2024 8th Scientific School Dynamics of Complex Networks and their Applications (DCNA) | 979-8-3503-9083-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/DCNA63495.2024.10718624

HTM Spatial Pooler – a Nonparametric Interpretable Feature Selection Algorithm? An Introductory Exploration

Muhammad Salman Kabir
Innopolis University
Innopolis, Russia
m.kabir@innopolis.university

Semen Kurkin
Baltic Center for Neurotechnology and Artificial Intelligence
Immanuel Kant Baltic Federal University,
Kaliningrad, Russia
Orcid: 0000-0002-3438-5717

Rositsa Paunova
Plovdiv Medical University
Plovdiv, Bulgaria
Rositsa.Paunova@mu-plovdiv.bg

Drozdstoy Stoyanov
Plovdiv Medical University
Plovdiv, Bulgaria
drozdstoy.Stoyanov@mu-plovdiv.bg

Alexander Hramov
Baltic Center for Neurotechnology and Artificial Intelligence
Immanuel Kant Baltic Federal University,
Kaliningrad, Russia
hramovae@gmail.com

***Abstract*—Neuroimaging data analysis suffers from the curse of dimensionality. Feature selection plays an important role in mitigating this curse. Here we explore the potential of the Hierarchical Temporal Memory (HTM) Spatial Pooler (SP) as a nonparametric interpretable feature selection algorithm for high-dimensional neuroimaging data. Our results indicate that SP demonstrates comparable performance to other feature selection algorithms in terms of feature stability and classification performance.**

***Index Terms*—Feature selection, Hierarchical temporal memory, Spatial pooler, Neuroimaging data, Machine learning, Interpretability**

I. Introduction

Analysis of neuroimaging data presents a significant challenge because of its high dimensionality. For instance, functional connectivity vectors derived from BOLD signals across 200 brain parcellations comprise of 19,900 unique connections. This high-dimensional feature space, coupled with typically limited sample size in neuroimaging studies, often leads to overfitting in machine learning models [1]–[4]. Additionally, such high-dimensional data hinder effective visualization of group differences [5]–[7].

To address these challenges, researchers employ several dimension reduction and feature selection algorithms [1]. Commonly used dimension reduction algorithms include Uniform Manifold Approximation and Projection (UMAP), t-Distributed Stochastic Neighbor Embedding (t-SNE), Principal Component Analysis (PCA) etc. While these techniques effectively reduce dimensionality, they often compromise result interpretability [8]. Alternatively, feature selection algorithms maintain interpretability by choosing a subset of original features based on specific criteria [9]. Common approaches include selecting features based on ANOVA F-test score, chi-squared score, mutual information score etc. As machine learning models are increasingly applied to neuroimaging studies, the need for interpretability has grown significantly to ensure that these systems provide not only accurate predictions but also meaningful insights into the underlying neural mechanisms, ultimately enhancing trust and applicability in clinical settings [10]–[13].

Hierarchical Temporal Memory (HTM) is a physiological-interpretable machine learning model that is inspired by the structural and functional properties of neocortex [14]. Neocortex is the region of the mammalian brain responsible for higher-order cognitive functions such as vision, hearing, motor control, language, and planning. Biological evidence suggests that neocortex implements a common set of algorithms across its neural circuitry despite the diversity of tasks it performs [15]. HTM provides a theoretical framework to understand and model the neocortex's capabilities and its underlying algorithm [16]. HTM architecture comprises four interconnected components namely encoder, spatial pooler, temporal memory and classifier that work together to learn from input data (Fig. 1).

This study explores the potential of Spatial Pooler (SP) [16], a component in HTM architecture, as a nonparametric interpretable feature selection algorithm. We computed the stability of features selected by SP from two neuroimaging

This work was supported by Russian Science Foundation (Grant No. 23-71-30010).

979-8-3503-9083-4/24/$31.00 ©2024 IEEE

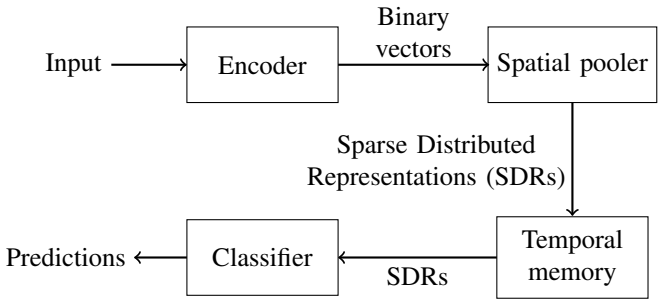

Fig. 1. **HTM architecture.** An HTM system comprises of an encoder, SP, temporal memory and a classifier.

datasets and compared obtained stability with several established feature selection algorithms. Additionally, we assessed the classification performance of machine learning classifiers using the SP-selected features to validate the effectiveness.

Our exploration aimed to:

- Evaluate the stability of features selected by the SP,
- Assess the classification performance of machine learning models using SP-selected features,
- Compare the SP performance with other feature selection algorithms.

The rest of paper is divided as follow: section II briefly describes the working of SP, section III elaborates the evaluation methodology, section IV presents the results and section V concludes the paper.

II. HTM Spatial Pooler

HTM is effective because it uses Sparse Distributed Representations (SDRs) in its operation [17]. Researchers have observed the existence of SDRs in several cortical regions, including auditory, visual, and somatosensory regions [18]. The purpose of SP in HTM framework is to encode stream of sensory inputs into SDRs. The design of spatial pooler incorporates several neurobiological principles such as competitive Hebbian learning, homeostatic excitability regulation, sensory cortex connection topology and activity-dependent structural plasticity [16].

The SP operates on the binary vectors generated by the encoder and assigns columns to these binary vectors. These columns correspond to dendritic segments of neocortex neurons. Each column has a set of synapses with associated permanence values, initialized randomly. Active synapses (synapses that are connected to active bits in the input binary vector) inhibit neighboring columns through a process known as Hebbian learning [19], [20]. This inhibition mechanism results in the creation of SDRs consisting of active columns. The SP's operation is governed by thresholds that determine whether a column is considered active or inactive.

Fig. 2 shows how SP takes in a binary input and converts it into an SDR. The input to the SP is a binary vector, which may or may not be sparse. A binary matrix represents the connected synapses in the SP. This matrix, while not necessarily sparse, typically exhibits sparsity in practice. The process involves a dot multiplication between input and binary matrix, resulting in a vector of overlap counts. The subsequent inhibition step determines the column winners, which form the output. Specifically, the indexes corresponding to the highest overlap counts become the active bits in the output SDR.

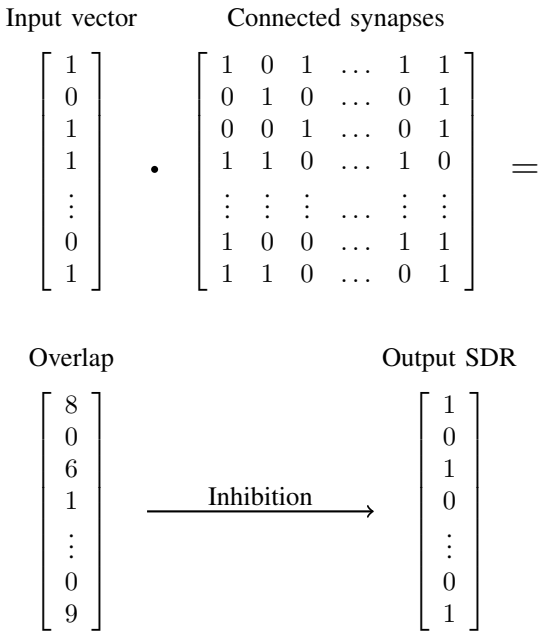

Fig. 2. **SP mechanism.** The binary encoded input vector is dot multiplied by the connected synapses to produce overlap values, which then undergo inhibition to generate the final SDR output.

III. Evaluation

Our analysis used two neuroimaging datasets to explore the feasibility of SP as a feature selection algorithm; Preprocessed Autism Brain Imaging Data Exchange (ABIDE) dataset [21] and custom-curated rs-fMRI of Major Depressive Disorder [22]. For the rest of study, this rs-fMRI dataset will be addressed as MDD. We chose these datasets because several previous studies successfully classified participants based on features present in them [22]–[24].

The preprocessed ABIDE dataset comprises neuroimaging data collected for Autism Spectrum Disorder (ASD) research. Preprocessing removed noise, corrected artifacts, and aligned data across imaging centers and acquisition protocols. This preprocessing makes the data more suitable for comparative and statistical analyses. Table I details the characteristics

of the preprocessed ABIDE dataset used in this study. [21] describes these characteristics in details. The MDD dataset included 140 participants: 70 with major depressive disorder and 70 healthy controls. [22], [25], [26] describe the collection, processing and characteristics of MDD dataset.

TABLE I
CHARACTERISTICS OF PREPROCESSED ABIDE DATASET

Autistic participants	391
Neurotypical participants	391
Neuroimaging data	rs-fMRI
Preprocessing pipeline	CPAC
Preprocessing strategy	filt_global
Regions of interest	Craddock 200 atlas

For each participant, we computed Pearson Correlation Coefficient [27] among all pairs of parcellations to obtain functional brain connectivity matrix. Since obtained connectivity matrix were diagonally symmetrical, we considered only the unique connections among parcellations. This resulted in vectors of sizes 19,900 and 13,530 for each participant in ABIDE and MDD dataset respectively. The functional connectivity vector of each participant was then passed through an encoder which assigned a value of 1 for a correlation value of above 0 and 0 otherwise.

Table II outlines the hyperparameters setting for the SPs used in our study [28]. We employed a global inhibition mechanism, set potential radius to the size of input vector and set percentage of potential synapses to 100. We operated under the assumption that each column in the SP was connected to the entire input space.

TABLE II
HYPERPARAMETERS SETTING FOR TRAINING SPS

Hyperparameters	**ABIDE**	**MDD**
Input size	19900	13530
Potential radius	19900	13530
Columns size	500	500
Boost strength	300	100
Output sparsity (%)	3	3
Percentage of potential synapses	100	100
Stimulus threshold	10	10
Increase in the permanence of active synapses	0.14	0.14
Decrease in the permanence of inactive synapses	0.02	0.02
Permanence threshold for a synapse to be consider connected (connected permanence)	0.95	0.30

We employed 10-fold cross validation strategy to evaluate the performance of SP on different subsets of our data. This strategy ensured that the obtained results were consistent and not dependent on particular data split [29]. In each iteration, we fed 9 folds of data to the SP for training and once trained, identified the indexes of synapses (associated with SDRs) whose permanence were $\geq$ connected permanence. These indexes were considered to be the indexes of selected features.

We compared the performance of SP as a feature selection algorithm against chi-squared, ANOVA F-test and mutual info scores based feature selection algorithms [9]. These algorithms were implemented using the *SelectKBest* method available in *scikit-learn*, python [30]. Additionally, we computed Jaccard stability index, Kuncheva index and Dice-Sorensen index [31] to measure the stability of SP-selected features.

Furthermore, we trained three classifiers namely Support Vector Machine (kernel set to radial biases function), Random Forest and Logistic Regressor [32] to evaluate the classification performance of SP-selected features. Accuracy, recall, precision and F1-score [33] were used to measure the classification performance. Default parameters were used to train these classifiers [30].

We employed Friedman test [34] to verify the statistical significance of stability indexes and performance scores. Bonferroni correction was further applied to overcome potential Type I errors.

IV. RESULTS AND DISCUSSION

The trained SPs generated SDRs with a fixed sparsity of 3%. Table III details the summary statistics of entropy of SPs, sparsity and activation frequency of SDRs before and after training. These statistics verified that SPs learnt well on the given data [16]. Figures 3 and 4 show the sparsity of the data and SP output for ABIDE and MDD dataset respectively.

TABLE III
SUMMARY STATISTICS OF TRAINED SPS BEFORE (BEF.) AND AFTER (AFT.) TRAINING

Dataset	**Measure**	**Time**	**Min**	**Mean**	**Std**	**Max**
ABIDE	Entropy	Bef.	-	0	-	-
		Aft.	-	0.9999	-	-
	Sparsity	Bef.	-inf	1234.57	35.1364	inf
		Aft.	0.03	0.03	$1.1104e^{-6}$	0.03
	Act. Freq	Bef.	1234.54	1234.56	0.0032	1234.57
		Aft.	0.0299	0.03	0.0004	0.0313
MDD	Entropy	Bef.	-	0	-	-
		Aft.	-	0.9985	-	-
	Sparsity	Bef.	-inf	1234.57	35.1364	inf
		Aft.	0.03	0.03	$2.5999e^{-6}$	0.03
	Act. Freq	Bef.	1234.54	1234.56	0.0032	1234.57
		Aft.	0.0238	0.02999	0.0033	0.0317

Table IV presents the stability indexes of employed feature selection methods for preprocessed ABIDE and MDD dataset.

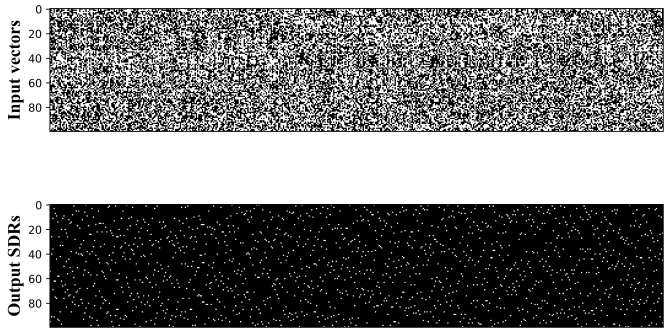

Fig. 3. **Sparsity in ABIDE.** The top plot shows the sparsity of the input, while the bottom plot represents the sparsity of the output SDRs.

Fig. 4. **Sparsity in MDD.** The top plot shows the sparsity of the input, while the bottom plot represents the sparsity of the output SDRs.

Friedman test revealed a significant difference among the stability indexes ($p < 0.05$, Bonferroni corrected). SP trained on ABIDE selected features that have stability close to other methods. However, in case of MDD, selected features seemed to be less stable. One possible reason behind this low stability is because of the small size of MDD dataset.

TABLE IV
MEAN STABILITY INDEXES OF FEATURE SELECTION ALGORITHMS OVER 10 FOLDS

Dataset	Stability index / Feature selection	Jaccard index	Kuncheva index	Dice-Sorensen index
ABIDE	**SP**	**0.5769**	**0.5238**	**0.7166**
	Chi-squared	0.6121	0.5979	0.7593
	ANOVA f-test	0.7170	0.7300	0.8351
	Mutual info	0.5549	0.5184	0.7137
MDD	**SP**	**0.2149**	**0.2967**	**0.3493**
	Chi-squared	0.4867	0.6869	0.6544
	ANOVA F-test	0.5926	0.8009	0.7436
	Mutual info	0.3852	0.5623	0.5560

Similarly, table V, VI and VII present the classification performance obtained with and without feature selection. Friedman test revealed no statistical difference among these performance scores ($p > 0.05$, Bonferroni corrected). It is evident from these scores that SP-selected features provided the same classification performance as that of other algorithms.

TABLE V
MEAN OF SVM PERFORMANCE SCORES OVER 10 FOLDS

Dataset	Performance score / Feature selection (FS)	Accuracy	Precision	Recall	F1 score
ABIDE	Without FS	0.6804	0.6839	0.6747	0.6777
	SP	**0.6765**	**0.6836**	**0.6649**	**0.6723**
	Chi-squared	0.6727	0.6753	0.6652	0.6690
	ANOVA F-test	0.6637	0.6668	0.6596	0.6611
	Mutual info	0.6586	0.6588	0.6587	0.6579
MDD	Without FS	0.5643	0.5532	0.5439	0.5345
	SP	**0.5857**	**0.5634**	**0.6251**	**0.5773**
	Chi-squared	0.5714	0.5757	0.5176	0.5286
	ANOVA F-test	0.5929	0.5857	0.5412	0.5451
	Mutual info	0.6143	0.5992	0.5773	0.5729

TABLE VI
MEAN OF RANDOM FOREST PERFORMANCE SCORES OVER 10 FOLDS

Dataset	Performance score / Feature selection (FS)	Accuracy	Precision	Recall	F1 score
ABIDE	Without FS	0.6178	0.6248	0.5857	0.6021
	SP	**0.6100**	**0.6115**	**0.6084**	**0.6071**
	Chi-squared	0.6241	0.6382	0.5971	0.6117
	ANOVA F-test	0.6497	0.6608	0.6165	0.6352
	Mutual info	0.6254	0.6385	0.5921	0.6104
MDD	Without FS	0.5500	0.5411	0.5419	0.5155
	SP	**0.6143**	**0.6004**	**0.6695**	**0.6064**
	Chi-squared	0.5857	0.5747	0.6245	0.5810
	ANOVA F-test	0.5786	0.5681	0.5787	0.5562
	Mutual info	0.6500	0.6277	0.6473	0.6201

TABLE VII
MEAN OF LOGISTIC REGRESSOR PERFORMANCE SCORES OVER 10 FOLDS

Dataset	Performance score / Feature selection (FS)	Accuracy	Precision	Recall	F1 score
ABIDE	Without FS	0.6548	0.6526	0.6538	0.6522
	SP	**0.6535**	**0.6577**	**0.6294**	**0.6425**
	Chi-squared	0.6624	0.6594	0.6638	0.6608
	ANOVA F-test	0.6547	0.6516	0.6557	0.6528
	Mutual info	0.6535	0.6537	0.6541	0.6517
MDD	Without FS	0.6357	0.5971	0.6830	0.6305
	SP	**0.6143**	**0.5809**	**0.6638**	**0.6095**
	Chi-squared	0.5643	0.5731	0.5608	0.5407
	ANOVA F-test	0.6000	0.6020	0.6296	0.5960
	Mutual info	0.6286	0.6096	0.6563	0.6138

V. CONCLUSION

We demonstrated the potential of using the HTM SP as a nonparametric interpretable feature selection algorithm

for high-dimensional neuroimaging data. Our introductory investigation suggests that the SP can select features with stability comparable to established methods when applied to the ABIDE dataset. However, the SP-selected features show lower stability for the smaller MDD dataset, indicating a potential sensitivity to sample size. Classification performance using SP-selected features is comparable to that of other feature selection algorithms across both datasets. We believe that our exploration could contribute to the ongoing efforts to develop robust and interpretable feature selection methods for neuroimaging data.

REFERENCES

[1] B. Mwangi, T. S. Tian, and J. C. Soares, "A review of feature reduction techniques in neuroimaging," *Neuroinformatics*, vol. 12, no. 2, p. 229–244, Apr. 2014.

[2] A. Chepurova, A. Hramov, and S. Kurkin, "Motor imagery: How to assess, improve its performance, and apply it for psychosis diagnostics," *Diagnostics*, vol. 12, no. 4, p. 949, 2022.

[3] O. E. Karpov, E. N. Pitsik, S. A. Kurkin, V. A. Maksimenko, A. V. Gusev, N. N. Shusharina, and A. E. Hramov, "Analysis of publication activity and research trends in the field of ai medical applications: Network approach," *International Journal of Environmental Research and Public Health*, vol. 20, no. 7, p. 5335, 2023.

[4] O. E. Karpov, V. V. Grubov, V. A. Maksimenko, S. A. Kurkin, N. M. Smirnov, N. P. Utyashev, D. A. Andrikov, N. N. Shusharina, and A. E. Hramov, "Extreme value theory inspires explainable machine learning approach for seizure detection," *Scientific Reports*, vol. 12, no. 1, p. 11474, 2022.

[5] E. A. Allen, E. B. Erhardt, and V. D. Calhoun, "Data visualization in the neurosciences: overcoming the curse of dimensionality," *Neuron*, vol. 74, no. 4, p. 603–608, May 2012.

[6] S. A. Kurkin, V. V. Grubov, V. A. Maksimenko, E. N. Pitsik, M. V. Khramova, and A. E. Hramov, "System for monitoring and adjusting the learning process of primary schoolchildren based on the eeg data analysis," *Informatsionno-upravliaiushchie sistemy [Information and Control Systems]*, no. 5, pp. 50–61, 2020.

[7] S. Kurkin, E. Pitsik, and N. Frolov, "Artificial intelligence systems for classifying eeg responses to imaginary and real movements of operators," in *Saratov Fall Meeting 2018: Computations and Data Analysis: from Nanoscale Tools to Brain Functions*, vol. 11067. SPIE, 2019, pp. 42–47.

[8] H. A. Chipman and H. Gu, "Interpretable dimension reduction," *Journal of Applied Statistics*, vol. 32, no. 9, p. 969–987, Nov. 2005.

[9] J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, and H. Liu, "Feature selection: A data perspective," *ACM Comput. Surv.*, vol. 50, no. 6, pp. 94:1–94:45, Dec. 2017.

[10] E. Thibeau-Sutre, S. Collin, N. Burgos, and O. Colliot, *Interpretability of Machine Learning Methods Applied to Neuroimaging*. New York, NY: Springer US, 2023, p. 655–704. [Online]. Available: https://doi.org/10.1007/978-1-0716-3195-9_22

[11] N. Frolov, M. S. Kabir, V. Maksimenko, and A. Hramov, "Machine learning evaluates changes in functional connectivity under a prolonged cognitive load," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 31, no. 10, 2021.

[12] M. S. Kabir and S. Kurkin, "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. 63–67.

[13] M. S. Kabir, S. Kurkin, G. Portnova, O. Martynova, Z. Wang, and A. Hramov, "Contrastive machine learning reveals in eeg resting-state network salient features specific to autism spectrum disorder," *Chaos, Solitons & Fractals*, vol. 185, p. 115123, Aug. 2024.

[14] X. Chen, W. Wang, and W. Li, "An overview of hierarchical temporal memory: A new neocortex algorithm," in *2012 Proceedings of International Conference on Modelling, Identification and Control*, Jun. 2012, p. 1004–1010. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6260285

[15] J. Hawkins, S. Ahmad, and Y. Cui, "A theory of how columns in the neocortex enable learning the structure of the world," *Frontiers in Neural Circuits*, vol. 11, p. 81, Oct. 2017.

[16] Y. Cui, S. Ahmad, and J. Hawkins, "The htm spatial pooler—a neocortical algorithm for online sparse distributed coding," *Frontiers in Computational Neuroscience*, vol. 11, Nov. 2017. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fncom.2017.00111

[17] S. Ahmad and J. Hawkins, "How do neurons operate on sparse distributed representations? a mathematical theory of sparsity, neurons and active dendrites," no. arXiv:1601.00720, May 2016, arXiv:1601.00720 [cs, q-bio]. [Online]. Available: http://arxiv.org/abs/1601.00720

[18] S. Crochet, J. F. A. Poulet, Y. Kremer, and C. C. H. Petersen, "Synaptic mechanisms underlying sparse coding of active touch," *Neuron*, vol. 69, no. 6, p. 1160–1175, Mar. 2011.

[19] D. O. Hebb, *The organization of behavior; a neuropsychological theory*, ser. The organization of behavior; a neuropsychological theory. Oxford, England: Wiley, 1949.

[20] S. Song, K. D. Miller, and L. F. Abbott, "Competitive hebbian learning through spike-timing-dependent synaptic plasticity," *Nature Neuroscience*, vol. 3, no. 9, p. 919–926, Sep. 2000.

[21] C. Craddock, Y. Benhajali, C. Chu, F. Chouinard, A. Evans, A. Jakab, B. S. Khundrakpam, J. D. Lewis, Q. Li, M. Milham *et al.*, "The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives," *Frontiers in Neuroinformatics*, vol. 7, no. 27, p. 5, 2013.

[22] E. N. Pitsik, V. A. Maximenko, S. A. Kurkin, A. P. Sergeev, D. Stoyanov, R. Paunova, S. Kandilarova, D. Simeonova, and A. E. Hramov, "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*, vol. 167, p. 113041, Feb. 2023.

[23] M. Khodatars, A. Shoeibi, D. Sadeghi, N. Ghaasemi, M. Jafari, P. Moridian, A. Khadem, R. Alizadehsani, A. Zare, Y. Kong, A. Khosravi, S. Nahavandi, S. Hussain, U. R. Acharya, and M. Berk, "Deep learning for neuroimaging-based diagnosis and rehabilitation of autism spectrum disorder: A review," *Computers in Biology and Medicine*, vol. 139, p. 104949, Dec. 2021.

[24] D. Stoyanov, V. Khorev, R. Paunova, S. Kandilarova, D. Simeonova, A. Badarin, A. Hramov, and S. Kurkin, "Resting-state functional connectivity impairment in patients with major depressive episode," *International Journal of Environmental Research and Public Health*, vol. 19, no. 21, p. 14045, 2022.

[25] A. V. Andreev, S. A. Kurkin, D. Stoyanov, A. A. Badarin, R. Paunova, and A. E. Hramov, "Toward interpretability of machine learning methods for the classification of patients with major depressive disorder based on functional network measures," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 33, no. 6, 2023.

[26] A. N. Pisarchik, A. V. Andreev, S. A. Kurkin, D. Stoyanov, A. A. Badarin, R. Paunova, and A. E. Hramov, "Topology switching during window thresholding fmri-based functional networks of patients with major depressive disorder: Consensus network approach," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 33, no. 9, 2023.

[27] A. M. Bastos and J.-M. Schoffelen, "A tutorial review of functional connectivity analysis methods and their interpretational pitfalls," *Frontiers in Systems Neuroscience*, vol. 9, Jan. 2016. [Online]. Available: https://www.frontiersin.org/journals/systems-neuroscience/articles/10.3389/fnsys.2015.00175/full

[28] M. Otahal, D. Keeney, D. McDougall *et al.*, "HTM.core implementation of Hierarchical Temporal Memory," https://github.com/htm-community/htm.core/, 2019.

[29] T. Fushiki, "Estimation of prediction error by using k-fold cross-validation," *Statistics and Computing*, vol. 21, no. 2, p. 137–146, Apr 2011. [Online]. Available: https://doi.org/10.1007/s11222-009-9153-8

[30] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, A. Muller, J. Nothman, G. Louppe, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in python," no. arXiv:1201.0490, Jun. 2018, arXiv:1201.0490 [cs]. [Online]. Available: http://arxiv.org/abs/1201.0490

[31] U. M. Khaire and R. Dhanalakshmi, "Stability of feature selection algorithm: A review," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 4, p. 1060–1073, Apr. 2022.

[32] A. C. Müller and S. Guido, *Introduction to Machine Learning with Python: A Guide for Data Scientists*. O'Reilly Media, Inc., Sep. 2016.
[33] H. M and S. M.N, "A review on evaluation metrics for data classification evaluations," *International Journal of Data Mining & Knowledge Management Process*, vol. 5, no. 2, p. 01–11, Mar. 2015.
[34] M. R. Sheldon, M. J. Fillyaw, and W. D. Thompson, "The use and interpretation of the friedman test in the analysis of ordinal-scale data in repeated measures designs," *Physiotherapy Research International*, vol. 1, no. 4, p. 221–228, 1996.