

UC Irvine

UC Irvine Previously Published Works

Title

A Novel Neighborhood Rough Set-Based Feature Selection Method and Its Application to Biomarker Identification of Schizophrenia

Permalink

<https://escholarship.org/uc/item/43v2f3dd>

Journal

IEEE Journal of Biomedical and Health Informatics, 27(1)

ISSN

2168-2194

Authors

Xing, Ying
Kochunov, Peter
van Erp, Theo GM
[et al.](#)

Publication Date

2023

DOI

10.1109/jbhi.2022.3212479

Peer reviewed



HHS Public Access

Author manuscript

IEEE J Biomed Health Inform. Author manuscript; available in PMC 2023 April 07.

Published in final edited form as:

IEEE J Biomed Health Inform. 2023 January ; 27(1): 215–226. doi:10.1109/JBHI.2022.3212479.

A Novel Neighborhood Rough Set-Based Feature Selection Method and Its Application to Biomarker Identification of Schizophrenia

Ying Xing,

School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China

Peter Kochunov,

Maryland Psychiatric Research Center and Department of Psychiatry, University of Maryland, School of Medicine, Baltimore, MD 21201 USA

Theo G.M. van Erp,

Department of Psychiatry and Human Behavior, School of Medicine, University of California, Irvine, CA 92617 USA, and also with the Center for the Neurobiology of Learning and Memory, University of California, Irvine, CA 92617 USA

Tianzhou Ma,

Department of Epidemiology and Biostatistics, University of Maryland, College Park, MD 20740 USA

Vince D. Calhoun,

Tri-Institutional Center for Translational Research in Neuroimaging and Data Science (TReNDS), Georgia State University, Georgia Institute of Technology, Emory University, Atlanta, GA 30030 USA

Yuhui Du

School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China

Abstract

Feature selection can disclose biomarkers of mental disorders that have unclear biological mechanisms. Although neighborhood rough set (NRS) has been applied to discover important sparse features, it has hardly ever been utilized in neuroimaging-based biomarker identification, probably due to the inadequate feature evaluation metric and incomplete information provided under a single-granularity. Here, we propose a new NRS-based feature selection method and successfully identify brain functional connectivity biomarkers of schizophrenia (SZ) using functional magnetic resonance imaging (fMRI) data. Specifically, we develop a new weighted metric based on NRS combined with information entropy to evaluate the capacity of features in distinguishing different groups. Inspired by multi-granularity information maximization theory, we further take advantage of the complementary information from different neighborhood sizes via a multi-granularity fusion to obtain the most discriminative and stable features. For validation, we

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Corresponding author: Yuhui Du, duyuhui@sxu.edu.cn.

compare our method with six popular feature selection methods using three public omics datasets as well as resting-state fMRI data of 393 SZ patients and 429 healthy controls. Results show that our method obtained higher classification accuracies on both omics data (100.0%, 88.6%, and 72.2% for three omics datasets, respectively) and fMRI data (93.9% for main dataset, and 76.3% and 83.8% for two independent datasets, respectively). Moreover, our findings reveal biologically meaningful substrates of SZ, notably involving the connectivity between the thalamus and superior temporal gyrus as well as between the postcentral gyrus and calcarine gyrus. Taken together, we propose a new NRS-based feature selection method that shows the potential of exploring effective and sparse neuroimaging-based biomarkers of mental disorders.

Keywords

Feature selection; fMRI; neighborhood rough set; information entropy; multi-granularity; schizophrenia

I. Introduction

Researchers in the neuroscience field inevitably suffer from the curse of dimensionality due to the high dimensionality of neuroimaging data. Consequently, feature selection that eliminates irrelevant and redundant features and only remains important features becomes a necessary step to mine the most useful information [1], [2]. Indeed, feature selection plays a key role in identifying meaningful biomarkers that characterize the brain impairments of mental disorders, thus providing insights on the neural substrates as well as the diagnosis and prediction of mental disorders [3].

In recent years, various feature selection methods have been utilized to identify biomarkers of mental disorders using neuroimaging data. Widely applied feature selection methods are generally categorized into unsupervised and supervised methods based on whether the diagnosis labels are utilized or not. If the category label information is unavailable, unsupervised feature selection methods, such as unsupervised correlation-based feature selection (UCFS) [4], depend on statistical measures to remove the redundant features. However, the identified features from such methods could fail to indicate the stable group difference due to the lack of guidance from category labels. In contrast, supervised feature selection methods, primarily including similarity-based, information entropy-based, and hybrid methods [5], [6], can benefit from the guidance of category labels. Similarity-based feature selection methods, such as Fisher score (FiSc) [7], [8] and ReliefF [9], [10], evaluate feature importance through the ability of features to preserve data similarity which can be inferred from label information [11]. However, most similarity-based methods cannot handle feature redundancy [5]. Information entropy-based feature selection methods, such as minimum redundancy maximum relevance method [8], [12], are proposed to maximize relevance between features and labels and minimize redundancy among features. However, information entropy-based methods can only handle discrete data, hence data discretization processes are required beforehand for continuous features, which could result in information loss [6]. Hybrid feature selection methods, such as support vector machine with recursive feature elimination (SVM-RFE) [13], [14], aim to provide an aggregated result by ranking

features based on classification accuracy. However, the features selected by hybrid methods are likely sensitive to the used classifiers [6]. In order to overcome the shortcomings of the above methods, more advanced techniques are demanded to obtain more powerful biomarkers.

Neighborhood rough set (NRS) is a mathematical tool with a profound theoretical foundation and has been used for the goal of feature selection under different granularities (neighborhood sizes) [15], [16]. NRS-based feature selection is a supervised method with strong interpretability that can effectively reduce redundant and irrelevant features and is independent of specific classifiers. Hu first proposed the neighborhood dependency degree maximization (NDDM) feature selection method to assess feature importance using consistent samples whose neighbors belong to the same class, however, the remaining inconsistent samples were completely overlooked [17]. To overcome the limitation, several feature selection methods combined with information entropy, which is suitable for measuring data uncertainty, have emerged [18]. Chen et al. and Sun et al. proposed the entropy gain-based gene selection (EGGS) [19] and decision neighborhood entropy with fisher score-based feature selection algorithm (FSDNE) [20] by maximizing joint neighborhood entropy and decision neighborhood entropy, respectively. Although EGGS and FSDNE further introduced inconsistent samples to assess the importance of features, the dominant role of consistent samples was ignored. Moreover, determining the neighborhood size in the aforementioned NRS-based feature selection methods is challenging, and such single-granularity methods are prone to yield unstable features. To integrate various complementary information derived from different granularities, multi-granularity ideas were introduced in NRS-based feature selection methods [21], [22]. Based on multiple feature ranks obtained by the NDDM method at different granularities, an integrated feature order indicating feature importance was obtained via the cross-entropy Monte Carlo algorithm, which unfortunately lacks the guidance of label information [22]. Liu integrated discriminative features obtained by the NDDM at the finest and the coarsest granularities (i.e., the smallest and largest neighborhood sizes), however, such method inevitably missed some important features at remaining granularities [23]. In short, constructing an appropriate metric to measure the importance of features and determining the neighborhood size remain challenges for NRS-based feature selection methods.

In response to the aforementioned challenges, in this paper, we propose a new multi-granularity NRS-based feature selection method and apply it to identify biomarkers of schizophrenia (SZ) using neuroimaging data. In detail, we first construct a new feature evaluation metric to balance the contribution of all samples in evaluating the discriminative ability of features and meanwhile emphasize the dominant role of consistent samples through a weighting strategy. And then, we come up with a weighted neighborhood rough set combined with entropy (WNRE) method based on the new metric to select important features under a certain granularity. More importantly, to avoid the unstable outcomes derived from a single-granularity method with a specific neighborhood size, we further extend WNRE and propose a multi-granularity WNRE-based (MGWNRE) feature selection method which integrates features under different neighborhood sizes according to their classification ability to obtain more stable and discriminative features. We first verify the effectiveness and superiority of MGWNRE method over six state-of-the-art methods

on three main omics datasets since they are suitable for algorithm evaluation thanks to more accurate ground-truth class labels. Furthermore, we employ our method to explore biomarkers of schizophrenia using functional network connectivity (FNC) features derived from functional magnetic resonance imaging (fMRI) data. Importantly, we not only utilize a main fMRI dataset to explore the sparse and important FNC features in discriminating SZ patients from healthy controls (HCs) but also demonstrate the generalization ability of the selected features on two fully independent datasets. Our results show that our method yields more discriminative features and higher classification accuracies than the six competing methods on both the main datasets and the two independent datasets. Taken together, we propose a promising NRS-based feature selection method and identify potential biomarkers of SZ.

II. Method and Materials

In this section, we first briefly review the NRS theory and information entropy. Then, we propose a single-granularity WNRE feature selection method, which effectively evaluates the importance of features by taking advantage of all available samples to improve the discriminative performance of the selected features. To combine complementary predictive powers provided by different granularities, we further extend WNRE to the MGWNRE feature selection method to ensure the stability of selected features. To verify our method, we employ three omics datasets to evaluate the feasibility of our methods and also apply our method to fMRI datasets for discovering biomarkers of SZ. In the experiments, we compare our method with six competing feature selection methods.

A. Preliminary Knowledge

We first introduce some basic concepts about NRS theory [17], [20], which are the basis of the proposed methods. Let a neighborhood decision system is denoted as $NS = \langle U, A, \{d\}, V, f, \delta \rangle$. Here, U is a nonempty finite set of all samples. A is a nonempty finite set of conditional features (i.e., features used for classification). d is a decision feature (i.e., class label in classification). V is the union set including both conditional and decision features' values across all samples, so $V = \cup_{a \in \{A \cup \{d\}\}} f(x_i, a)$, where $f(x_j, a)$ represents the value of sample x_j with respect to feature a . δ is a neighborhood size used to reflect the relations between different samples according to conditional features. B ($B \subseteq A$) is a conditional feature subset.

Equivalence class and neighborhood set are two key sample sets for any sample $x_j \in U$. Equivalence class of x_j (with respect to decision feature d) consists of samples with the same class label. And samples in the same equivalence class are indiscernible (equivalent) on d . Formally, the equivalence class of x_j can be defined by

$$[x_j]_d = \{x_i \mid x_i \in U, f(x_i, d) = f(x_j, d)\}. \quad (1)$$

Given a feature subset B , neighborhood set of x_j contains its similar and indiscernible (close) samples whose distance from x_j are less than δ . Mathematically, the neighborhood set of x_j is defined as

$$n_B^{\delta}(\mathbf{x}_i) = \{\mathbf{x}_j \mid \mathbf{x}_j \in U, \Delta_B(\mathbf{x}_i, \mathbf{x}_j) \leq \delta\}, \quad (2)$$

where $\Delta_B(\mathbf{x}_i, \mathbf{x}_j)$ is a distance function to measure the distance (dissimilarity) between samples \mathbf{x}_i and \mathbf{x}_j on a feature subset B . In this paper, we use Euclidean distance [24].

NRS-based feature selection aims to extract a conditional feature subset including discriminative and important features, so that the equivalence class can be approximately represented by neighborhood set. Thus, neighborhood dependency degree is proposed to evaluate the approximating ability by counting samples whose neighborhood set consistently belongs to its equivalence class. Such samples are called consistent samples because they accurately belong to a certain class using feature subset B , whereas the remaining inconsistent samples do not. Therefore, the significance of conditional feature subset B can be characterized by neighborhood dependency degree as

$$\gamma_B^{\delta}(\mathbf{d}) = \frac{|\{\mathbf{x}_i \mid \mathbf{x}_i \in U, n_B^{\delta}(\mathbf{x}_i) \subseteq [\mathbf{x}_i]_{\mathbf{d}}\}|}{|U|}, \quad (3)$$

where the $|\cdot|$ indicates the cardinality of a set, and the value of $\gamma_B^{\delta}(\mathbf{d})$ ranges from 0 to 1. In our method, we propose to use the neighborhood dependency degree to weigh the importance of consistent samples.

Information entropy [5] is an effective mathematical tool to determine whether a conditional feature should be retained according to the reduction of conditional entropy and is introduced to alleviate the defect that neighborhood dependency degree ignores the ability of features in distinguishing inconsistent samples. Conditional entropy [25] is widely utilized to take advantage of all samples to evaluate the discriminative ability of discrete value features. However, for continuous value features, the indispensable discretization processes may result in information loss. Neighborhood rough set theory is suitable for continuous and discrete value features but cannot leverage the inconsistent samples. Therefore, we aim to combine these two theories and propose a new metric to measure the importance of features according to all samples.

More detailed information about NRS and information entropy can be found in literature [5], [17], [20], [26].

B. Our Proposed Weighted Neighborhood Rough Set Combined With Entropy (WNRE) Feature Selection Method

A key in feature selection is to select discriminative features according to a feature evaluation metric. Inspired by prior work (i.e., NDDM [17], EGGS [19], and FSDNE [20]), we define a new metric to characterize the discriminative ability of feature subsets using all samples and propose a new feature selection method named WNRE. Considering the dominant role of consistent samples, we take neighborhood dependency degree as a weight to emphasize the capability of feature subsets in yielding consistent samples.

In our method, given neighborhood size δ and decision feature \mathbf{d} , the importance of any conditional feature subset B is defined by

$$J_{\delta}(\mathbf{B}, \mathbf{d}) = -\gamma_{\delta}^{\delta}(\mathbf{d}) \sum_{i=1}^{|\mathcal{U}|} \log \left(\frac{|\mathbf{n}_{\mathbf{B}}^{\delta}(\mathbf{x}_i) \cap [\mathbf{x}_i]_{\mathbf{d}}|}{|\mathcal{U}|} \cdot \frac{|\mathbf{n}_{\mathbf{B}}^{\delta}(\mathbf{x}_i) \cap [\mathbf{x}_i]_{\mathbf{d}}|}{|\mathbf{n}_{\mathbf{B}}^{\delta}(\mathbf{x}_i)|} \right). \quad (4)$$

The two items in brackets, namely, neighborhood accuracy and neighborhood credibility degree, are used to assess the discriminative ability of \mathbf{B} on all samples. The former item evaluates the approximation accuracy of the neighborhood set to the equivalence class using all samples. The latter reflects purity to assess the sample proportion of the same class in the neighborhood. In our method, the neighborhood dependency degree is used as a weight value to emphasize the discriminative ability of feature subsets on consistent samples. In sum, we propose a feature evaluation metric to comprehensively evaluate the capacity of feature subsets in distinguishing the whole sample set while highlighting the role of consistent samples.

A larger $J_{\delta}(\mathbf{B}, \mathbf{d})$ means that the related features in \mathbf{B} are more discriminative. Hence, we define a significance measurement of a given conditional feature $\mathbf{a}_j \in \mathbf{A} - \mathbf{B}$ as

$$\text{Sig}(\mathbf{a}_j, \mathbf{B}, \mathbf{d}) = J_{\delta}(\mathbf{B} \cup \{\mathbf{a}_j\}, \mathbf{d}) - J_{\delta}(\mathbf{B}, \mathbf{d}). \quad (5)$$

Here, $\text{Sig}(\mathbf{a}_j, \mathbf{B}, \mathbf{d})$ represents the increased contribution of feature subset \mathbf{B} after adding a feature \mathbf{a}_j . If $\text{Sig}(\mathbf{a}_j, \mathbf{B}, \mathbf{d}) > 0$, it means that feature \mathbf{a}_j is essential, otherwise, it is deleted because it is a redundant feature (when $\text{Sig}(\mathbf{a}_j, \mathbf{B}, \mathbf{d}) = 0$) or irrelevant feature (when $\text{Sig}(\mathbf{a}_j, \mathbf{B}, \mathbf{d}) < 0$). The larger the value of $\text{Sig}(\mathbf{a}_j, \mathbf{B}, \mathbf{d})$, the more important feature \mathbf{a}_j is. In our method, we perform a heuristic algorithm to search for the discriminative feature subset. In the process, we start from an empty feature set and add features iteratively through maximizing $\text{Sig}(\mathbf{a}_j, \mathbf{B}, \mathbf{d})$. This operation terminates when $\text{Sig}(\mathbf{a}_j, \mathbf{B}, \mathbf{d})$ stops increasing.

In addition, given that the feature selection process is time-consuming on high-dimensional data, we first employ the FiSc method [7] to preselect a candidate feature subset to speed up the subsequent feature selection process. Herein, we select the top l candidate discriminate features as input to the subsequent WNRE feature selection process. The WNRE algorithm is described in Algorithm 1.

As shown in Algorithm 1, given the original conditional feature set \mathbf{A} , decision feature \mathbf{d} , and neighborhood size δ , we first initialize the expected discriminative feature subset \mathbf{B}^* as an empty set. Then, we select the top l features to constitute the candidate feature subset \mathbf{A}' based on the FiSc method. Next, the most important feature according to the evaluation metric (4–5) is iteratively added into the subset \mathbf{B}^* one by one. After incorporating a feature, \mathbf{B}^* is judged to see if it meets the terminal condition. If the condition is not satisfied, the algorithm continues to iterate and amalgamate the important features, otherwise outputs the selected feature subset \mathbf{B}^* .

Algorithm 1:

WNRE.

 Input: A neighborhood decision system $NS = \langle U, A, \{d\}, V, f, \delta \rangle$.

 Output: Excepted feature subset B^* .

```

1:   Initialize  $B^* = \emptyset$ .
2:   Select top- $l$  features from  $A$  to constitute the candidate feature subset  $A'$  based on FiSc method.
3:   do {
4:     for  $\forall a_j \in A'$ 
5:       Compute  $Sig(a_j, B^*, d)$  according to formula (4–5).
6:     end for
7:     Select the feature  $a_k$  satisfying  $Sig(a_k, B^*, d) = \max(Sig(a_j, B^*, d))$ .
8:      $B^* = B^* \cup \{a_k\}, A' = A' - \{a_k\}$ .
9:   } until  $Sig(a_k, B^*, d) < \epsilon^{-10}$ 

```

Since WNRE faces difficulty in determining neighborhood size, we further propose a multi-granularity fusion strategy to overcome its shortcoming.

C. Our Proposed Multi-Granularity WNRE-Based (MGWNRE) Feature Selection Method

Based on our proposed WNRE method, an extended feature selection method, abbreviated by MGWNRE, is proposed to merge the features selected by multiple WNRE processes under different neighborhood sizes and yield complementary and stable important features. The algorithm is outlined in Algorithm 2.

As shown in Algorithm 2, given initial conditional feature set A , decision feature d , and a set of neighborhood sizes Θ , we perform multiple WNRE feature selection processes and merge the feature subsets that result in a satisfying performance. Specifically, for arbitrary δ , we can obtain a feature subset B^* after executing WNRE algorithm. Then, we apply a classifier to evaluate the significance of B^* and obtain associated comprehensive classification performance (CCP, i.e., mean of classification accuracy, sensitivity, and specificity). After performing this operation for each δ , we merge feature subsets whose CCP is better than the average level and output the final discriminative features.

D. Performance Evaluation of MGWNRE and Comparisons With Other Methods

In this section, we perform two experiments to verify the effectiveness of our methods and identify stable biomarkers of SZ. In the first experiment, we evaluate the effectiveness of our methods on three public omics datasets with classification labels to assess if the selected features are discriminative in classifying different classes. In the second experiment, our methods are applied to fMRI data to identify discriminative features (that indicate biomarkers) that can separate SZ patients from HCs. The generalization ability and stability of the identified biomarkers are further evaluated using fully independent fMRI datasets. Furthermore, we implement six competing feature selection methods for a comprehensive comparison in the two experiments.

Algorithm 2:

MGWNRE.

Input: A neighborhood decision system $NS = \langle U, A, \{d\}, V, f, \delta \rangle$, a set of neighborhood sizes $\Theta = \{\delta_1, \delta_2, \dots, \delta_q\}$, $|\Theta| = q$.

Output: Excepted feature subset B^{**} .

1. Initialize $B^{**} = \emptyset$.
- 2: **for** $\forall \delta_i \in \Theta$
- 3: Execute WNRE and obtain a feature subset B_i^* with current δ_i .
- 4: Compute CCP_i related to B_i^* .
- 5: **end for**
- 6: Compute average $\overline{CCP} = \frac{1}{q} \sum_i^q CCP_i$.
- 7: **for** $\forall CCP_i$
- 8: **if** $CCP_i \geq \overline{CCP}$
- 9: $B^{**} = B^{**} \cup B_i^*$.
- 10: **end if**
- 11: **end for**

1) Experimental Pipelines for Our Proposed Methods and Competing

Methods: Fig. 1 shows the experimental pipelines for our proposed methods (i.e., WNRE and MGWNRE) and the competing methods (i.e., NDDM, EGGS, FSDNE, UCFS, ReliefF, and SVM-RFE). Fig. 1(a) displays the experimental pipeline of each method using the main datasets, including three omics datasets and an fMRI dataset, which aims to validate the distinguishing ability of the features selected by each method. Fig. 1(b) displays the experimental pipeline of each method on two fMRI independent datasets, which focuses on validating the generalization ability of the features selected by each method from the main fMRI dataset.

As shown in Fig. 1(a), using each main dataset, each method is carried out under an outer 10-fold cross-validation with an embedded inner 5-fold cross-validation pipeline to ensure that the selected features are reliable and unbiased. Considering the existence of intrinsic parameters in each method, the inner cross-validation aims to search the optimal parameters and select discriminative features, and the outer cross-validation aims to evaluate the distinguishing ability of the selected features on unseen data. Specifically, we perform the following steps for each method on each main dataset. For the outer cross-validation, we divide data into ten equal folds randomly. Nine of ten folds are used as the outer training data to search for discriminative features and train an outer classifier. The remaining fold is used as the outer testing data to evaluate the features according to the classification performance of the trained classifier. It is worth pointing out that we first employ FiSc method on the outer training data to preselect features whose fisher scores are greater than the average score as the candidate features to improve computation efficiency. For the inner cross-validation, outer training data with candidate features are divided into five equal folds. Given specific parameters for each method, four of five folds are used as the inner

training data to select features and train the inner classifier with the selected features, and the remaining inner one fold is used to evaluate the selected features via the above-mentioned comprehensive classification performance of the trained inner classifiers.

To be fair, we search the optimal parameters according to the classification results of inner cross-validation to select the most discriminative features for each method. In detail, according to the comprehensive classification performance of the trained inner classifiers in the inner cross-validation, we use optimal neighborhood size for WNRE, NDDM, EGGs, and FSDNE, and optimal numbers of selected features for UCFS, ReliefF, and SVM-RFE. As such, for each method, we can obtain ten discriminative feature subsets and associated classification performance (i.e., accuracy, sensitivity, and specificity). Finally, the classification performances of our proposed methods and the competing methods are compared.

To verify the generalization ability of the features selected by each method from the main fMRI dataset, the top-10 features with the highest occurring frequency in the ten discriminative feature subsets are utilized on independent fMRI datasets for a strict classification test, respectively, as shown in Fig. 1(b). In detail, we perform another classification experiment under 10-fold cross-validation using independent datasets based on the top-10 discriminative features selected by each method. In each run, nine folds are used to train a classifier, and the remaining one fold is used to test the performance based on the classifier. The classification performances from different methods are compared.

Here, we employ SVM with a linear kernel [14], K nearest neighbor with K of 3 (3NN) [19], and AdaBoost [27] as the outer classifiers for separate tests to verify the classification performance and take SVM as the inner classifier.

2) Performance Evaluation Experiment on Omics Datasets:

Omics data that is helpful for tumor and cancer diagnosis have been widely used in evaluating the performance of feature selection methods [19], [20]. In the first experiment, we employ our two methods on three public omics datasets to examine if the features selected by our methods have powerful abilities in classifying different classes. Small-round-blue-cell tumor (SR-BCT) dataset involves 2308 features and 83 samples, including four childhood tumor subtypes [20]. Colon dataset involves 2000 features and 62 samples, including patients with colon cancer and healthy samples [20]. Breast invasive carcinoma (BRCA) dataset involves 6780 features and 305 samples, including four cancer subtypes [28].

As shown in Fig. 1(a), using the three omics datasets, respectively, we perform our methods and competing methods under the unbiased 10-fold cross-validation pipeline to compare the average accuracy, sensitivity, specificity, and the number of selected features of each method.

3) Performance Evaluation Experiment on fMRI Datasets:

FNC features derived from fMRI data potentially can reveal the aberrant brain activity of mental disorders such as SZ, which is a common and serious illness [29]. Identifying

biomarkers for SZ is expected to help understand the underlying mechanism of the disorder and assist in diagnosis.

In the second experiment, FNC features derived from resting-state fMRI data of 100 SZ patients and 127 age-matched HCs collected from the Function Biomedical Informatics Research Network (FBIRN)¹ are taken as the main fMRI dataset which is used to validate that our methods can select the significant features in distinguishing HC and SZ groups. Then, 163 SZ patients and 149 age-matched HCs from Bipolar-Schizophrenia Network on Intermediate Phenotypes (BSNIP)² [30] and 130 SZ patients and 153 age-matched HCs from Maryland Psychiatric Research Center (MPRC)³ are taken as two independent datasets which are separately utilized to evaluate the generalization ability of the features selected from the main fMRI dataset. Here, 1378 FNC features revealing interactions between brain networks are derived from the fMRI data as the initial features for each subject based on the previously developed NeuroMark method [31], [32] and toolbox (<http://www.yuhuidu.com/>).

Using the FNC features, we first carry out our methods on the FBIRN dataset under the 10-fold cross-validation pipeline (Fig. 1(a)) to examine the ability of selected FNC features in distinguishing HC and SZ groups. In the pipeline, the outer cross-validation is leveraged to train and evaluate the outer classifier based on the discriminative FNC features selected from the associated inner cross-validation. In the experiment, we also evaluate the classification performance of the six competing methods using the same pipeline.

Based on the respective top-10 discriminative FNC features selected by each method from FBIRN dataset, we perform another classification test on independent datasets (BSNIP and MPRC datasets) under a 10-fold cross-validation experimental pipeline (Fig. 1(b)). In this pipeline, a classifier is trained using nine folds data, and the remaining one fold data is then used to test the classification performance based on the classifier. Finally, we compare the classification performance of our methods with that of the competing methods on the two datasets. Considering that we only use a small number (i.e., 10) of FNC features selected by each method from FBIRN dataset to perform the classification test on independent datasets, we would see which feature selection method is more stable according to the classification performance.

The selected FNC features which work well in the HC-SZ classification may represent meaningful biomarkers for SZ. Therefore, using each fMRI dataset (both main and independent datasets), we summarize and analyze the top-10 high-frequency FNC features selected by our MGWNRE method. First, we perform a two-sample t-test to test if there are significant differences with respect to each selected FNC feature between the HC and SZ groups. Then, we compute the average connectivity strength of each selected FNC feature for each group (HC or SZ) to show the differences between the groups. We aim to investigate if the group differences are consistent across different datasets in terms of the ten FNC features.

¹The reference number of the IRB approval for FBIRN data is HS No.2009-7128.

²The original source of BSNIP data can be found at <https://nda.nih.gov/>.

³The reference number of the IRB approval for MPRC data is HP-00045716.

4) Competing Methods and Comprehensive Comparison of MGWNRE With Other Methods on All Datasets:

For a comparison with our methods, we choose six popular feature selection methods, including three NRS-based methods (i.e., NDDM, EGGS, and FSDNE) and three non-NRS-based methods (i.e., UCFS, ReliefF, and SVM-RFE). Since there are multiple datasets and competing methods in our work, it is difficult to judge which method works best overall. Therefore, we compute an overall ranking on all used datasets for these methods. In detail, we first rank all methods according to the comprehensive classification performance for each dataset. Then, we calculate the average ranking for each method on the six datasets and obtain an overall ranking of these methods. Moreover, we compute the standard deviation of the comprehensive classification performance for each method on the six datasets to reflect the stability of classification performance.

III. Results

A. Results on Omics Datasets in the First Experiment

Fig. 2 and Table I display the experimental results from the different feature selection methods on the omics datasets, including SRBCT, Colon, and BRCA. It should be noted that the number of candidate features preselected by FiSc method was 654, 629, and 2077 for SRBCT, Colon, and BRCA datasets, respectively. Using the SRBCT dataset, the classification results based on different classifiers (i.e., SVM, 3NN, and AdaBoost classifiers) are shown in Fig. 2(a)–(c), and it can be seen that based on each classifier, the proposed MGWNRE consistently obtained superior classification results, and WNRE also reached promising results compared with the competing methods. As outlined in Table I, among the eight methods, MGWNRE achieved the best performance with an average accuracy of 100%, sensitivity of 100%, and specificity of 100% based on both SVM and 3NN classifiers. Based on AdaBoost classifier, MGWNRE also obtained the highest average accuracy (i.e., 95.4%) and highest average sensitivity (i.e., 97.5%). In addition, from the table, we can see that WNRE also achieved good classification results which were better than that of the most competing methods based on the three classifiers. In short, MGWNRE selected a small number of features (i.e., 53 features, only 2.3% of the entire features) and achieved satisfactory classification results.

The experimental results using the Colon dataset based on the three classifiers are presented in Fig. 2(d)–(f) and Table I. From the figure, we can see that the classification results of MGWNRE were remarkably better than that of other competing methods based on SVM and 3NN classifiers. From the table, it can be seen specifically that among all methods, MGWNRE attained the highest average accuracy of 88.3% based on SVM classifier and 88.6% based on 3NN classifier, and its classification performance based on AdaBoost classifier was also better than that of most competing methods. In addition, WNRE obtained acceptable classification performances compared with other methods based on different classifiers. In summary, MGWNRE achieved an excellent distinguishing ability using relatively few features (i.e., 70 features, 3.5% of the entire features).

Fig. 2(g)–(i) and Table I display the results on the BRCA dataset using the three classifiers. Compared with the competing methods, MGWNRE selected a small number of important features (i.e., 72 features, only 1% of the entire features) and obtained the highest average accuracy of 72.2%, sensitivity of 90.0%, and specificity of 68.4% when using SVM classifier, and satisfactory classification results when using other two classifiers. In addition, the proposed WNRE method also achieved promising results relative to the competing methods.

B. Results on fMRI Datasets in the Second Experiment

Using fMRI data of SZ patients and HCs, the classification results of the main dataset (i.e., FBIRN) based on the three classifiers are presented in Fig. 3(a)–(c) and Table II. It should be noted that the number of candidate features preselected by FiSc method was 430 for FBIRN dataset. The figure provides an overview of the classification performance for each feature selection method, and we can see that both MGWNRE and WNRE got outstanding results. From the table, it is seen that compared with the competing methods, MGWNRE identified the fewest features (i.e., 21 features, only 1.5% of the entire features) and reached the best average accuracy and specificity based on each classifier. Based on 3NN classifier, MGWNRE obtained the highest average accuracy of 93.9%, sensitivity of 93.8%, and specificity of 94.0%. In addition, WNRE also obtained outstanding classification results compared with the competing methods based on each classifier.

Using the top-10 high-frequency FNC features selected by each method on the FBIRN dataset, the experimental results of the two independent datasets (i.e., BSNIP and MPRC datasets) are shown in Fig. 3(d)–(i) and Table II. As displayed in Fig. 3(d)–(f) and Table II, compared with the competing methods, MGWNRE obtained the best average accuracy and sensitivity based on 3NN and AdaBoost classifiers, and outperformed five of six competing methods based on SVM classifier which achieved an average accuracy of 76.3% on BSNIP dataset. Besides, WNRE also got an acceptable classification performance on BSNIP dataset. Regarding the results on MPRC dataset (Fig. 3(g)–(i) and Table II), MGWNRE got excellent classification results with an average accuracy of 83.8%, sensitivity of 84.3%, and specificity of 83.1% based on SVM classifier, which outperformed all six competing methods. Based on 3NN and AdaBoost classifiers, MGWNRE also obtained promising classification results. In addition, WNRE was superior to most competing methods based on each classifier on MPRC dataset. To summarize, using only the top-10 FNC features selected by our proposed methods on the main fMRI dataset, the SZ patients and HCs in the fully independent datasets can be effectively distinguished.

To show the group difference of the top-10 high-frequency FNC features selected by MGWNRE from the main fMRI dataset, Table III includes the P-values and T-values of each FNC feature obtained by performing a two-sample t-test between the HC and SZ groups in the FBIRN, BSNIP, and MPRC datasets. For FBIRN and BSNIP datasets, our results suggest that there were significant differences (P -value < 0.05) between the two groups for all the ten FNC features. For MPRC dataset, nine of ten FNC features showed significant differences between groups.

More interestingly, Fig. 4 and Table III both support that the changes in the connectivity strengths of FNCs in the SZ group relative to the HC group were very consistent across the three datasets. Among the ten FNC features, the SZ group showed higher connectivity strengths than the HC group in four FNCs, including the connectivity between the thalamus and superior temporal gyrus, middle temporal gyrus and insula, middle temporal gyrus and anterior cingulate cortex, and right middle occipital gyrus and hippocampus. In contrast, the connectivity strengths of the remaining six FNCs were lower in the SZ group relative to the HC group. In sum, our method found the most significant FNCs that showed consistent changing trends in SZ across multiple datasets.

C. Results of the Comprehensive Comparison of MGWNRE With Other Methods on All Datasets

To compare the comprehensive classification performance of the feature selection methods based on the three classifiers, Fig. 5 shows the overall ranking of each method on the multiple datasets, including SRBCT, Colon, BRCA, FBIRN, BSNIP, and MPRC. The proposed MGWNRE not only ranked in the top but also had lower standard deviations than the other six competing methods based on each classifier. WNRE also obtained the suboptimal ranking based on the three classifiers. Taken together, the selected features by our methods are stable and discriminative in terms of classification performance.

IV. Discussions

In the neuroscience field, the identification of meaningful biomarkers of mental disorders from neuroimaging data is an important topic for the mission of precision diagnostics. Feature selection technologies help select meaningful features to construct an efficient and interpretable model with strong generalization capability [5]. Recently, feature selection has shown the potential in characterizing the aberrant brain functional connectivity of schizophrenia using fMRI data, however, the underlying neurological mechanism of the disorder remains unclear [33]. Neighborhood rough set-based feature selection methods have been shown to appropriately identify sparse and more discriminative features, but they are rarely used in biomarker exploration for brain disorders.

Aiming at constructing a more effective feature evaluation metric and overcoming the shortcoming of the determination of neighborhood size in traditional NRS-based feature selection methods, we propose a new feature selection method, i.e., MGWNRE, to identify the sparse and important features that can distinguish different classes well. Herein, based on NRS theory and information entropy, we develop a feature evaluation metric that not only assesses the importance of features using both consistent and inconsistent samples but also highlights the dominant contribution of consistent samples. Based on the novel metric, we propose WNRE and extend it to a multi-granularity level (i.e., MGWNRE) to minimize the need in determining a prior neighborhood size, thus further ensuring the distinguishing capability as well as stability of the selected features. Moreover, we manifest the feasibility and effectiveness of the discriminative features selected by MGWNRE through comprehensive evaluations on omics and fMRI datasets. It is known that all supervised feature selection methods involved in the paper are for classification tasks in

which the selected features are used for distinguishing different groups. However, sometimes a prediction task with continuous output also can be converted to a classification problem with discrete labels before using these methods [34].

Comparing classification performance with the six state-of-the-art feature selection methods on the three omics datasets, our results demonstrated that MGWNRE can identify better discriminative features to separate different classes. In detail, the improved classification results using the features selected by MGWNRE indicated that the proposed method can effectively identify a small number of discriminative features to classify different groups, and present the superiority of MGWNRE over the competing methods. In addition, we found that the classification performances of compared NRS-based methods were generally inferior to non-NRS-based methods, although NRS-based methods can select fewer features. Impressively, our MGWNRE method not only outperformed the NRS-based methods, which indicated that the proposed feature importance metric more appropriately characterized the discriminative ability of features, but also outperformed non-NRS-based methods. In short, this experiment verified the practicability of MGWNRE in identifying the discriminative features.

Findings from the fMRI datasets further support that the features detected by MGWNRE can distinguish SZ patients from HCs well. Specifically, MGWNRE method achieved a promising average accuracy of 93.9% based on 3NN classifier on the main fMRI dataset, whereas all competing methods had accuracies below 90%. Existing systematic reviews reported that using a sample size similar to our work, the classification accuracies of most previous studies classifying SZ patients and HCs were less than 90% [35], [36]. In addition, only using the top-10 high-frequency FNC features selected from the main fMRI dataset, MGWNRE also obtained satisfactory classification accuracies of 76.3% and 83.8% based on SVM classifier for the two independent cohorts, respectively. While the classification accuracies of the competing methods ranged from 65.1% to 81.7% based on SVM classifier on the independent cohorts. In previous studies differentiating SZ patients from HCs, the accuracies on completely independent datasets were around 70% [37]. The high classification accuracy on independent datasets in our work demonstrated the robustness and stability of the features selected by our MGWNRE. In short, our proposed feature importance metric in WNRE can efficiently represent the discriminative ability of features. In the meantime, MGWNRE further improves the single-granularity WNRE method by providing complementary information at different granularity levels.

Furthermore, our work highlights a small number of discriminative functional connectivity that contributes to distinguish HCs and SZ patients. Based on our MGWNRE, we analyzed top-10 important FNC features, primarily including the thalamus, cerebellum, temporal (e.g., middle and superior temporal gyrus), parietal (e.g., left and right postcentral gyrus, and superior parietal lobule), occipital gyrus (e.g., right middle and inferior occipital gyrus), and calcarine gyrus regions, which showed abnormalities in SZ patients compared with HCs. Specifically, SZ patients showed significantly altered thalamus-related functional connectivity, including lower thalamic functional connectivity with the cerebellum and caudate, and significantly higher thalamic functional connectivity with the superior temporal gyrus. Indeed, it is well known that the thalamus plays a key role in information processing

in the brain and shows significant abnormalities in SZ. Previous studies have reported the analogous aberrant functional connectivity in SZ patients. Using resting-state seed-based functional connectivity, Ferri et al. found that SZ patients had smaller thalamic connectivity with the cerebellum and greater thalamic connectivity with multiple sensory-motor regions, including the superior temporal gyrus [29]. Also using the seed-based functional connectivity, Huang et al. studied the functional dysconnectivity pattern of salience network in first-episode SZ patients which had shown a notable reduced functional connectivity between the thalamus and caudate head compared with HCs [38]. In addition, we found that SZ patients presented lower connectivity between the superior parietal lobule and inferior occipital gyrus compared with HCs, which is consistent with a previous study that also reported lower connectivity between the superior parietal and occipital cortex [39]. More importantly, new findings have been observed in our paper: (1) significantly lower connectivity between postcentral gyrus and calcarine gyrus in SZ patients (compared with HCs); (2) higher connectivity between middle temporal gyrus and insula, between middle temporal gyrus and anterior cingulate cortex, and between right middle occipital gyrus and hippocampus in SZ patients (compared with HCs). In terms of the first new finding, we have good reason to speculate that the interaction between sensory function and visual processing might be disrupted in SZ because of the reduced connectivity between the postcentral gyrus (the primary sensory receiving area of touch) and the calcarine gyrus (closely related to the visual processing) [40], [41]. In recent research, Ramsay et al. also highlighted the relationship between the low-level visual sensory discrimination impairments and the cognitive disruptions of mental disorders including SZ [42]. For the second new finding, as is well known that the middle temporal gyrus is involved in language and semantic memory processing, and the insula as well as anterior cingulate cortex are both closely linked to cognitive and emotional processing [43], [44], [45], our findings further indicated that the information communication between language and emotional processing might be extremely active in SZ. A study indirectly supporting our findings reported that SZ patients have shown a larger negative response to negative words than HCs, which indicated the anomalies in the interactions between semantic and emotion processing of SZ [46]. Additionally, studies have shown that the occipital cortex is associated with the maintenance of visuospatial information, and the hippocampus plays a crucial role in memory and learning [39], [47]. The higher connectivity between the occipital gyrus and hippocampus that we found in this paper would support that vision abnormality (e.g., hallucinations) in SZ may be associated with memory impairments. Taken together, we found important brain functional impairments in SZ by using our proposed method, as the few related features selected by our method can distinguish the SZ and HC groups well using different datasets.

V. Conclusion

In this paper, a new NRS-based feature selection method, named MGWNRE, is proposed to identify the sparse and important features that can distinguish different classes well and is applied to explore biomarkers revealing brain functional impairments in SZ. A new feature evaluation metric is developed in MGWNRE, which not only makes full use of all samples when assessing feature importance but also emphasizes the contribution of consistent samples in the assessment. The use of sufficient data in assessing the features'

importance ensures that more complete and reliable important features can be selected in our method. More importantly, the multi-granularity idea in MGWNRE incorporates complementarily important features selected at different granularities, further improving the stability and generalization ability of the selected features. We verify the effectiveness and superiority of the proposed MGWNRE method over six competing methods on three well-known omics datasets and three fMRI datasets. Furthermore, the reliable biomarkers of SZ are identified based on MGWNRE, which reveals brain functional impairments of SZ. However, it should be noted that more computation time is an inherent shortcoming of all NRS-based feature selection methods, although the computation efficiency of our proposed methods is comparable to previous NRS-based methods. In addition, single-modal data could not fully characterize all abnormalities of SZ, and we expect to further validate the proposed method by integrating multimodal data.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China through YHD under Grants 62076157 and 61703253, in part by the Fund Program for the Scientific Activities of Selected Returned Overseas Professionals in Shanxi Province through YHD, in part by the 1331 Engineering Project of Shanxi Province of China, in part by the National Institutes of Health under Grants R01MH118695 and NIH R01MH123610, and in part by NSF under Grant 2112455 through VDC.

References

- [1]. Rondina JM et al. , “SCoRS—A method based on stability for feature selection and mapping in neuroimaging,” *IEEE Trans. Med. Imag*, vol. 33, no. 1, pp. 85–98, Jan. 2014.
- [2]. Xia CH et al. , “Linked dimensions of psychopathology and connectivity in functional brain networks,” *Nature Commun*, vol. 9, no. 1, pp. 1–14, 2018. [PubMed: 29317637]
- [3]. Rueda A, González FA, and Romero E, “Extracting salient brain patterns for imaging-based classification of neurodegenerative diseases,” *IEEE Trans. Med. Imag*, vol. 33, no. 6, pp. 1262–1274, Jun. 2014.
- [4]. Zheng Q, Delingette H, Fung K, Petersen SE, and Ayache N, “Pathological cluster identification by unsupervised analysis in 3,822 U.K. biobank cardiac MRIs,” *Front. Cardiovasc. Med*, vol. 7, 2020, Art. no. 539788.
- [5]. Li J et al. , “Feature selection: A data perspective,” *ACM Comput. Surv*, vol. 50, no. 6, pp. 1–45, 2017.
- [6]. Cai J and Luo J, Wang S, and Yang S, “Feature selection in machine learning: A new perspective,” *Neurocomputing*, vol. 300, pp. 70–79, 2018.
- [7]. Gu Q, Li Z, and Han J, “Generalized fisher score for feature selection,” in *Proc. Conf. UAI*, 2012, pp. 266–273.
- [8]. Zhang T, Zhao Z, Zhang C, Zhang J, Jin Z, and Li L, “Classification of early and late mild cognitive impairment using functional brain network of resting-state fMRI,” *Front. Psychiatry*, vol. 10, 2019, Art. no. 572.
- [9]. Kira K and Rendell LA, “A practical approach to feature selection,” in *Proc. Int. Conf. Mach. Learn.*, 1992, pp. 249–256.
- [10]. Huang Z, Yang C, Zhou X, and Huang T, “A hybrid feature selection method based on binary state transition algorithm and ReliefF,” *IEEE J. Biomed. Health Inform*, vol. 23, no. 5, pp. 1888–1898, Sep. 2019. [PubMed: 30281502]
- [11]. Zhao Z, Wang L, Liu H, and Ye J, “On similarity preserving feature selection,” *IEEE Trans. Knowl. Data Eng*, vol. 25, no. 3, pp. 619–632, Mar. 2013.
- [12]. Peng H, Long F, and Ding C, “Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy,” *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005. [PubMed: 16119262]

- [13]. Bron EE, Smits M, Niessen WJ, and Klein S, “Feature selection based on the SVM weight vector for classification of dementia,” *IEEE J. Biomed. Health. Inform.*, vol. 19, no. 5, pp. 1617–1626, Sep. 2015. [PubMed: 25974958]
- [14]. Du Y, Hao H, Wang S, Pearlson GD, and Calhoun VD, “Identifying commonality and specificity across psychosis sub-groups via classification based on features from dynamic connectivity analysis,” *NeuroImage: Clin.*, vol. 27, 2020, Art. no. 102284.
- [15]. Hu Q, Pedrycz W, Yu D, and Lang J, “Selecting discrete and continuous features based on neighborhood decision error minimization,” *Trans. Syst. Man Cybern. Part B-Cybern.*, vol. 40, no. 1, pp. 137–150, 2010.
- [16]. Yong L, Wenliang H, Yunliang J, and Zhiyong Z, “Quick attribute reduct algorithm for neighborhood rough set model,” *Inf. Sci.*, vol. 271, pp. 65–81, 2014.
- [17]. Hu Q, Yu D, Liu J, and Wu C, “Neighborhood rough set based heterogeneous feature subset selection,” *Inf. Sci.*, vol. 178, no. 18, pp. 3577–3594, 2008.
- [18]. Sun L, Zhang X, Xu J, and Zhang S, “An attribute reduction method using neighborhood entropy measures in neighborhood rough sets,” *Entropy*, vol. 21, no. 2, 2019, Art. no. 155.
- [19]. Chen Y, Zhang Z, Zheng J, Ma Y, and Xue Y, “Gene selection for tumor classification using neighborhood rough sets and entropy measures,” *J. Biomed. Inform.*, vol. 67, pp. 59–68, 2017. [PubMed: 28215562]
- [20]. Sun L, Zhang X, Qian Y, Xu J, and Zhang S, “Feature selection using neighborhood entropy-based uncertainty measures for gene expression data classification,” *Inf. Sci.*, vol. 502, pp. 18–41, 2019.
- [21]. Li Y, Cai M, Zhou J, and Li Q, “Accelerated multi-granularity reduction based on neighborhood rough sets,” *Appl. Intell.*, pp. 1–16, 2022.
- [22]. Lin Y, Li J, Lin P, Lin G, and Chen J, “Feature selection via neighborhood multi-granulation fusion,” *Knowl.-Based Syst.*, vol. 67, pp. 162–168, 2014.
- [23]. Liu K, Yang X, Fujita H, Liu D, Yang X, and Qian Y, “An efficient selector for multi-granularity attribute reduction,” *Inf. Sci.*, vol. 505, pp. 457–472, 2019.
- [24]. Wilson DR and Martinez TR, “Improved heterogeneous distance functions,” *J. Artif. Intell. Res.*, vol. 6, pp. 1–34, 1997.
- [25]. Dai J, Wang W, Tian H, and Liu L, “Attribute selection based on a new conditional entropy for incomplete decision systems,” *Knowl.-Based Syst.*, vol. 39, pp. 207–213, 2013.
- [26]. Shannon CE, “A mathematical theory of communication,” *Bell system Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.
- [27]. Freund Y and Schapire RE, “A decision-theoretic generalization of on-line learning and an application to boosting,” *J. Comput. System Sci.*, vol. 55, no. 1, pp. 119–139, 1997.
- [28]. Goldman M et al. , “The UCSC cancer genomics browser: Update 2015,” *Nucleic Acids Res.*, vol. 43, no. D1, pp. D812–D817, 2015. [PubMed: 25392408]
- [29]. Ferri J et al. , “Resting-state thalamic dysconnectivity in schizophrenia and relationships with symptoms,” *Psychol. Med.*, vol. 48, no. 15, pp. 2492–2499, 2018. [PubMed: 29444726]
- [30]. Tamminga CA, Pearlson G, Keshavan M, Sweeney J, Clementz B, and Thaker G, “Bipolar and schizophrenia network for intermediate phenotypes: Outcomes across the psychosis continuum,” *Schizophrenia Bull.*, vol. 40, no. Suppl_2, pp. S131–S137, 2014.
- [31]. Duand YFan Y, “Group information guided ICA for fMRI data analysis,” *Neuroimage*, vol. 69, pp. 157–197, 2013.. [PubMed: 23194820]
- [32]. Du Y et al. , “NeuroMark: An automated and adaptive ICA based pipeline to identify reproducible fMRI markers of brain disorders,” *NeuroImage: Clin.*, vol. 28, 2020, Art. no. 102375.
- [33]. Jafri MJ, Pearlson GD, Stevens M, and Calhoun VD, “A method for functional network connectivity among spatially independent resting-state components in schizophrenia,” *Neuroimage*, vol. 39, no. 4, pp. 1666–1681, 2008. [PubMed: 18082428]
- [34]. El-Manzalawy Y, Hsieh T-Y, Shivakumar M, Kim D, and Honavar V, “Min-redundancy and max-relevance multi-view feature selection for predicting ovarian cancer survival using multi-omics data,” *BMC Med. Genomic.*, vol. 11, no. 3, pp. 19–31, 2018.

- [35]. Steardo L Jr. et al. , “Application of support vector machine on fMRI data as biomarkers in schizophrenia diagnosis: A systematic review,” *Front. Psychiatry*, vol. 11, 2020, Art. no. 588.
- [36]. Rashid B and Calhoun V, “Towards a brain-based predictome of mental illness,” *Hum. Brain Mapping*, vol. 41, no. 12, pp. 3468–3535, 2020.
- [37]. Cai XL et al. , “Generalizability of machine learning for classification of schizophrenia based on resting-state functional MRI data,” *Hum. Brain Mapping*, vol. 41, no. 1, pp. 172–184, 2020.
- [38]. Huang H et al. , “Aberrant resting-state functional connectivity of salience network in first-episode schizophrenia,” *Brain Imag. Behav*, vol. 14, no. 5, pp. 1350–1360, 2020.
- [39]. Henseler I, Falkai P, and Gruber O, “Disturbed functional connectivity within brain networks subserving domain-specific subcomponents of working memory in schizophrenia: Relation to performance and clinical symptoms,” *J. Psychiatr. Res*, vol. 44, no. 6, pp. 364–372, 2010. [PubMed: 19837416]
- [40]. Ploner M, Schmitz F, Freund H-J, and Schnitzler A, “Differential organization of touch and pain in human primary somatosensory cortex,” *J. Neurophysiol*, vol. 83, no. 3, pp. 1770–1776, 2000. [PubMed: 10712498]
- [41]. Li RR et al. , “Altered functional connectivity strength and its correlations with cognitive function in subjects with ultra-high risk for psychosis at rest,” *CNS Neurosci. Therapeutics*, vol. 24, no. 12, pp. 1140–1148, 2018.
- [42]. Ramsay IS, Schallmo M-P, Biagianti B, Fisher M, Vinogradov S, and Sponheim SR, “Deficits in auditory and visual sensory discrimination reflect a genetic liability for psychosis and predict disruptions in global cognitive functioning,” *Front. Psychiatry*, vol. 11, 2020, Art. no. 638.
- [43]. Bush G, Luu P, and Posner MI, “Cognitive and emotional influences in anterior cingulate cortex,” *Trends Cogn. Sci*, vol. 4, no. 6, pp. 215–222, 2000. [PubMed: 10827444]
- [44]. Onitsuka T et al. , “Middle and inferior temporal gyrus gray matter volume abnormalities in chronic schizophrenia: An MRI study,” *Amer. J. Psychiatry*, vol. 161, no. 9, pp. 1603–1611, 2004. [PubMed: 15337650]
- [45]. Wylie KP and Tregellas JR, “The role of the insula in schizophrenia,” *Schizophrenia Res*, vol. 123, no. 2–3, pp. 93–104, 2010.
- [46]. Klumpp H, Keller J, Miller GA, Casas BR, Best JL, and Deldin PJ, “Semantic processing of emotional words in depression and schizophrenia,” *Int. J. Psychophysiol*, vol. 75, no. 2, pp. 211–215, 2010. [PubMed: 20006969]
- [47]. Anand KS and Dhikav V, “Hippocampus in health and disease: An overview,” *Ann. Indian Acad. Neurol*, vol. 15, no. 4, 2012, Art. no. 239.

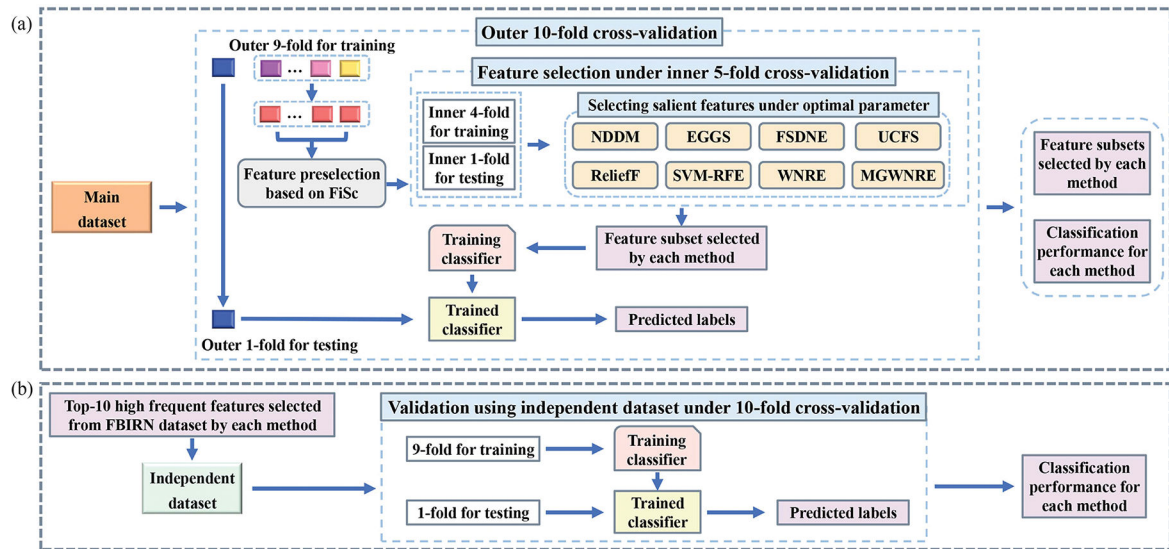


Fig. 1.

The experimental pipelines for the proposed methods (i.e., WNRE and MGWNRE) and the competing methods (i.e., NDDM, EGGS, FSDNE, UCFS, ReliefF, and SVM-RFE). (a) Denotes the outer 10-fold cross-validation with embedded inner 5-fold cross-validation experimental pipeline of each method on main datasets (i.e., the SRBCT, Colon, BRCA, and FBIRN datasets). (b) Denotes a 10-fold cross-validation experimental pipeline on independent datasets (i.e., the BSNIP and MPRC datasets) using the top-10 high-frequency features selected from the FBIRN dataset by each method. FiSc is the abbreviation for Fisher score-based feature selection method, which is used to preselect candidate features to improve computing efficiency. Classification performance includes accuracy, sensitivity, and specificity.

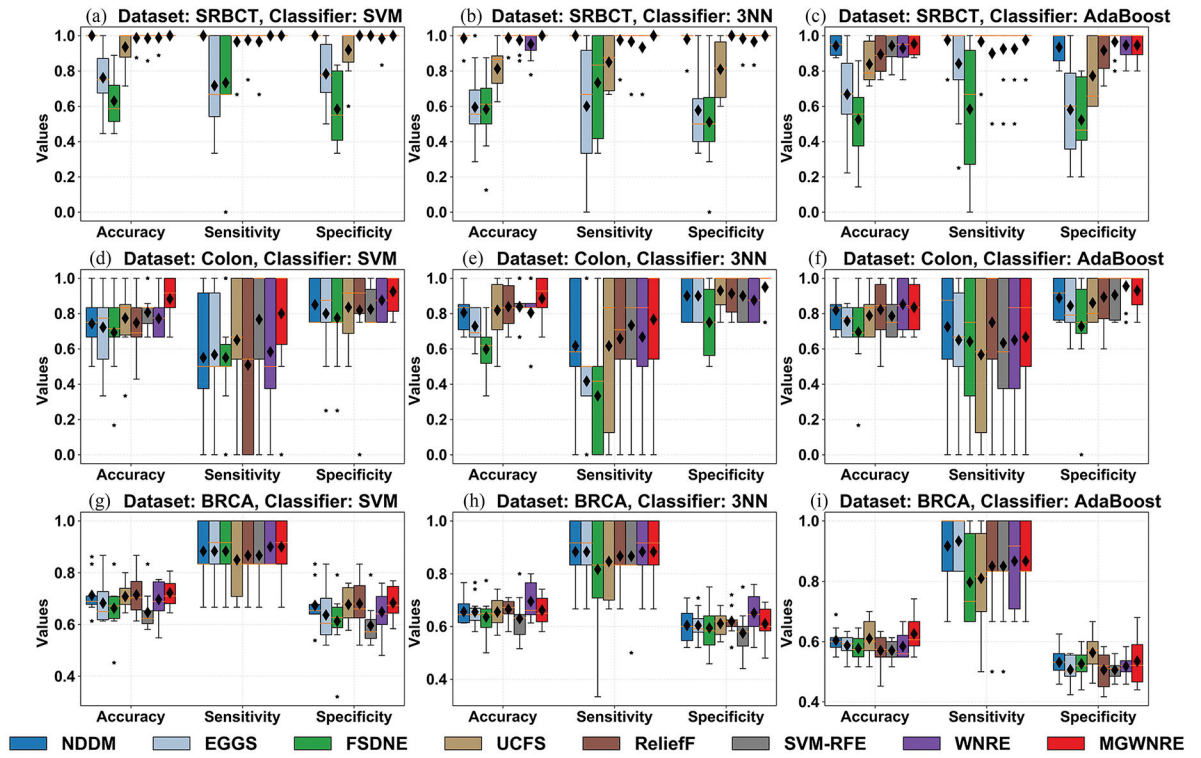


Fig. 2. Classification performance of various methods on omics datasets. (a–c) Show the classification performance on SRBCT dataset based on SVM, 3NN, and AdaBoost classifiers, respectively. (d–f) Show the classification performance on Colon dataset based on SVM, 3NN, and AdaBoost classifiers, respectively. (g–i) Show the classification performance on BRCA dataset based on SVM, 3NN, and AdaBoost classifiers, respectively. “◆” means the average value of 10-fold cross-validation results.

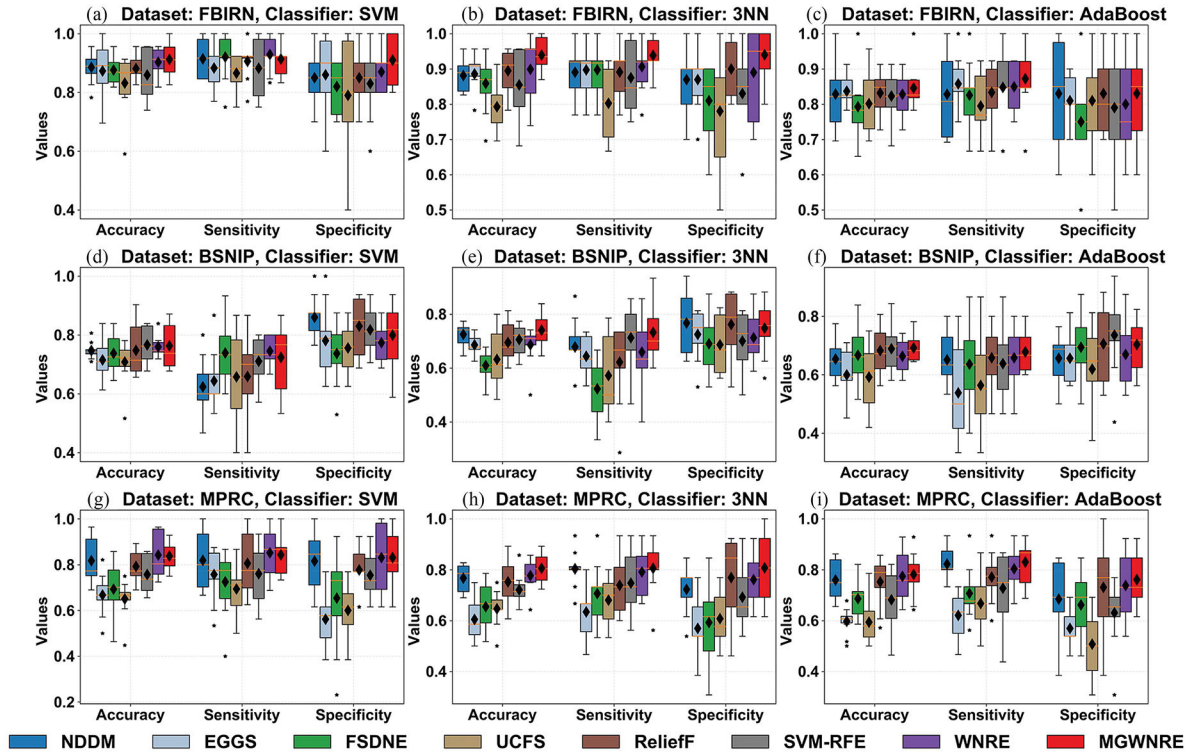


Fig. 3. Classification performance of various methods on fMRI datasets. (a–c) Show the classification performance on the main dataset (i.e., FBIRN) based on SVM, 3NN, and AdaBoost classifiers, respectively. (d–f) Show the classification performance on the independent dataset (i.e., BSNIP) based on SVM, 3NN, and AdaBoost classifiers, respectively. (g–i) Show the classification performance on the other independent dataset (i.e., MPRC) based on SVM, 3NN, and AdaBoost classifiers, respectively. “♦” means the average value of 10-fold cross-validation results.

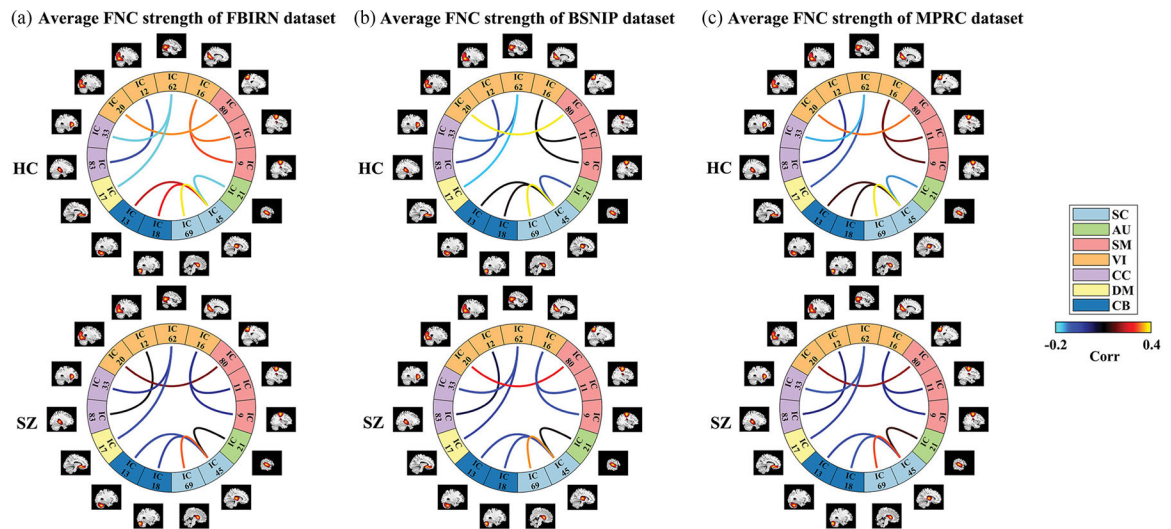


Fig. 4.

Average FNC strength of HC and SZ groups on the features selected by MGWNRE for (a) FBIRN dataset, (b) BSNIP dataset, and (c) MPRC dataset. The brain is divided into seven brain functional domains, including sub-cortical (SC), auditory (AU), sensorimotor (SM), visual (VI), cognitive-control (CC), default-mode (DM), and cerebellar (CB) domains. “Corr” represents the connectivity strength between different ICs. IC denotes the independent component representing one brain functional network.

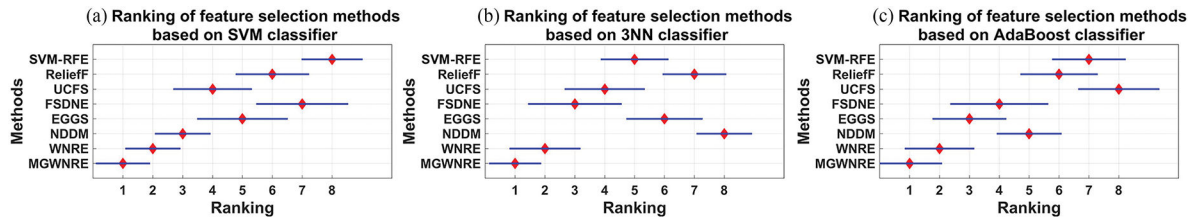


Fig. 5. Algorithm ranking chart of the comprehensive classification performance based on (a) SVM classifier, (b) 3NN classifier, and (c) AdaBoost classifier for different feature selection methods on all datasets, including SRBCT, Colon, BRCA, FBIRN, BSNIIP, and MPRC datasets. For each method, the red diamond symbol indicates the ranking, and the length of the blue horizontal line represents the standard deviation of the comprehensive classification performances on multiple datasets.

Table 1

Classification Performance and the Number of Selected Features of Various Methods on Omics Datasets, Including SRBCT, Colon, and BRCA. THE Used Classifiers Include SVM, 3NN, and AdaBoost

Datasets	Methods	SVM			3NN			AdaBoost			Num
		Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	
SRBCT	NDDM	100.0±0.0	100.0±0.0	100.0±0.0	98.6±4.5	100.0±0.0	98.0±6.3	94.3±6.1	97.5±7.9	93.5±8.6	54
	EGGS	76.1±17.8	71.7±27.3	78.3±17.5	59.4±22.2	60.0±34.4	57.7±24.3	66.7±22.6	84.2±27.3	58.0±26.2	25
	FSDNE	62.9±15.3	73.3±30.6	58.3±20.9	58.4±22.6	73.3±30.6	51.0±27.2	52.5±21.1	58.3±36.0	52.2±23.4	9
	UCFS	93.5±9.6	96.7±10.5	92.0±14.0	81.3±12.4	85.0±16.1	80.9±16.5	83.8±12.1	96.7±10.5	77.1±20.0	30
	ReliefF	98.8±4.0	97.5±7.9	100.0±0.0	98.8±4.0	97.5±7.9	100.0±0.0	89.4±10.2	90.0±21.1	91.7±11.2	53
	SVM-RFE	98.6±4.5	96.7±10.5	100.0±0.0	97.5±5.4	96.7±10.5	98.3±5.3	94.3±8.0	92.5±16.9	96.6±7.4	20
Colon	WNRE	98.9±3.5	100.0±0.0	98.3±5.3	95.2±8.1	93.3±14.1	96.7±7.0	92.9±8.5	92.5±16.9	94.6±8.9	114
	MGWNRE	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	95.4±6.0	97.5±7.9	94.6±8.9	53
	NDDM	74.3±15.3	55.0±37.7	85.0±12.9	80.5±10.8	61.7±39.3	90.0±12.9	81.9±12.4	72.5±35.1	88.8±12.0	36
	EGGS	72.1±22.2	56.7±37.0	80.0±25.8	72.9±9.7	41.7±28.6	90.0±12.9	75.7±8.7	65.0±30.9	84.3±14.6	23
	FSDNE	69.3±22.8	55.0±29.4	77.5±24.9	59.8±17.5	33.3±33.3	75.0±20.4	69.5±22.3	64.2±39.7	72.7±29.3	10
	UCFS	77.4±19.4	65.0±38.8	83.5±19.4	81.9±16.6	61.7±45.8	93.0±11.4	78.8±11.5	56.7±43.9	86.0±12.2	82
BRCA	ReliefF	74.8±17.4	50.8±48.2	81.8±30.6	83.8±13.0	65.8±39.0	91.3±11.4	82.1±16.7	75.0±36.2	89.3±14.9	98
	SVM-RFE	80.7±10.1	76.7±34.4	82.5±12.1	83.8±11.2	73.3±33.5	90.0±12.9	78.6±13.8	63.3±35.8	90.5±12.3	20
	WNRE	77.1±11.3	58.3±40.3	87.5±13.2	80.5±17.3	66.7±40.8	87.5±13.2	85.2±14.6	65.0±41.9	95.5±9.6	81
	MGWNRE	88.3±13.7	80.0±35.0	92.5±12.1	88.6±13.6	76.7±34.4	95.0±10.5	83.6±13.6	66.7±40.8	93.0±11.4	70
	NDDM	71.3±7.7	88.3±11.2	67.4±8.3	65.6±5.1	88.3±13.7	60.4±6.9	60.4±4.1	91.7±11.8	53.1±5.0	72
	EGGS	68.2±8.2	88.3±13.7	63.7±10.5	65.6±5.9	88.3±13.7	60.3±5.8	58.7±3.5	93.3±8.6	50.6±4.9	142
Colon	FSDNE	66.3±9.7	88.3±13.7	61.2±12.2	63.6±7.6	81.7±21.4	59.5±9.7	57.7±4.1	79.7±15.3	52.6±5.2	57
	UCFS	70.8±4.6	85.0±14.6	67.7±7.0	65.6±5.7	84.7±14.7	61.1±5.0	61.0±6.2	81.0±16.6	56.3±5.9	339
	ReliefF	71.5±7.7	86.7±13.1	68.0±9.7	66.6±4.0	86.7±13.1	61.9±5.4	57.1±5.7	85.0±16.6	50.6±5.9	198
	SVM-RFE	64.6±7.6	86.7±13.1	59.6±8.1	63.0±8.1	86.7±15.3	57.5±8.4	57.1±3.6	85.0±16.6	50.6±3.3	126
	WNRE	69.6±7.3	90.0±8.6	64.9±8.6	69.5±7.0	88.3±13.7	65.2±7.9	58.4±4.6	86.7±15.3	51.8±3.7	144
	MGWNRE	72.2±5.5	90.0±11.7	68.4±6.5	66.2±5.8	88.3±13.7	61.1±6.3	62.6±6.4	86.7±13.1	53.5±8.1	72

Bold numbers represent the best classification results among the eight feature selection methods. Num indicates the average number of selected features by each method.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table II

Classification Performance and the Number of Selected Features of Various Methods on fMRI Datasets, Including FBIRN, BSNIP, and MPRC. THE Used Classifiers Include SVM, 3NN, and AdaBoost

Datasets	Methods	SVM			3NN			AdaBoost			Num
		Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	
FBIRN	NDDM	88.6±5.5	91.4±6.7	85.0±8.5	88.1±4.5	89.0±6.6	87.0±9.5	82.9±9.7	82.8±12.5	83.0±14.9	24
	EGGS	87.3±9.7	88.3±8.4	86.0±13.5	88.6±5.8	89.7±7.4	87.0±8.2	83.7±4.1	85.8±7.3	81.0±7.4	66
	FSDNE	87.6±5.1	92.1±7.7	82.0±9.2	85.9±8.0	89.7±7.4	81.0±11.0	79.3±9.0	82.6±9.3	75.0±12.7	82
	UCFS	83.2±9.8	86.5±6.7	79.0±19.7	79.3±6.2	80.3±10.1	78.0±16.9	80.2±8.8	79.5±9.4	81.0±12.9	139
	ReliefF	88.1±4.1	90.6±6.0	85.0±8.5	89.5±6.2	89.1±8.4	90.0±8.2	83.2±5.7	83.3±8.3	83.0±11.6	56
	SVM-RFE	85.9±8.7	88.3±10.0	83.0±11.6	85.5±10.1	87.5±10.0	83.0±14.2	82.3±6.5	84.8±10.8	79.0±14.5	33
BSNIP	WIN RE	90.3±5.0	92.9±5.9	87.0±6.7	89.9±9.0	90.6±7.0	89.0±13.7	82.8±6.1	84.9±7.2	80.0±14.1	55
	MGWNRE	91.2±5.4	91.3±5.9	91.0±8.8	93.9±5.1	93.8±4.9	94.0±7.0	84.6±6.3	87.2±9.6	83.0±13.4	21
	NDDM	74.7±2.7	62.4±9.7	86.0±7.3	72.4±4.2	67.9±10.2	76.7±11.3	65.4±7.1	65.1±9.2	65.6±8.5	10
	EGGS	71.5±6.5	64.4±9.7	78.1±12.1	68.6±3.4	64.4±7.3	72.5±9.2	59.9±9.2	53.8±16.8	65.6±8.4	10
	FSDNE	73.7±6.3	73.9±11.0	73.7±10.1	61.0±5.2	52.3±11.4	69.0±11.5	66.7±9.2	63.6±16.0	69.4±11.8	10
	UCFS	70.8±7.6	65.8±14.6	75.6±9.2	63.2±10.8	57.2±14.2	68.7±10.7	59.2±11.2	56.4±15.8	61.9±14.8	10
MPRC	ReliefF	74.8±9.0	65.9±14.4	83.0±10.0	69.5±7.2	62.2±14.6	76.1±12.4	68.3±7.8	65.8±11.0	70.6±13.6	10
	SVM-RFE	76.6±6.2	71.0±8.2	81.7±7.2	70.5±5.8	71.2±11.6	70.0±10.7	68.9±8.1	63.8±11.7	73.5±13.3	10
	WNRE	76.0±3.4	74.5±5.3	77.4±6.2	68.6±7.1	65.9±13.0	71.2±9.3	66.3±5.9	65.8±9.5	66.9±10.7	10
	MGWNRE	76.3±7.8	72.4±12.3	80.0±12.1	74.0±5.9	73.2±10.3	74.8±9.6	69.2±4.6	67.8±6.7	70.4±9.2	10
	NDDM	81.7±9.5	81.9±12.3	81.5±12.1	76.7±5.3	80.4±7.1	72.3±9.0	76.0±8.1	82.3±6.4	68.5±16.0	10
	EGGS	66.8±8.9	75.8±11.4	56.2±12.6	60.5±8.4	63.4±10.6	56.9±11.6	59.7±5.3	62.1±9.6	56.9±6.5	10
FBIRN	FSDNE	69.3±12.1	72.5±14.6	65.4±20.9	65.4±9.2	70.6±11.9	59.2±15.8	68.6±7.2	70.8±10.2	66.2±13.2	10
	UCFS	65.1±8.1	69.4±10.6	60.0±12.5	64.7±6.9	68.0±9.7	60.8±11.7	59.4±8.5	66.8±10.9	50.8±13.2	10
	ReliefF	79.2±7.1	80.6±13.4	77.7±9.9	75.3±8.6	73.9±11.4	76.9±17.0	75.2±9.4	77.1±9.5	73.1±17.5	10
	SVM-RFE	75.7±8.4	76.0±13.3	75.4±9.5	72.1±6.0	74.8±11.8	69.2±13.1	68.2±11.3	72.7±14.2	63.1±13.5	10
	WNRE	84.2±10.1	85.2±10.3	83.1±14.9	77.7±6.0	79.1±10.1	76.2±9.2	77.4±9.7	80.4±7.6	73.8±14.6	10
	MGWNRE	83.8±6.1	84.3±8.9	83.1±11.4	80.6±6.3	80.6±10.0	80.8±13.2	78.1±8.7	83.0±8.7	76.2±11.7	10

Bold numbers represent the best classification results among the eight feature selection methods. Num indicates the average number of selected features by each method.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Group Differences of the top-10 Important FNC Features Selected By MGWNRE on Different Datasets, Including FBIRN, BSNIP, and MPRC Datasets. For Each Dataset, Two-Sample T-Tests Were Used to Examine the Group Differences for Each Selected FNC Feature, and the P-Values and T-Values of HC VS. SZ are Shown

TABLE III

Brain functional network (IC ID, brain functional domain)	Brain functional network (IC ID, brain functional domain)	FBIRN HC vs. SZ		BSNIP HC vs. SZ		MPRC HC vs. SZ	
		P-value	T-value	P-value	T-value	P-value	T-value
Caudate (69, SC)	Thalamus (45, SC)	1.03e-20	10.33	8.27e-06	4.53	7.14e-05	4.03
Thalamus (45, SC)	Superior temporal gyms (21, AU)	1.33e-29	-13.13	4.28e-17	-8.92	2.49e-23	-10.91
Thalamus (45, SC)	Cerebellum (13, CB)	1.30e-24	11.57	3.39e-14	7.95	9.48e-18	9.18
Thalamus (45, SC)	Cerebellum (18, CB)	8.67e-26	11.94	1.40e-18	9.38	1.67e-24	11.25
Left postcentral gyrus (9, SM)	Calcarine gyrus (16, VI)	6.13e-26	11.99	1.33e-14	8.09	1.52e-11	7.04
Right postcentral gyrus (11, SM)	Calcarine gyrus (16, VI)	1.24e-25	11.89	5.56e-15	8.22	8.21e-12	7.14
Superior parietal lobule (80, SM)	Inferior occipital gyrus (20, VI)	9.15e-08	5.52	1.79e-08	5.78	9.23e-07	5.02
Middle temporal gyrus (62, VI)	Insula (33, CC)	2.36e-11	-7.04	4.67e-03	-2.85	6.34e-03	-2.75
Middle temporal gyrus (62, VI)	Anterior cingulate cortex (17, DM)	7.49e-15	-8.34	1.40e-07	-5.39	1.48e-02	-2.45
Right middle occipital gyrus (12, VI)	Hippocampus (83, CC)	2.97e-11	-7.00	1.20e-07	-5.42	1.51e-01	-1.44

Here, IC denotes the independent component. The brain is divided into seven brain functional domains, including sub-cortical (SC), auditory (AU), sensorimotor (SM), visual (VI), cognitive-control (CC), default-mode (DM), and cerebellar (CB) domains.