse of dimensionality. Npj Digit Med 2021;4(1):1–8. http://dx.doi.org/10.1038/s41746-021-00521-5.

[21] Mwangi B, Tian TS, Soares JC. A review of feature reduction techniques in neuroimaging. Neuroinformatics 2014;12(2):229–44. http://dx.doi.org/10.1007/s12021-013-9204-3.

[22] Kernbach JM, Ort J, Hakvoort K, Clusmann H, Delev D, Neuloh G. Dimensionality reduction: Foundations and applications in clinical neuroscience. Acta Neurochir Suppl 2022;134:59–63. http://dx.doi.org/10.1007/978-3-030-85292-4-8.

[23] Li F, Wang G, Jiang L, Yao D, Xu P, Ma X, Dong D, He B. Disease-specific resting-state eeg network variations in schizophrenia revealed by the contrastive machine learning. Brain Res Bull 2023;202:110744. http://dx.doi.org/10.1016/j.brainresbull.2023.110744.

[24] Aglinskas A, Hartshorne JK, Anzellotti S. Contrastive machine learning reveals the structure of neuroanatomical variation within autism. Science 2022;376(6597):1070–4. http://dx.doi.org/10.1126/science.abm2461.

[25] Kabir MS, Kurkin S, Portnova G, Martynova O, Wang Z, Hramov A. Contrastive machine learning reveals in eeg resting-state network salient features specific to autism spectrum disorder. Chaos Solitons Fractals 2024;185:115123. http://dx.doi.org/10.1016/j.chaos.2024.115123.

[26] Tong X, Xie H, Fonzo GA, Zhao K, Satterthwaite TD, Carlisle NB, Zhang Y. Symptom dimensions of resting-state electroencephalographic functional connectivity in autism. Nat Ment Health 2024;2(3):287–98. http://dx.doi.org/10.1038/s44220-023-00195-w.

[27] Li J, Cheng K, Wang S, Morstatter F, Trevino RP, Tang J, Liu H. Feature selection: A data perspective. ACM Comput Surv 2017;50(6):94:1–45. http://dx.doi.org/10.1145/3136625.

[28] Andreev AV, Kurkin SA, Stoyanov D, Badarin AA, Paunova R, Hramov AE. Toward interpretability of machine learning methods for the classification of patients with major depressive disorder based on functional network measures. Chaos: An Interdiscip J Nonlinear Sci 2023;33(6).

[29] Frolov N, Kabir MS, Maksimenko V, Hramov A. Machine learning evaluates changes in functional connectivity under a prolonged cognitive load. Chaos: An Interdiscip J Nonlinear Sci 2021;31(10).

[30] Kabir MS, Kurkin S. Combination of machine learning and functional networks concept for diagnosis of autism spectrum disorder. In: 2022 fourth international conference neurotechnologies and neurointerfaces. CNN, IEEE; 2022, p. 63–7.

[31] Chipman HA, Gu H. Interpretable dimension reduction. J Appl Stat 2005;32(9):969–87. http://dx.doi.org/10.1080/02664760500168648.

[32] Le-Khac PH, Healy G, Smeaton AF. Contrastive representation learning: A framework and review. IEEE Access 2020;8:193907–34. http://dx.doi.org/10.1109/ACCESS.2020.3031549.

[33] Abid A, Zhang MJ, Bagaria VK, Zou J. Exploring patterns enriched in a dataset with contrastive principal component analysis. Nat Commun 2018;9(1):2134. http://dx.doi.org/10.1038/s41467-018-04608-8.

[34] Aglinskas A, Schwartz E, Anzellotti S. Disentangling disorder-specific variation is key for precision psychiatry in autism. Front Behav Neurosci 2023;17. http://dx.doi.org/10.3389/fnbeh.2023.1121017, https://www.frontiersin.org/journals/behavioral-neuroscience/articles/10.3389/fnbeh.2023.1121017/full.

[35] Ding C, Sun Y, Li K, Xie S, Yan H, Li P, Yan J, Chen J, Wang H, Wang H, Chen Y, Yang Y, Lv L, Zhang H, Lu L, Zhang D, Chen Y, Zhang Z, Jiang T, Liu B. Disorder-specific neurodynamic features in schizophrenia inferred by neurodynamic embedded contrastive variational autoencoder model. Transl Psychiatry 2024;14(1):1–14. http://dx.doi.org/10.1038/s41398-024-03200-7.

[36] Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, Hergueta T, Baker R, Dunbar GC, et al. The mini-international neuropsychiatric interview (mini): the development and validation of a structured diagnostic psychiatric interview for dsm-iv and icd-10. J Clin Psychiatry 1998;59(20):22–33.

[37] Müller MJ, Himmerich H, Kienzle B, Szegedi A. Differentiating moderate and severe depression using the montgomery–åsberg depression rating scale (madrs). J Affect Disord 2003;77(3):255–60.

[38] Montgomery SA, Åsberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry 1979;134(4):382–9.

[39] Stoyanov D, Kandilarova S, Aryutova K, Paunova R, Todeva-Radneva A, Latypova A, Kherif F. Multivariate analysis of structural and functional neuroimaging can inform psychiatric differential diagnosis. Diagnostics 2020;11(1):19.

[40] Penny William D, Friston Karl J, Ashburner John T, Kiebel Stefan J, Nichols Thomas E. Statistical parametric mapping: the analysis of functional brain images. Elsevier; 2011.
[41] Rolls ET, Huang CC, Lin CP, Feng J, Joliot M. Automated anatomical labelling atlas 3. Neuroimage 2020;206:116189.
[42] Cui Y, Ahmad S, Hawkins J. The htm spatial pooler—a neocortical algorithm for online sparse distributed coding. Front Comput Neurosci 2017;11. http://dx.doi.org/10.3389/fncom.2017.00111, https://www.frontiersin.org/articles/10.3389/fncom.2017.00111.
[43] Chen X, Wang W, Li W. An overview of hierarchical temporal memory: A new neocortex algorithm. In: 2012 proceedings of international conference on modelling. Identification and Control; 2012, p. 1004–10, https://ieeexplore.ieee.org/abstract/document/6260285.
[44] Hawkins J, Ahmad S, Cui Y. A theory of how columns in the neocortex enable learning the structure of the world. Front Neural Circuits 2017;11:81. http://dx.doi.org/10.3389/fncir.2017.00081.
[45] Crochet S, Poulet JFA, Kremer Y, Petersen CCH. Synaptic mechanisms underlying sparse coding of active touch. Neuron 2011;69(6):1160–75. http://dx.doi.org/10.1016/j.neuron.2011.02.022.
[46] Ahmad S, Hawkins J. How do neurons operate on sparse distributed representations? a mathematical theory of sparsity, neurons and active dendrites (arxiv:1601.00720). 2016, http://dx.doi.org/10.48550/arXiv.1601.00720, arXiv:1601.00720 [cs, q-bio].
[47] Hebb DO. The organization of behavior; a neuropsychological theory. In: The organization of behavior; a neuropsychological theory. Oxford, England: Wiley; 1949.
[48] Song S, Miller KD, Abbott LF. Competitive hebbian learning through spike-timing-dependent synaptic plasticity. Nature Neurosci 2000;3(9):919–26. http://dx.doi.org/10.1038/78829.
[49] Kabir MS, Kurkin S, Paunova R, Stoyanov D, Hramov A. Htm spatial pooler - a nonparametric interpretable feature selection algorithm? an introductory exploration. In: 2024 8th scientific school dynamics of complex networks and their applications. DCNA, 2024, p. 106–11. http://dx.doi.org/10.1109/DCNA63495.2024.10718624, https://ieeexplore.ieee.org/abstract/document/10718624.
[50] Fushiki T. Estimation of prediction error by using k-fold cross-validation. Stat Comput 2011;21(2):137–46. http://dx.doi.org/10.1007/s11222-009-9153-8.
[51] Rasmussen CE. Gaussian processes in machine learning. Berlin, Heidelberg: Springer; 2004, p. 63–71. http://dx.doi.org/10.1007/978-3-540-28650-9_4.
[52] M H, n SM. A review on evaluation metrics for data classification evaluations, international journal of data mining & knowledge management process. 2015, https://www.semanticscholar.org/paper/A-Review-On-Evaluation-Metrics-For-Data-Evaluations-H.-M.N/d79e64f90000c1594dc6e12908ecdcddaac49459.
[53] Kaiser RH, Andrews-Hanna JR, Wager TD, Pizzagalli DA. Large-scale network dysfunction in major depressive disorder: a meta-analysis of resting-state functional connectivity. JAMA Psychiatry 2015;72(6):603–11.
[54] Mulders PC, van Eijndhoven PF, Schene AH, Beckmann CF, Tendolkar I. Resting-state functional connectivity in major depressive disorder: a review. Neurosci Biobehav Rev 2015;56:330–44.
[55] Menon V. Large-scale brain networks and psychopathology: a unifying triple network model. Trends Cogn Sci 2011;15(10):483–506.
[56] Drysdale AT, Grosenick L, Downar J, Dunlop K, Mansouri F, Meng Y, Fetcho RN, Zebley B, Oathes DJ, Etkin A, et al. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. Nature Med 2017;23(1):28–38.
[57] Li W, Wang Y, Ward BD, Antuono PG, Li SJ, Goveas JS. Intrinsic inter-network brain dysfunction correlates with symptom dimensions in late-life depression. J Psychiatr Res 2017;87:71–80.
[58] Walther S, Federspiel A, Horn H, Razavi N, Wiest R, Dierks T, Strik W, Müller TJ. Resting state cerebral blood flow and objective motor activity reveal basal ganglia dysfunction in schizophrenia. Psychiatry Res: Neuroimaging 2011;192(2):117–24.
[59] Northoff G, Wiebking C, Feinberg T, Panksepp J. The 'resting-state hypothesis' of major depressive disorder—a translational subcortical–cortical framework for a system disorder. Neurosci Biobehav Rev 2011;35(9):1929–45.
[60] Wang L, Dai Z, Peng H, Tan L, Ding Y, He Z, Zhang Y, Xia M, Li Z, Li W, et al. Overlapping and segregated resting-state functional connectivity in patients with major depressive disorder with and without childhood neglect. Hum Brain Mapp 2014;35(4):1154–66.
[61] Sheline YI, Price JL, Yan Z, Mintun MA. Resting-state functional mri in depression unmasks increased connectivity between networks via the dorsal nexus. Proc Natl Acad Sci 2010;107(24):11020–5.
[62] Guo W, Liu F, Xue Z, Gao K, Liu Z, Xiao C, Chen H, Zhao J. Decreased interhemispheric coordination in treatment-resistant depression: a resting-state fmri study. PloS One 2013;8(8):e71368.
[63] Cheng W, Rolls ET, Qiu J, Xie X, Lyu W, Li Y, Huang CC, Yang AC, Tsai SJ, Lyu F, et al. Functional connectivity of the human amygdala in health and in depression. Soc Cogn Affect Neurosci 2018;13(6):557–68.
[64] Rosselli M, Lang M, Arruda F. Executive dysfunction in depressive disorders. In: Dysexecutive syndromes: clinical and experimental perspectives. 2019, p. 241–59.
[65] Aglinskas A, Hartshorne JK, Anzellotti S. Contrastive machine learning reveals the structure of neuroanatomical variation within autism. Science 2022;376(6597):1070–4.
[66] Kaiser RH, Whitfield-Gabrieli S, Dillon DG, Goer F, Beltzer M, Minkel J, Smoski M, Dichter G, Pizzagalli DA. Dynamic resting-state functional connectivity in major depression. Neuropsychopharmacology 2016;41(7):1822–30.