

Selection and Stability of Functional Connectivity Features for Classification of Brain Disorders

Aniruddha Saha, Soujanya Hazra, and Sanjay Ghosh, *Senior Member, IEEE*

Abstract—Brain disorders are an umbrella term for a group of neurological and psychiatric conditions that have a major effect on thinking, feeling, and acting. These conditions encompass a wide range of conditions. The illnesses in question pose significant difficulties not only for individuals, but also for healthcare systems all across the world. In this study, we explore the capability of explainable machine learning for classification of people who suffer from brain disorders. This is accomplished by the utilization of brain connection map, also referred as *connectome*, derived from functional magnetic resonance imaging (fMRI) data. In order to analyze features that are based on the connectome, we investigated several different feature selection procedures. These strategies included the Least Absolute Shrinkage and Selection Operator (LASSO), Relief, and Analysis of Variance (ANOVA), in addition to a logistic regression (LR) classifier. First and foremost, the purpose was to evaluate and contrast the classification accuracy of different feature selection methods in terms of distinguishing healthy controls from sick or injured individuals. The evaluation of the stability of the traits that were chosen was the second objective. The identification of the regions of the brain that have an effect on the classification was the third main objective. When applied to the UCLA dataset, the LASSO approach, which is our most effective strategy, produced a classification accuracy of 91.85% and a stability index of 0.74, which is greater than the results obtained by other approaches: Relief and ANOVA. These methods are effective in locating trustworthy biomarkers, which adds to the development of connectome-based classification in the context of issues that impact the brain.

Index Terms—Biomarker, brain disorders, feature stability, functional connectivity, logistic regression classifier.

I. INTRODUCTION

BRAIN disorders represent a diverse group of neurological and psychiatric conditions that severely impact an individual's cognitive functions, emotional regulation, and perception of reality [1]. Characterized by symptoms such as cognitive impairments, mood disturbances, abnormal behavior, and disruptions in sensory processing, these conditions pose complex challenges in both diagnosis and treatment [2]. The exact etiologies of many brain disorders remain elusive, although it is widely acknowledged that a multifactorial interplay of genetic, environmental, and neurobiological factors contributes to their onset [3]. Brain connectivity map or connectome is distinct in several ways, presenting both unique challenges and opportunities when applied in machine learning. Unlike conventional medical images, which have a grid-like structure where each pixel is only adjacent to nearby

pixels, connectomes possess a more complex topology. In connectome data, each brain region, represented as a node in the network, can be connected to any other region, either structurally or functionally [4]. It is a matrix-like structure where the size of the matrix, i.e., rows or columns, represents the number of brain regions or Region of Interest (ROI) under study. These connections are represented by binary or weighted edges between nodes. This network structure encapsulates essential information about brain connectivity that is not directly captured by other imaging modalities.

Structural magnetic resonance imaging (sMRI) is a popular tool for studying brain alterations, since it offers comprehensive anatomical images of the brain [5]. It makes it possible to build structural connectomes, which record the brain's physical structure. Functional MRI (fMRI) measures blood-oxygen-level-dependent signals, which indicate regional blood flow changes as proxies for neuronal co-activation [4]. The synchronized activations across spatially different brain areas constitute the foundation of functional connectivity (FC). It is a network-level representation commonly employed to investigate illnesses including Alzheimer's disease [6], major depressive disorder [7], autism [8], and schizophrenia [4]. Although fMRI is the predominant method for FC analysis, alternative techniques like electroencephalography (EEG) [7] and magnetoencephalography (MEG) [9] provide supplementary insights with superior temporal resolution. FC data is highly dimensional and complicated, so successful classification models require great care in feature selection and stability. It needs to make sure the biomarkers found are reliable and have an impact on the nervous system.

In this paper, we investigate three different feature selection procedures to classify brain disorders based on functional connectivity. We also analyze the stability of these selected features in terms of how consistently the same features are selected across different iterations, datasets, or models. Our research expands on previous efforts to develop stable biomarkers in brain connectomes [10] to assess the classification of brain disease based on functional connectivity. Unlike classification-only approaches, we prioritize feature stability, interpretability, and reproducible findings. In particular, the key contributions of this paper are:

- We evaluate and compare three popular feature selection algorithms, LASSO, Relief, and ANOVA, on two fMRI datasets, analyzing a total of 16,769 functional connectivity features per subject (3,403 in Dataset 1 and 13,366 in Dataset 2).
- We specifically quantify feature stability across cross-validation folds using the *Kuncheva index* and *Jaccard index* to ensure biomarker discovery reproducibility. The

This research is supported by the Faculty Start-up Research Grant (FSRG), IIT Kharagpur, awarded to Dr. Sanjay Ghosh.

Aniruddha Saha, Soujanya Hazra, and Sanjay Ghosh are with the Department of Electrical Engineering, Indian Institute of Technology Kharagpur, WB 721302, India (e-mail: sanjay.ghosh@ee.iitkgp.ac.in).

average stability was achieved by LASSO (*Kuncheva*: 0.74 and *Jaccard*: 0.69).

- We illustrate the generalizability of our findings on two different fMRI datasets with a total of 226 subjects (54 in Dataset 1 and 172 in Dataset 2), demonstrating consistent classification accuracy and feature robustness trends. LASSO achieved up to 91.85% accuracy and 91.98% F1-score.

This paper is summarized as follows. Section II discusses neuro-imaging functional connectivity analysis, feature selection, and stability assessment. Section III covers the datasets, feature extraction pipeline, classification model, feature selection techniques, and stability metrics. Section IV presents the experimental setup, classification findings, and feature stability analysis across datasets. Section V analyzes explainability by identifying consistently identified brain areas that contribute to classification. Section VI closes the paper and suggests future research.

II. RELATED WORK

Over the past decade, numerous studies have investigated the correlation between brain functional connectivity (FC) and various neurological and mental problems. Researchers identified, selected, and interpreted neuro-imaging characteristics from fMRI [4], EEG [7], and MEG [9] for diagnostic and prognostic reasons using machine learning and deep learning. These systems demonstrate that feature selection, graph-based representations, and multi-modal fusion improve brain disease accuracy and interpretability. Qu et al. [11] introduced a deep graphical approach to extract distinguishing features from functional brain networks. A nonlinear network fusion method using message-passing mechanisms was used to assess multimodal brain connectivity [12]. Jie et al. [13] improved categorization using topological and statistical FC features across neuroimaging tasks. According to the authors in this study [14], spontaneous emotional brain states evolve and affect mental health. Tawhid and colleagues introduced GENet [15], a generic neural model for EEG-based neurological diseases. Dai et al. [16] suggested a transfer learning method that aligns features across domains to enhance the classification of psychiatric disorders. Authors in [17] employed a region-selected graph convolutional network (GCN) to dynamically identify disorder-specific patterns instead of relying on fixed brain areas. A multi-scale dynamic graph technique to model FC fluctuations over time was introduced in another study [1]. Multimodal integration and graph-based learning have been the primary areas of attention in research about the classification of Alzheimer's disease. An approach to learning that is self-paced and makes use of structural and functional imaging was proposed in [6]. The application of EEG-based graph neural network (GNN) was demonstrated in [18], which showed that the selection of the FC metric had a considerable impact on classification performance. Cao and collaborators [19] introduced a dynamic connectivity-based approach for more precise detection of mild cognitive impairment (MCI). In a parallel study, [20] presented an ordinal pattern kernel that captures brain signal temporal patterns for MCI classification.

Kang et al. [21] identified depressed patients using residual neural networks trained on certain frequency bands and brain areas. Another study [22] presented an attention-based fusion model for depression diagnosis using multimodal MRI data. The work by Hasanzadeh and colleagues [7] utilized graph theory to study directed FC networks from EEG signals. Guo et al. [23] built whole-brain effective connection networks from resting-state fMRI to find reliable depression related patterns. The innovative schizophrenia diagnosis framework, Schizo-Net [24], uses multimodal deep learning and EEG-based connection biomarkers. A study [25] provided a strategy for selecting joint channels and connectivity applied to fNIRS data for stress detection in decision-making scenarios. The authors [8] utilized machine learning algorithms on MEG data to detect autism in children, with a focus on early diagnosis. In order to highlight the temporal dynamics linked to autism, Jamal et al. [26] used supervised learning to analyze synchrostate-derived EEG characteristics. The recent paper [27] examined how dynamic FC changes with age could predict brain age. Their connection analysis method may detect early cognitive deterioration.

Neuroimaging investigations require feature selection due to the connectome data's high dimensionality. Ji and Yao [28] suggested a CNN-graphical Lasso model for sparse, topologically significant features. The authors in [10] developed a stability-driven framework to find repeatable biomarkers from functional connectomes in schizophrenia, highlighting the necessity for rigorous feature selection. In the same direction, the study in [29] demonstrated the effectiveness of a feature selection technique based on support vector machines (SVMs) that uses the weight vector to categorize dementia in high-dimensional neuroimaging data. Jie et al. [30] used whole-brain FC and FOBA with SVM to distinguish bipolar disorder from serious depression. Using a majority-voting method, Holker and Susan [31] ranked EEG characteristics that were important for detecting alcohol use disorders. Ali et al. [32] introduced a correlation-filter-based method for selecting informative channels in EEG-fNIRS systems within a hybrid architecture. Using infinite feature selection, authors [33] were able to uncover changes in connection following a stroke. In the study [34], the authors utilized a hybrid method using feature fusion and selection to identify workloads based on EEG data. Functional connection in clinical neuroscience is shown by the wide range of illnesses and methods in the literature. Most of these studies favor classification accuracy over feature stability and reproducibility, which is crucial for reliable biomarkers. Our work assesses the discriminative strength and consistency of selected features across cross-validation folds using Kuncheva and Jaccard indices. This dual goal improves explainable AI brain illness diagnosis interpretability and reliability.

III. MATERIALS AND METHOD

In this paper, we have used a vectorized connectome as our input feature vector for our model. As the number of features is much larger than the data samples, effective feature selection must be used. Here we have evaluated feature selection techniques like embedded feature selection, filter-based

Symbols	Significance
i	Subject index
n	Total number of subjects.
x_i	Input feature of i -th subject.
d	Number of ROI's
$p = \frac{d(d-1)}{2}$	Length of each feature vector.
$x_{ij}, j = 1, 2, \dots, p$	j -th feature component of i -th subject.
$y_i \in \{0, 1\}$	Class of i -th subject.
-----	-----
$w \in \mathbb{R}^{q \times 1}$	Weight vector in LR computation.
b	Intercept parameter in LR computation.
-----	-----
α	Intercept parameter in LASSO computation.
β	Weight assignment vector of length p in LASSO.
λ	Regularization parameter in LASSO.
$v \in \mathbb{R}^{p \times 1}$	Weight assignment vector of length p in Relief.
F_j	F-Score for j -th feature using ANOVA.

TABLE I

SUMMARY OF THE VARIABLES AND DEFINITIONS USED IN THIS PAPER.

feature selection, and wrapper-based feature selection. We have used logistic regression (LR) as our backbone classifier for classifying Schizophrenia from healthy patients.

A. Datasets

Here we have used two types of datasets, for our experiment, the descriptions of which are given below:

1) *Dataset 1*: The dataset used in our experiments is downloaded from [35]. The cohort consists of a schizophrenic group of 27 subjects and a control group of 27 healthy subjects. All of the data was collected at the Service of General Psychiatry at the Lausanne University Hospital. The schizophrenic patients were diagnosed with schizophrenia and schizoaffective disorders after meeting the DSM-IV criteria. Control subjects had no history of neurological disease. All 54 subjects had given written consent following the institutional guidelines approved by the Ethics Committee of Clinical Research of the Faculty of Biology and Medicine, University of Lausanne, Switzerland.

A 3 Tesla Siemens Trio scanner equipped with a 32-channel head coil was used to acquire imaging data from all participants. Functional imaging was performed using a gradient echo planar imaging (EPI) sequence optimized for BOLD contrast, with an in-plane resolution and slice thickness of 3.3 mm, along with a 0.3-mm gap. The echo time (TE) was 30 ms, and the repetition time (TR) was 1,920 ms, resulting in 280 images per subject. Functional connectomes were derived from fMRI BOLD time-series by computing the absolute Pearson correlation between the time-courses of individual brain regions. Here we have considered 83 regions of interest(ROI) for our study.

2) *Dataset 2*: This dataset consists of fMRI data acquired at the University of California, Los Angeles (UCLA) Consortium for Neuropsychiatric Phenomics LA5c study, which contains

50 schizophrenic subjects and 122 control subjects [36]. Each fMRI data set, having $60 \times 60 \times 30 = 108,000$ voxels is preprocessed through: functional realignment, slice timing correction, outlier detection, segmentation, normalization, functional smoothing, and temporal bandpass filtering. The brain is parcellated into 164 regions-of-interest (ROI) in accordance with AAL3 Atlas [37] and Harvard-Oxford Cortical Atlas [38]. The connectivity between each pair of ROIs is measured in terms of Pearson correlation coefficient (PCC) [39], [40]. The finally-processed functional connectomes are publicly available from Sunil et al. [41].

B. Logistic Regression

Here we have used logistic regression [42] as our backbone classifier. Because it works well when the dataset is small to moderately sized and doesn't require large computational resources [43]. Given a set of data examples $\mathbf{x}_i \in \mathbb{R}^p, i = 1, \dots, n$, and a vector $\mathbf{y} \in \{0, 1\}^n$ representing the group membership of data points, logistic regression aims to model the probability that each data point belongs to a particular class. The logistic regression model finds the best-fitting hyperplane that separates the two classes by maximizing the likelihood of the observed data. The mathematical formulation can be written as a maximum likelihood estimation problem:

$$\max_{\mathbf{w}, b} \sum_{i=1}^n (y_i \log(\sigma(\mathbf{w}^T \mathbf{x}_i + b)) + (1 - y_i) \log(1 - \sigma(\mathbf{w}^T \mathbf{x}_i + b))), \quad (1)$$

where $\sigma(z) = \frac{1}{1+e^{-z}}$ is the logistic function, $\mathbf{w} \in \mathbb{R}^p$ is a vector of coefficients, and b is the intercept term. A classifier that generalizes well is then found by controlling the complexity of the model, often by regularizing $\mathbf{w}$ to prevent overfitting, typically using L_2 regularization. For the sake of clarity, we have summarized the symbols in Table I.

C. Feature selection algorithms

1) *LASSO*: In this work, we utilize the least absolute shrinkage and selection operator (LASSO) as our method for both regularization and feature selection. LASSO [44] is particularly effective in scenarios with high-dimensional data, where the number of features might exceed the number of observations.

LASSO modifies the ordinary least squares objective by adding a regularization term that is the L_1 norm of the coefficients. This addition encourages sparsity in the coefficients of the model, effectively performing feature selection by driving some coefficients to exactly zero. The mathematical formulation for LASSO is given by an unconstrained optimization problem:

$$\min_{\beta} \left\{ \frac{1}{2n} \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j \mathbf{x}_{ij})^2 + \lambda \sum_{j=1}^p |\beta_j| \right\} \quad (2)$$

The choice of λ in (2) is critical in LASSO as it determines the level of regularization; higher values lead to greater regularization. Typically, the regularization parameter is selected

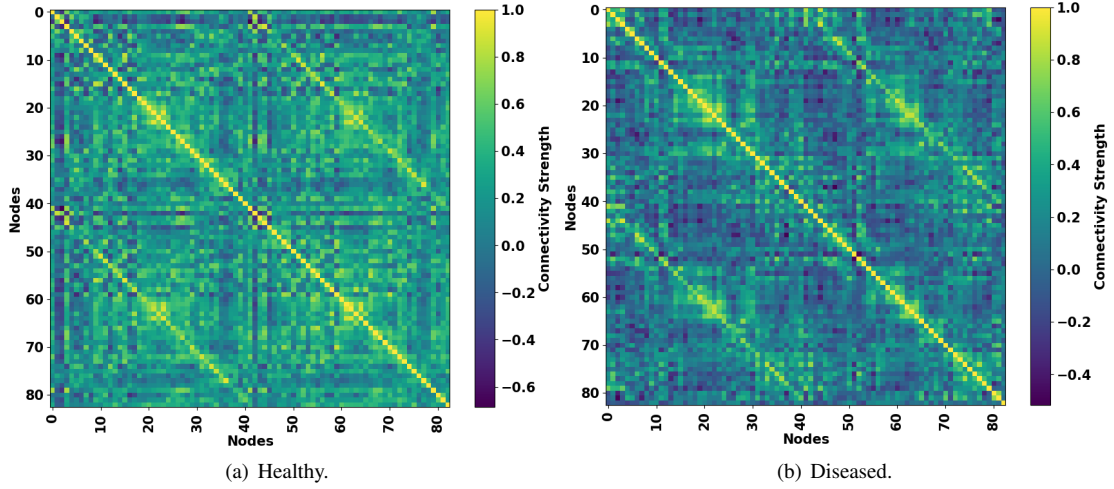


Fig. 1. Example of functional connectomes of healthy and diseased subjects from Dataset 1 (provided by University of Lausanne, Switzerland). A symmetric matrix shows the pairwise Pearson correlation between 83 brain areas in each subject’s connectome. The matrix values are thresholded absolute correlations; warmer colors indicate greater functional relationships.

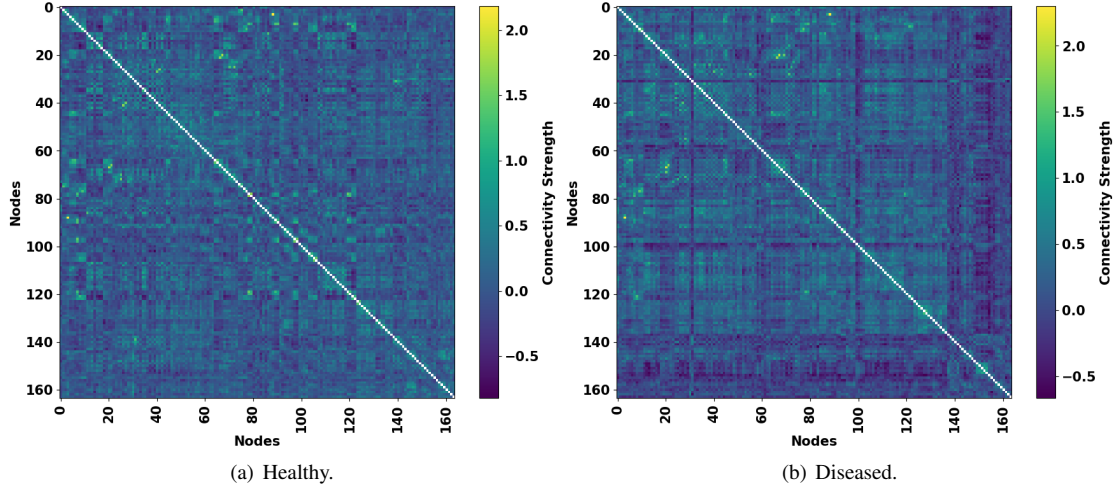


Fig. 2. Example of functional connectomes of healthy and diseased subjects from Dataset 2 (provided by UCLA, USA). The matrices denote the pairwise association among 164 brain regions. These high-resolution connectomes reveal more intricate inter-regional interactions than Dataset 1. The visual disparities between the two groups demonstrate fundamental abnormalities in connection patterns pertinent to classification.

through a cross-validation approach to balance between model complexity and data fidelity.

2) *Relief*: Given a randomly selected subject x_l , we first find the *hit set* $\mathcal{H}$ and *miss set* $\mathcal{M}$ of it. For example, $\mathcal{H}$ is the collection of the rest of the healthy subjects if the randomly selected subject x_l is healthy. Similarly, $\mathcal{M}$ is the collection of all diseased subjects if the subject x_l is diseased.

$$x_l^{hit} = \operatorname{argmin}_{x_i \in \mathcal{H}} \|x_l - x_i\|^2. \quad (3)$$

$$x_l^{miss} = \operatorname{argmin}_{x_i \in \mathcal{M}} \|x_l - x_i\|^2. \quad (4)$$

We initialize the weights v_j for $j = 1, 2, \dots, p$ with zeros. Then the aggregate weight for the j -th feature is given by

$$v_j = v_j - (x_{lj} - x_{lj}^{hit})^2 + (x_{lj} - x_{lj}^{miss})^2, \quad (5)$$

where x_{lj} is the j -th feature component of l -th subject, x_{lj}^{hit} is the j -th feature component of nearest neighbor from its hit set $\mathcal{H}$, and x_{lj}^{miss} is the j -th feature component of Euclidean distance nearest neighbor from its miss set $\mathcal{M}$.

Notice that the weight of any given feature decreases if it differs from that feature in nearby instances of the same class more than in nearby instances of the other class, and increases in the reverse case. After a certain number of iterations, divide each element of the weight vector by the number of iterations. This becomes the relevance vector. There is no hard stopping condition based on convergence or accuracy; instead, the number of iterations is a hyperparameter that controls how well the feature relevance estimates generalize.

3) *Analysis of Variance (ANOVA)*: Analysis of variance (ANOVA) compares the variance between groups (i.e., different classes) and the variance within groups (i.e., samples in the same class). The goal is to determine whether a feature’s values vary significantly across different classes. The key steps

are as follows:

- 1) Calculate the mean of each class: For a given j -th feature, the mean value for each class is computed as:

$$\begin{aligned} x_{0j}^H &= \frac{1}{n_H} \sum_{i=1}^{n_H} x_{ij}, & x_{1j}^D &= \frac{1}{n_D} \sum_{i=1}^{n_D} x_{ij}, \\ \bar{x}_{jj} &= \frac{1}{n} \sum_{i=1}^n x_{ij}, & n &= n_H + n_D, \end{aligned} \quad (6)$$

where

- n_H is the number of healthy subjects (class 0).
- n_D is the number of diseased subjects (class 1).
- n is the total number of subjects.
- x_{0j}^H is the mean of j -th feature-component for class 0 which is of healthy subjects.
- x_{1j}^D is the mean of j -th feature-component for class 1 which is of diseased subjects.
- $\bar{x}_{jj}$ is the mean j -th feature-component for all subjects.

- 2) Compute between-class variance (SSB):

The variance between classes, for the j th feature, also called the Sum of Squares Between (SSB), is given by:

$$SSB_j = n_H(x_{0j}^H - \bar{x}_{jj})^2 + n_D(x_{1j}^D - \bar{x}_{jj})^2$$

For binary classification, we have mean square between groups (MSB) for the j -th component of the feature vector is:

$$MSB_j = SSB_j.$$

- 3) Compute within-class variance: For the j -th feature, the variance within each class, referred to as, sum of squares within (SSW), is:

$$SSW_j = \sum_{i=1}^{n_H} (x_{ij} - x_{0j}^H)^2 + \sum_{i=1}^{n_D} (x_{ij} - x_{1j}^D)^2$$

For binary classification, we have the mean square within groups (MSW) for j -th component of feature vector as:

$$MSW_j = \frac{SSW_j}{n - 1}$$

- 4) Compute F-statistic: Finally, the F-statistic for the j -th component of the feature vector is calculated as follows:

$$F_j = \frac{MSB_j}{MSW_j}. \quad (7)$$

A higher F_j value indicates that the feature is more useful for classification.

D. Stability Analysis

The stability of selected features is a critical aspect in feature selection tasks. It measures how consistently the same features are selected across different iterations, datasets, or models. A stable feature selection method indicates robustness and reliability. Here we have used the Kuncheva stability index and the Jaccard Index to assess the stability of the features that are selected by the algorithms.

1) *Kuncheva Index*: The Kuncheva index (KI) [45] is a statistical measure used to assess the stability and consistency of feature selection across different subsets of data in machine learning and data analysis. In our experiments, we examine the stability of total p components in an FC feature of a particular subject. We exclude self-loops, which means that the diagonal non-zero values of the connectivity matrices are discarded.

Suppose that two subsets of feature components are j_θ and j_ϕ . Each subset has k elements. The total number of components in the feature vector of each subject is p . The larger the intersection between the subsets, the higher the value of the consistency in selection of the feature-component. Let r be the cardinality of the intersection of the two subsets. The maximum r equals k is achieved when both j_θ and j_ϕ are identical subsets. The Kuncheva index (KI) [45] is given by:

$$K(j_\theta, j_\phi) = \frac{r - (k^2/p)}{k - (k^2/p)}, \quad (8)$$

where $r = |j_\theta \cap j_\phi|$ and $k = |j_\theta| = |j_\phi|$, the size of each subset of feature components.

In our experiments, we draw $\lambda = 100$ subsamplings to obtain a respective feature signature, i.e. sequence of indices of selected features. The overall stability index [45] for a sequence of signatures is defined as the average of all pairwise stability indices on k subsamplings:

$$K_{\text{avg}} = \frac{2}{\lambda(\lambda - 1)} \sum_{\theta=1}^{\lambda-1} \sum_{\phi=\theta+1}^{\lambda} K(j_\theta, j_\phi). \quad (9)$$

Note that a higher value of K in (8) indicates a higher stability.

2) *Jaccard Index*: Jaccard's index (JI) [46] measures the average similarity from all pairwise selected feature subsets:

$$J_0(j_\theta, j_\phi) = \frac{|j_\theta \cap j_\phi|}{|j_\theta \cup j_\phi|} \quad (10)$$

$$J_{\text{avg}} = \frac{2}{\lambda(\lambda - 1)} \sum_{\theta=1}^{\lambda-1} \sum_{\phi=\theta+1}^{\lambda} J(j_\theta, j_\phi). \quad (11)$$

The stability index J_0 in (10) lies in the range $[0, 1]$. A high value close to 1 implies the algorithm is stable [47].

IV. EXPERIMENTAL RESULTS

A. Experimental Settings

1) *Nested 5-Fold Cross-Validation Framework*: The process utilizes two levels of cross-validation: the external loop for unbiased performance estimation and the internal loop for model tuning and parameter optimization [10]. We use the cross-validation experimental scheme as in [10].

External Loop: The dataset is divided into 5 folds. This external CV loop is shuffled 20 times, generating 100 subsamplings of the original dataset. In each shuffle, 80% of the data (four folds) serves as the training set for the internal CV, while the remaining 20% (one fold) acts as the testing set to provide an unbiased evaluation of the model.

Internal Loop: Within the training set of the external CV, the internal CV loop is applied to further split the data into training and validation subsets. Four folds are used as the training set, and one fold (held-out) is used as the validation set to tune model parameters and select features.

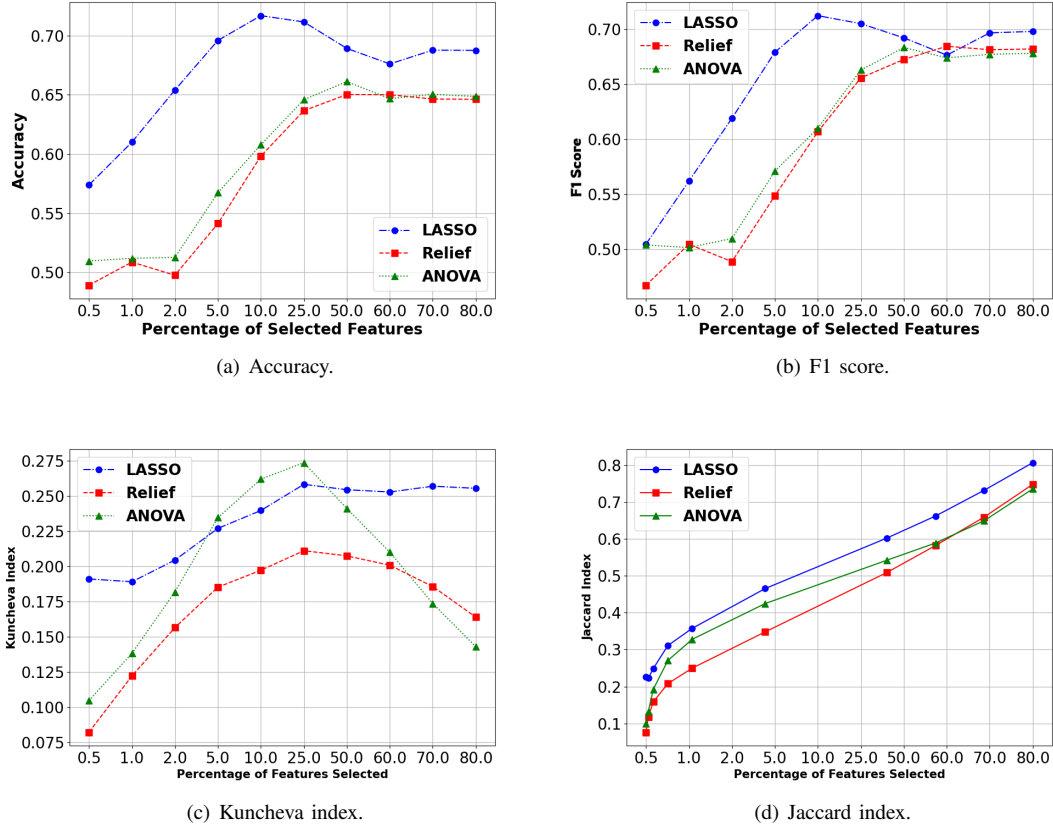


Fig. 3. Classification performance and feature selection stability on Dataset 1 (University of Lausanne). This dataset includes 54 functional connectomes vectorized into 3,403 characteristics from 27 healthy controls and 27 schizophrenic patients. LASSO, Relief, and ANOVA are tested at different percentages of selected characteristics in subfigures. LASSO consistently surpasses others in classification accuracy and F1 score. Kuncheva index and Jaccard index assess feature stability across 100 cross-validation subsamplings. LASSO has the highest stability across both criteria, indicating more reproducible biomarker selection.

Methods	Accuracy	Precision	Recall	F1-Score
LASSO	70	72.40	67.27	66.42
Relief	54.12	54.01	58.27	55
ANOVA	56.32	57.43	60.36	56.25

TABLE II

DATASET 1. CLASSIFICATION USING TOP 5% COMPONENTS IN FC.

Methods	Accuracy	Precision	Recall	F1-Score
LASSO	71.67	73.49	71.67	71.26
Relief	60	59.46	64.87	60.56
ANOVA	60.55	61.74	64.47	61.01

TABLE III

DATASET 1. CLASSIFICATION USING TOP 10% COMPONENTS IN FC.

2) *Data Augmentation*: We have used the synthetic minority oversampling technique (SMOTE) [48] for augmenting the data in Dataset 2. The main principle behind SMOTE is, it selects samples from the minority class, determines the k-nearest neighbors, and finally creates a new sample by interpolating between a selected sample and one of its neighbors:

$$x_{\text{new}} = x_{\text{original}} + \lambda(x_{\text{neighbor}} - x_{\text{original}}) \quad (12)$$

where λ is a random number between 0 and 1. The above data augmentation step is for class balancing in Dataset 2.

B. Results on Classification of Brain Disorders

In this section, we present a detailed analysis of the performance metrics derived from various feature selection

techniques applied within our study. In particular, we evaluate the accuracy of the models that incorporate these feature selection methods: LASSO, Relief, and ANOVA. To quantitatively measure the accuracy, we plotted the classification accuracies for each feature selection method. This allowed us to compare how each method impacts the predictive performance of the models. The graphical representations were designed to provide a clear and comparative perspective of the efficacy of each feature selection approach under varying conditions.

1) *Dataset 1. University of Lausanne*:: This dataset has functional connectomes of 27 healthy subjects and 27 schizophrenia (diseased) patients. Each connectome has $d = 83$ brain regions; therefore, $\frac{83 \times (83-1)}{2} = 3,403$ components in the vectorised connectome of each subject. We first study the classification performance by three feature selection for

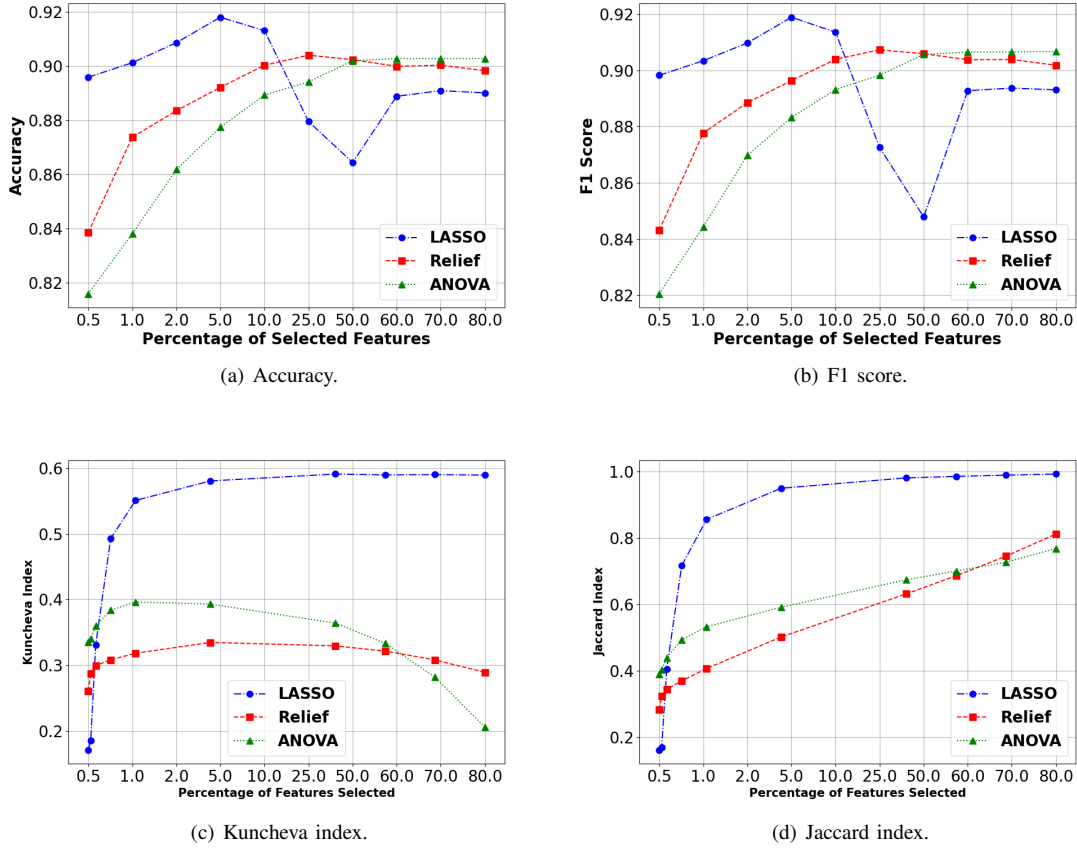


Fig. 4. Classification performance and feature selection stability on Dataset 2 (UCLA). The functional connectomes of 172 people (122 healthy controls and 50 schizophrenia patients) were vectorized into 13,366 features from 164 brain areas. The feature selection methods LASSO, Relief, and ANOVA are tested at different feature percentages. LASSO consistently outperforms in classification accuracy and F1 score. Kuncheva index and Jaccard index evaluate feature stability across 100 cross-validation subsamplings. LASSO has the strongest stability across both criteria, proving its reproducibility for large-scale connectome biomarker detection.

different percentage of selected components shown in Fig. 3. Notice in Fig. 3(a) and 3(b) that LASSO selection offers better accuracy and F1 score respectively than both Relief and ANOVA. In Table II, we report four metrics - accuracy, precision, recall, and F1 score when top 5% components are selected (around 170 from a total of 3403) to be input to the logistic regression classifier. Similarly, we report the results for top 10% components in Table III.

2) *Dataset 2. UCLA*:: This dataset has functional connectomes of 122 healthy subjects and 50 schizophrenia (diseased) patients. Each connectome has $d = 164$ brain regions; therefore, $\frac{164 \times (164 - 1)}{2} = 13,366$ components in the vectorised connectome of each subject. We first study the classification performance by three feature selection for different percentage of selected components shown in Fig. 4. Notice in Fig. 4(a) and 4(b) that LASSO selection offers better accuracy and F1 score respectively than both Relief and ANOVA when top 10% components are selected. In Table IV, we report four metrics - accuracy, precision, recall, and F1 score when top 5% components are selected (around 668 from a total of 13,366) to be input to the logistic regression classifier. Similarly, we report the results for the top 10% components in Table V.

In terms of classification accuracy and F1 score, LASSO

Methods	Accuracy	Precision	Recall	F1-Score
LASSO	91.85	92.22	91.79	91.98
Relief	89.25	87.73	92.02	88.89
ANOVA	87.95	85.58	91.61	88.98

TABLE IV
DATASET 2. CLASSIFICATION USING TOP 5% COMPONENTS IN FC.

consistently outperformed Relief and ANOVA across both datasets. LASSO used the top 10% of functional connectivity features to reach a peak accuracy of 71.67% and an F1 score of 71.26% for Dataset 1 (Tables II and III). With F1 values of 60% or below, relief and ANOVA fell behind. Using the top 5% of features, LASSO achieved 91.85% classification accuracy and 91.98% F1 score (Table IV); minimal variations were seen when increasing to 10% (Table V). These outcomes highlight the benefit of utilizing integrated techniques that simultaneously optimize feature selection and classification, such as LASSO.

C. Results on Stability of Features

In this section, we experiment on the stability in the feature selection process so that the method can retrieve consistent

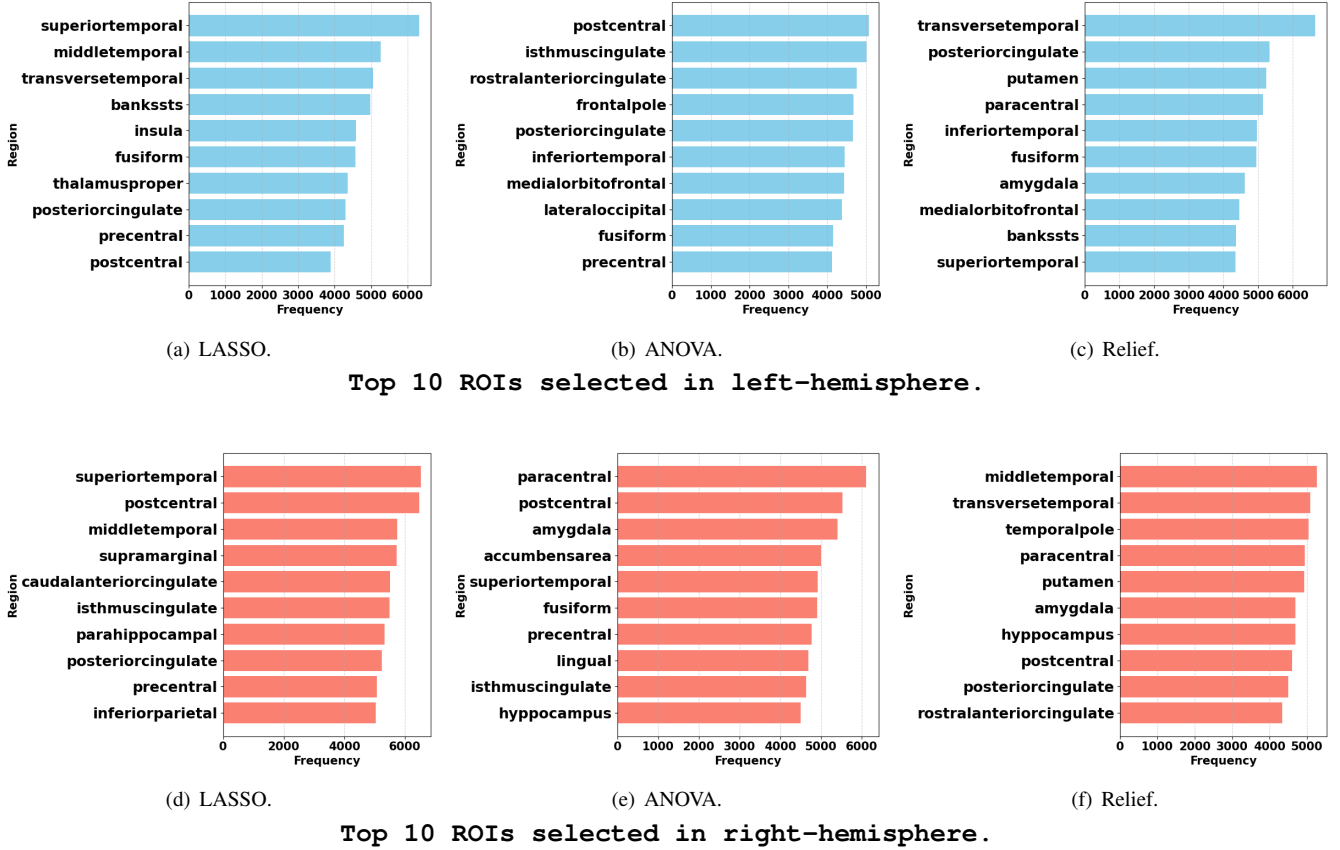


Fig. 5. Affected regions for dataset 1. In our experiment, we have 5-fold cross validation in each trial consisting of 7 different values of percentage of selected feature-component: [0.5, 1, 2, 5, 10, 25, 50]. We repeated the experiment over 10 trials. Therefore, for a particular brain region, the maximum possible occurrence will be: $10 \times 7 \times 5(d - 1) = 350 \times 82 = 28700$.

Methods	Accuracy	Precision	Recall	F1-Score
LASSO	91.25	91.75	91.22	91.56
Relief	90	88.48	92.67	90.56
ANOVA	89.12	87.43	91.61	89.14

TABLE V

DATASET 2. CLASSIFICATION USING TOP 10% COMPONENTS IN FC.

features across random subsamplings of the dataset. The stability of selected features is a critical aspect in feature selection tasks. This study helps us to measure how consistently the same features are selected across different iterations, datasets, or models. A stable feature selection method indicates robustness and reliability.

In the bottom row of Figure 3, we show the stability index as a function of the percentage of the feature-component selected for Dataset 1. Notice that LASSO and Relief selections have competitive performance in terms of the Kuncheva index, whereas LASSO consistently outperforms for the Jaccard index. In the bottom row of Figure 4, we show the stability index as a function of the percentage of the feature-component selected for Dataset 2. We found that LASSO consistently outperforms for indexes.

In addition to the graphical analysis, we present the mean stability ratings for each feature selection method based on

100 cross-validation subsamplings. On Dataset 1, LASSO had the most stability, with an average Kuncheva Index of around 0.69 and a Jaccard Similarity of 0.62. Relief achieved ratings of 0.64 (Kuncheva) and 0.59 (Jaccard), whereas ANOVA recorded scores of 0.61 and 0.55, respectively. In Dataset 2, the trend remained consistent, with LASSO producing the most stable feature subsets, with an average Kuncheva Index of 0.74 and a Jaccard Similarity of 0.69. Relief and ANOVA earned Kuncheva scores of 0.70 and 0.68, as well as Jaccard values of 0.65 and 0.63, respectively. These findings validate that LASSO not only promotes classification efficacy but also improves feature stability, hence affirming its appropriateness for reproducible biomarker identification.

V. IDENTIFICATION OF DISCRIMINATORY BRAIN REGIONS

In this work, we analyze how the varying brain regions influence the classification of brain disorders. We delve into a targeted analysis of the regions of interest (ROI) within the brain. By examining the degree of relevance attributed to specific ROIs, we identified what we refer to as the 'affected core'. This core consists of ROIs that consistently exhibit significant relevance, suggesting their critical role in the underlying condition being studied.

Following feature selection, we identified the most frequently selected brain regions across folds and random seeds.

Selected features, representing vectorized upper-triangular entries of the connectivity matrix, were mapped back to ROI pairs using the inverse of the standard upper-triangle index mapping. For each ROI, we computed its selection frequency as the number of times it appeared in any selected feature pair. From the top 10 ROIs selected by the feature selection methods in both hemispheres, we observed several regions with potential relevance to schizophrenia for both data sets in our experiments.

For the identification of 'affected core' regions, we apply an indicator function-based rule. In our experiment, we have 5-fold cross validation in each trial, consisting of 7 different values of percentage of selected feature-component: [0.5, 1, 2, 5, 10, 25, 50]. We repeated the experiment over 10 trials. Now, during feature selection at each trial and each validation, we update the frequency counter of regions r -th and t -th if the edge (r, t) is selected in the LASSO/Relief/ANOVA step. With initialization of $\delta_r = 0$ and $\delta_t = 0$, we update the frequency counter as follows:

$$\left. \begin{aligned} \delta_r &= \delta_r + 1 \\ \delta_t &= \delta_t + 1 \end{aligned} \right\} \text{ if edge } (r, t) \text{ is selected.} \quad (13)$$

In Fig. 5, we display the top 10 selected brain regions from the left hemisphere and the right hemisphere from Dataset 1 using LASSO, ANOVA, and Relief selections. Notice that for LASSO selection, the common (between left and right hemispheres) affected core regions are: (i) superior-temporal, (ii) middle-temporal, (iii) posterior-cingulate, (iv) pre-central, and (v) post-central. For ANOVA selection, the common 'affected core' brain regions are: post-central, isthmus-cingulate, and fusiform. For Relief selection, the common 'affected core' brain regions are: transverse-temporal, para-central, putamen, amygdala, and posterior-cingulate. Now considering higher classification performance and stability index of LASSO selection, we infer that the top 5 'affected core' brain regions are: superior-temporal, middle-temporal, posterior-cingulate, pre-central, and post-central.

VI. CONCLUSION

In this research, we studied the effectiveness of feature selection algorithms for functional connectivity-based disease categorization in the brain. We tested LASSO, Relief, and ANOVA on vectorized fMRI connectomes for classification and feature stability throughout cross-validation folds. We showed that LASSO consistently has the best accuracy, reproducibility trade-off using stability criteria like the *Kuncheva index* and *Jaccard index*. We found that mapping selected attributes to brain regions of interest enhanced both interpretability and performance evaluation. This allowed us to discover consistently selected and analytically acceptable brain regions that may contribute to disorder-related neural patterns. These results enable us to make connectome-based biomarker detection models clearer and dependable. Our results are validated using two separate fMRI datasets, comprising a total of 226 participants, indicating adaptability in both classification accuracy and feature stability. In future research, we intend to investigate the integration of multimodal data

sources, including EEG and MEG, to capture and enhance temporal dynamics. Furthermore, broadening this approach to include multiclass classification scenarios and other clinical populations (e.g., ADHD, bipolar disorder, early cognitive impairment) could further augment its application. Ultimately, integrating neurobiological priors or attention mechanisms into the feature selection process could produce more interpretable and robust models for therapeutic application.

ACKNOWLEDGMENT

The authors are grateful to the researchers who made the open-access neuroimaging datasets [10], [36], [41] that were used in this work. The University of Lausanne, Switzerland, shared Dataset 1, and UCLA made Dataset 2 open to everyone. These datasets made it possible to do this study without the collection of clinical data.

REFERENCES

- [1] Y. Ma, Q. Wang, L. Cao, L. Li, C. Zhang, L. Qiao, and M. Liu, "Multi-scale dynamic graph learning for brain disorder detection with functional MRI," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 3501–3512, 2023.
- [2] H. A. Whiteford, L. Degenhardt, J. Rehm, A. J. Baxter, A. J. Ferrari, H. E. Erskine, F. J. Charlson, R. E. Norman, A. D. Flaxman, N. Johns, R. Burstein, C. J. Murray, and T. Vos, "Global burden of disease attributable to mental and substance use disorders: findings from the global burden of disease study 2010," *The Lancet*, vol. 382, no. 9904, pp. 1575–1586, 2013.
- [3] J. Zhang, F. Feng, T. Han, X. Gong, and F. Duan, "Detection of autism spectrum disorder using fMRI functional connectivity with feature selection and deep learning," *Cognitive Computation*, vol. 15, no. 4, pp. 1106–1117, 2023.
- [4] M.-E. Lynall, D. S. Bassett, R. Kerwin, P. J. McKenna, M. Kitzbichler, U. Muller, and E. Bullmore, "Functional connectivity and brain networks in schizophrenia," *Journal of Neuroscience*, vol. 30, no. 28, pp. 9477–9487, 2010.
- [5] L. Cocchi, I. H. Harding, A. Lord, C. Pantelis, M. Yucel, and A. Zalesky, "Disruption of structure–function coupling in the schizophrenia connectome," *NeuroImage: Clinical*, vol. 4, pp. 779–787, 2014.
- [6] X. Hao, R. Wang, Y. Guo, Y. Xiao, M. Yu, M. Wang, W. Chen, and D. Zhang, "Multimodal self-paced locality-preserving learning for diagnosis of alzheimer's disease," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 15, no. 2, pp. 832–843, 2022.
- [7] F. Hasanzadeh, M. Mohebbi, and R. Rostami, "Graph theory analysis of directed functional brain networks in major depressive disorder based on EEG signal," *Journal of Neural Engineering*, vol. 17, no. 2, p. 026010, 2020.
- [8] K. Barik, K. Watanabe, J. Bhattacharya, and G. Saha, "Functional connectivity based machine learning approach for autism detection in young children using MEG signals," *Journal of Neural Engineering*, vol. 20, no. 2, p. 026012, 2023.
- [9] H. Jin, F. Abdelnour, P. Verma, B. S. Sipes, S. S. Nagarajan, and A. Raj, "Bayesian inference of frequency-specific functional connectivity in MEG imaging using a spectral graph model," *Imaging Neuroscience*, vol. 2, pp. 1–24, 10 2024.
- [10] L. Gutiérrez-Gómez, J. Vohryzek, B. Chiêm, P. S. Baumann, P. Conus, K. D. Cuenod, P. Hagmann, and J.-C. Delvenne, "Stable biomarker identification for predicting schizophrenia in the human connectome," *NeuroImage: Clinical*, 2020.
- [11] N. Chen, M. Guo, Y. Li, X. Hu, Z. Yao, and B. Hu, "Estimation of discriminative multimodal brain network connectivity using message-passing-based nonlinear network fusion," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 20, no. 4, pp. 2398–2406, 2021.
- [12] G. Qu, W. Hu, L. Xiao, J. Wang, Y. Bai, B. Patel, K. Zhang, and Y.-P. Wang, "Brain functional connectivity analysis via graphical deep learning," *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 5, pp. 1696–1706, 2021.

- [13] B. Jie, D. Zhang, W. Gao, Q. Wang, C.-Y. Wee, and D. Shen, "Integration of network topological and connectivity properties for neuroimaging classification," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 2, pp. 576–589, 2013.
- [14] P. A. Kragel, A. R. Hariri, and K. S. LaBar, "The temporal dynamics of spontaneous emotional brain states and their implications for mental health," *Journal of Cognitive Neuroscience*, vol. 34, no. 5, pp. 715–728, 2022.
- [15] M. N. A. Tawhid, S. Siuly, K. Wang, and H. Wang, "GENet: a generic neural network for detecting various neurological disorders from EEG," *IEEE Transactions on Cognitive and Developmental Systems*, 2024.
- [16] M. Dai, J. Su, Z. Fan, C. Wang, L. Peng, D. Hu, and L.-L. Zeng, "Feature and semantic matching multi-source domain adaptation for diagnostic classification of neuropsychiatric disorders," *IEEE Transactions on Cognitive and Developmental Systems*, 2025.
- [17] Z. Qin, Y. Li, X. Song, and L. Chai, "Classification of neuropsychiatric disorders via brain-region-selected graph convolutional network," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2025.
- [18] D. Klepl, F. He, M. Wu, D. J. Blackburn, and P. Sarrigiannis, "EEG-based graph neural network classification of alzheimer's disease: An empirical evaluation of functional connectivity methods," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 2651–2660, 2022.
- [19] T. Cao, R. Lin, Y. Zheng, D. Shen, and L. Xu, "A novel approach analysing the dynamic brain functional connectivity for improved MCI detection," *IEEE Transactions on Biomedical Engineering*, vol. 71, no. 1, pp. 207–216, 2023.
- [20] K. Ma, S. Huang, and D. Zhang, "Diagnosis of mild cognitive impairment with ordinal pattern kernel," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 1030–1040, 2022.
- [21] C. Kang, D. Novak, X. Yao, J. Xie, and Y. Hu, "Classifying and scoring major depressive disorders by residual neural networks on specific frequencies and brain regions," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 2964–2973, 2023.
- [22] G. Zheng, W. Zheng, Y. Zhang, J. Wang, M. Chen, Y. Wang, T. Cai, Z. Yao, and B. Hu, "An attention-based multi-modal MRI fusion model for major depressive disorder diagnosis," *Journal of Neural Engineering*, vol. 20, no. 6, p. 066005, 2023.
- [23] M. Guo, T. Wang, Z. Zhang, N. Chen, Y. Li, Y. Wang, Z. Yao, and B. Hu, "Diagnosis of major depressive disorder using whole-brain effective connectivity networks derived from resting-state functional MRI," *Journal of Neural Engineering*, vol. 17, no. 5, p. 056038, 2020.
- [24] N. Grover, A. Chharia, R. Upadhyay, and L. Longo, "Schizo-Net: A novel schizophrenia diagnosis framework using late fusion multi-modal deep learning on electroencephalogram-based brain connectivity indices," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 464–473, 2023.
- [25] M. Huang, X. Zhang, X. Chen, Y. Mai, X. Wu, J. Zhao, and Q. Feng, "Joint-channel-connectivity-based feature selection and classification on fNIRS for stress detection in decision-making," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 1858–1869, 2022.
- [26] W. Jamal, S. Das, I.-A. Oprescu, K. Maharatna, F. Apicella, and F. Sicca, "Classification of autism spectrum disorder using supervised learning of brain connectivity measures extracted from synchrostates," *Journal of Neural Engineering*, vol. 11, no. 4, p. 046019, 2014.
- [27] H. Han, J. Jiang, L. Gu, J. Q. Gan, and H. Wang, "Brain connectivity patterns derived from aging-related alterations in dynamic brain functional networks and their potential as features for brain age classification," *Journal of Neural Engineering*, vol. 21, no. 2, p. 026015, 2024.
- [28] J. Ji and Y. Yao, "Convolutional neural network with graphical lasso to extract sparse topological features for brain disease classification," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 6, pp. 2327–2338, 2020.
- [29] E. E. Bron, M. Smits, W. J. Niessen, and S. Klein, "Feature selection based on the SVM weight vector for classification of dementia," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 5, pp. 1617–1626, 2015.
- [30] N.-F. Jie, E. A. Osuch, M.-H. Zhu, M. Wammes, X.-Y. Ma, T.-Z. Jiang, J. Sui, and V. D. Calhoun, "Discriminating bipolar disorder from major depression using whole-brain functional connectivity: a feature selection analysis with SVM-FoBa algorithm," *Journal of Signal Processing Systems*, vol. 90, pp. 259–271, 2018.
- [31] R. Holker and S. Susan, "Quantitative EEG feature selection by majorityvoting for alcohol use disorder detection," *Proc. IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*, pp. 1–4, 2021.
- [32] M. U. Ali, A. Zafar, K. D. Kallu, H. Masood, M. M. N. Mannan, M. M. Ibrahim, S. Kim, and M. A. Khan, "Correlation-filter-based channel and feature selection framework for hybrid EEG-fNIRS BCI applications," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 6, pp. 3361–3370, 2023.
- [33] S. Obertino, G. Roffo, C. Granziera, and G. Menegaz, "Infinite feature selection on shore-based biomarkers reveals connectivity modulation after stroke," *Proc. 2016 International Workshop on Pattern Recognition in Neuroimaging (PRNI)*, pp. 1–4, 2016.
- [34] Z. Pei, H. Wang, A. Bezerianos, and J. Li, "EEG-based multiclass workload identification using feature fusion and selection," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–8, 2020.
- [35] J. Vohryzek, Y. Aleman-Gomez, A. Griffa, J. Raoul, M. Cleusix, P. S. Baumann, P. Conus, K. D. Cuenod, and P. Hagmann, "Structural and functional connectomes from 27 schizophrenic patients and 27 matched healthy adults [data set]," 2020.
- [36] R. A. Poldrack, E. Congdon, W. Triplett, K. Gorgolewski, K. Karlsgodt, J. Mumford, F. Sabb, N. Freimer, E. London, T. Cannon *et al.*, "A phenome-wide examination of neural and cognitive function," *Scientific Data*, vol. 3, no. 1, pp. 1–12, 2016.
- [37] E. T. Rolls, C.-C. Huang, C.-P. Lin, J. Feng, and M. Joliot, "Automated anatomical labelling atlas 3," *NeuroImage*, vol. 206, p. 116189, 2020.
- [38] D. Kennedy, C. Haselgrove, B. Fischl, J. Breeze, J. Frazier, L. Seidman, J. Goldstein, and B. Kosofsky, "Harvard-Oxford cortical and subcortical structural atlases," *Harvard Center for Morphometric Analysis*, 2016.
- [39] M. D. Fox, D. Zhang, A. Z. Snyder, and M. E. Raichle, "The global signal and observed anticorrelated resting state brain networks," *Journal of Neurophysiology*, vol. 101, no. 6, pp. 3270–3283, 2009.
- [40] K. Murphy, R. M. Birn, D. A. Handwerker, T. B. Jones, and P. A. Bandettini, "The impact of global signal regression on resting state correlations: are anti-correlated networks introduced?" *NeuroImage*, vol. 44, no. 3, pp. 893–905, 2009.
- [41] G. Sunil, S. Gowtham, A. Bose, S. Harish, and G. Srinivasa, "Graph neural network and machine learning analysis of functional neuroimaging for understanding schizophrenia," *BMC Neuroscience*, vol. 25, no. 1, p. 2, January 2024.
- [42] S. Dreiseitl and L. Ohno-Machado, "Logistic regression and artificial neural network classification models: a methodology review," *Journal of Biomedical Informatics*, vol. 35, no. 5, pp. 352–359, 2002.
- [43] I. Kurt, M. Ture, and A. T. Kurum, "Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease," *Expert Systems with Applications*, vol. 34, no. 1, pp. 366–374, 2008.
- [44] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, 2018.
- [45] L. I. Kuncheva, "A stability index for feature selection," *Proc. of the 25th IASTED International Multi-Conference: Artificial Intelligence and Applications (AIAP'07)*, 2007.
- [46] C. Mohana and K. Perumal, "A survey on feature selection stability measures," *International Journal of Computer and Information Technology*, vol. 5, no. 1, pp. 98–103, 2016.
- [47] M. Mohammadi, H. S. Noghabi, G. A. Hodtani, and H. R. Mashhadi, "Robust and stable gene selection via maximum–minimum correntropy criterion," *Genomics*, vol. 107, no. 2-3, pp. 83–87, 2016.
- [48] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.