Robust Exclusive Adaptive Sparse Feature Selection for Biomarker Discovery and Early Diagnosis of Neuropsychiatric Systemic Lupus Erythematosus

Tianhong Quan[1], Ye Yuan[2], Yu Luo[3], Teng Zhou[4,5(✉)], and Jing Qin[4]

[1] School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China
[2] College of Engineering, Shantou University, Shantou, China
[3] School of Computer Science and Technology, Guangdong University of Technology, Guangzhou, China
[4] Centre for Smart Health, The Hong Kong Polytechnic University, Hung Hom, Hong Kong
teng.zhou@polyu.edu.hk
[5] School of Cyberspace Security, Hainan University, Haikou, China

Abstract. The symptoms of neuropsychiatric systemic lupus erythematosus (NPSLE) are subtle and elusive at the early stages. ^{1}H-MRS (proton magnetic resonance spectrum) imaging technology can detect more detailed early appearances of NPSLE compared with conventional ones. However, the noises in ^{1}H-MRS data often bring bias in the diagnostic process. Moreover, the features of specific brain regions are positively correlated with a certain category but may be redundant for other categories. To overcome these issues, we propose a robust exclusive adaptive sparse feature selection (REASFS) algorithm for early diagnosis and biomarker discovery of NPSLE. Specifically, we employ generalized correntropic loss to address non-Gaussian noise and outliers. Then, we develop a generalized correntropy-induced exclusive $\ell_{2,1}$ regularization to adaptively accommodate various sparsity levels and preserve informative features. We conduct sufficient experiments on a benchmark NPSLE dataset, and the experimental results demonstrate the superiority of our proposed method compared with state-of-the-art ones.

Keywords: NPSLE · Feature Selection · non-Gaussian Noise

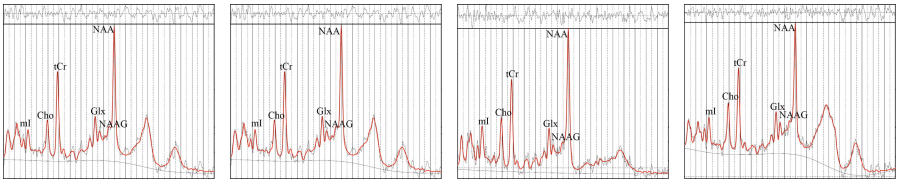
1 Introduction

Neuropsychiatric systemic lupus erythematosus (NPSLE) refers to a complex autoimmune disease that damages the brain nervous system of patients. The clinical symptoms of NPSLE include cognitive disorder, epilepsy, mental illness, etc., and patients with NPSLE have a nine-fold increased mortality compared to the general population [11]. Since the pathogenesis and mature treatment of NPSLE have not yet been found, it is extremely important to detect NPSLE at

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14224, pp. 127–135, 2023.
https://doi.org/10.1007/978-3-031-43904-9_13

its early stage and put better clinical interventions and treatments to prevent its progression. However, the high overlap of clinical symptoms with other psychiatric disorders and the absence of early non-invasive biomarkers make accurate diagnosis difficult and time-consuming [3].

Although conventional magnetic resonance imaging (MRI) tools are widely used to detect brain injuries and neuronal lesions, around 50% of patients with NPSLE present no brain abnormalities in structural MRI [17]. In fact, metabolic changes in many brain diseases precede pathomorphological changes, which indicates proton magnetic resonance spectroscopy (^{1}H-MRS) to be a more effective way to reflect the early appearance of NPSLE. ^{1}H-MRS is a non-invasive neuroimaging technology that can quantitatively analyze the concentration of metabolites and detect abnormal metabolism of the nervous system to reveal brain lesions. However, the complex noise caused by overlapping metabolite peaks, incomplete information on background components, and low signal-to-noise ratio (SNR) disturb the analysis results of this spectroscopic method [15]. Meanwhile, the individual differences in metabolism and the interaction between metabolites under low sample size make it difficult for traditional learning methods to distinguish NPSLE. Figure 1 shows spectra images of four participants including healthy controls (HC) and patients with NPSLE. It can be seen that the visual differences between patients with NPSLE and HCs in the spectra of the volumes are subtle. Therefore, it is crucial to develop effective learning algorithms to discover metabolic biomarkers and accurately diagnose NPSLE.



(a) HC1 (b) HC2 (c) NPSLE1 (d) NPSLE2

Fig. 1. Images (a) to (d) represent the spectra of the volumes obtained by LCModel for four individual participants. Abbreviation: N-acetylaspartate (NAA), N-acetylaspartylglutamate (NAAG), choline (Cho), total creatine (tCr), myo-inositol (mI), glutamate+glutamine (Glu + Gln = Glx).

The machine learning application for biomarker analysis and early diagnosis of NPSLE is at a nascent stage [4]. Most studies focus on the analysis of MR images using statistical or machine learning algorithms, such as Mann-Whitney U test [8], support vector machine (SVM) [7,24], ensemble model [16,22], etc. Generally, Machine learning algorithms based on the minimum mean square error (MMSE) criterion heavily rely on the assumption that noise is of Gaussian distribution. However, measurement-induced non-Gaussian noise in ^{1}H-MRS data undoubtedly limits the performance of MMSE-based machine learning methods.

On the other hand, for the discovery task of potential biomarkers, sparse coding-based methods (e.g., $\ell_{2,1}$ norm, $\ell_{2,0}$ norm, etc.) force row elements to zero that remove some valuable features [12,21]. More importantly, different brain regions have different functions and metabolite concentrations, which implies that the metabolic features for each brain region have different sparsity levels. Therefore, applying the same sparsity constraint to the metabolic features of all brain regions may not contribute to the improvement of the diagnostic performance of NPSLE.

In light of this, we propose a robust exclusive adaptive sparse feature selection (REASFS) algorithm to jointly address the aforementioned problems in biomarker discovery and early diagnosis of NPSLE. Specifically, we first extend our feature learning through generalized correntropic loss to handle data with complex non-Gaussian noise and outliers. We also present the mathematical analysis of the adaptive weighting mechanism of generalized correntropy. Then, we propose a novel regularization called generalized correntropy-induced exclusive $\ell_{2,1}$ to adaptively accommodate various sparsity levels and preserve informative features. The experimental results on a benchmark NPSLE dataset demonstrate the proposed method outperforms comparing methods in terms of early non-invasive biomarker discovery and early diagnosis.

2 Method

Dataset and Preprocessing: The T2-weighted MR images of 39 participants including 23 patients with NPSLE and 16 HCs were gathered from our affiliated hospital. All images were acquired at an average age of 30.6 years on a SIGNA 3.0T scanner with an eight-channel standard head coil. Then, the MR images were transformed into spectroscopy by multi-voxel ^{1}H-MRS based on a point-resolved spectral sequence (PRESS) with a two-dimensional multi-voxel technique. The collected spectroscopy data were preprocessed by a SAGE software package to correct the phase and frequency. An LCModel software was used to fit the spectra, correct the baseline, relaxation, and partial-volume effects, and quantify the concentration of metabolites. Finally, we used the absolute NAA concentration in single-voxel MRS as the standard to gain the absolute concentration of metabolites, and the NAA concentration of the corresponding voxel of multi-voxel ^{1}H-MRS was collected consistently. The spectra would be accepted if the SNR is greater than or equal to 10 and the metabolite concentration with standard deviations (SD) is less than or equal to 20%. The absolute metabolic concentrations, the corresponding ratio, and the linear combination of the spectra were extracted from different brain regions: RPCG, LPCG, RDT, LDT, RLN, LLN, RI, RPWM, and LPWM. A total of 117 metabolic features were extracted, and each brain region contained 13 metabolic features: Cr, phosphocreatine (PCr), Cr+PCr, NAA, NAAG, NAA+NAAG, NAA+NAAG/Cr+PCr, mI, mI/Cr+PCr, Cho+phosphocholine (PCh), Cho+PCh/Cr+PCr, Glu+Gln, and Glu+Gln/Cr+PCr.

2.1 Sparse Coding Framework

Given a data matrix $\mathbf{X} = [\mathbf{x}^1; \cdots; \mathbf{x}^n] \in \mathbb{R}^{n \times d}$ with n sample, the i-th row is represented by $\mathbf{x}^i$, and the corresponding label matrix is denoted as $\mathbf{Y} = [\mathbf{y}^1; \cdots; \mathbf{y}^n] \in \mathbb{R}^{n \times k}$, where $\mathbf{y}_i$ is one-hot vector. The Frobenius norm of projection matrix $\mathbf{W} \in \mathbb{R}^{d \times k}$ is $\|\mathbf{W}\|_F = \sqrt{\sum_{i=1}^{d} \sum_{j=1}^{k} W_{ij}^2}$. For sparse coding-based methods, the general problem can be formulated as

$$\min_{\mathbf{W}} J(\mathbf{W}) = L(\mathbf{W}) + \lambda R(\mathbf{W}), \tag{1}$$

where $L(\mathbf{W})$, $R(\mathbf{W})$, and λ are the loss function, the regularization term, and the hyperparameter, respectively. For least absolute shrinkage and selection operator (LASSO) [20], $L(\mathbf{W}) = \|\mathbf{Y} - \mathbf{X}\mathbf{W}\|_F^2$ and $R(\mathbf{W}) = \|\mathbf{W}\|_1$. For multi-task feature learning [6], $\ell_{2,1}$ norm is the most widely used regularization to select class-shared features via row sparsity, which is defined as $\|\mathbf{W}\|_{2,1} = \sum_{i=1}^{d} \|\mathbf{w}^i\|_2 = \sum_{i=1}^{d} \left(\sum_{j=1}^{k} W_{ij}^2 \right)^{1/2}$. Due to the row sparsity of $\ell_{2,1}$ norm, the features selected for all different classes are enforced to be exactly the same. Thus, the inflexibility of $\ell_{2,1}$ norm may lead to the deletion of meaningful features.

2.2 Generalized Correntropic Loss

Originating from information theoretic learning (ITL), the correntropy [2] is a local similarity measure between two random variables A and B, given by

$$V(A, B) = \mathbf{E}[k_\sigma(A, B)] = \int k_\sigma(a, b) dF_{AB}(a, b), \tag{2}$$

where $\mathbf{E}$, $k_\sigma(\cdot, \cdot)$, and $F_{AB}(a, b)$ denote the mathematical expectation, the Gaussian kernel, and the joint probability density function of (A, B), respectively.

When applying correntropy to the error criterion, the boundedness of the Gaussian kernel limits the disturbance of large errors caused by outliers on estimated parameters. However, the kernelized second-order statistic of correntropy is not suitable for all situations. Therefore, the generalized correntropy [1], a more flexible and powerful form of correntropy, was proposed by substituting the generalized Gaussian density (GGD) function for the Gaussian kernel in correntropy, and the GGD function is defined as

$$G_{\alpha,\beta}(e) = \frac{\alpha}{2\beta\Gamma\left(\frac{1}{\alpha}\right)} \exp\left(-|\frac{e}{\beta}|^\alpha \right) = \gamma \exp\left(-s|e|^\alpha\right), \tag{3}$$

where $\Gamma(\cdot)$ is the gamma function, α, $\beta > 0$ are the shape and bandwidth parameters, respectively, $s = 1/\beta^\alpha$ is the kernel parameter, $\gamma = \alpha/(2\beta\Gamma(1/\alpha))$ is the normalization factor. Specifically, when $\alpha = 2$ and $\alpha = 1$, GGD degenerates to Gaussian distribution and Laplacian distribution, respectively. As an adaptive similarity measure, generalized correntropy can be applied to machine learning

and adaptive systems [14]. The generalized correntropic loss function between A and B can be defined as

$$L_{GC}(A,B) = G_{\alpha,\beta}(0) - V_{\alpha,\beta}(A,B). \tag{4}$$

To analyze the adaptive weighting mechanism of generalized correntropy, we consider an alternative problem of (1), where $L(\mathbf{W}) = \hat{V}_{\alpha,\beta}(\mathbf{Y}, \mathbf{XW}) = \sum_{i=1}^{n} \exp(-s\|\mathbf{y}^i - \mathbf{x}^i\mathbf{W}\|_2^\alpha)$ and $R(\mathbf{W}) = \|\mathbf{W}\|^2 = \sum_{i=1}^{d} \|\mathbf{w}^i\|_2^2$. The optimal projection matrix $\mathbf{W}$ should satisfy $(\partial J(\mathbf{W})/\partial \mathbf{W}) = 0$, and we have

$$\mathbf{W} = \left(\mathbf{X}^T \boldsymbol{\Lambda} \mathbf{X} + \lambda \mathbf{I}\right)^{-1} \mathbf{X}^T \boldsymbol{\Lambda} \mathbf{Y}, \tag{5}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix with error-based diagonal elements $\Lambda_{ii} = \exp(-s\|\mathbf{y}^i - \mathbf{x}^i\mathbf{W}\|_2^\alpha)\|\mathbf{y}^i - \mathbf{x}^i\mathbf{W}\|_2^{\alpha-2}$ for adaptive sample weight.

2.3 Generalized Correntropy-Induced Exclusive $\ell_{2,1}$

To overcome the drawback of $\ell_{2,1}$ norm and achieve adaptive sparsity regularization on metabolic features of different brain regions, we propose a novel GCIE $\ell_{2,1}$. A flexible feature learning algorithm exclusive $\ell_{2,1}$ [9] is defined as

$$\min_{\mathbf{W}} J(\mathbf{W}) = L(\mathbf{W}) + \lambda_1 \|\mathbf{W}\|_{2,1} + \lambda_2 \|\mathbf{W}\|_{1,2}^2, \tag{6}$$

where $\|\mathbf{W}\|_{1,2}^2 = \sum_{i=1}^{d} \|\mathbf{w}^i\|_1^2 = \sum_{i=1}^{d} \left(\sum_{j=1}^{k} |W_{ij}|\right)^2$. Based on the exclusive $\ell_{2,1}$, we can not only removes the redundant features shared by all classes through row sparsity of $\ell_{2,1}$ norm but also selects different discriminative features for each class through exclusive sparsity of $\ell_{1,2}$ norm. Then we propose to introduce generalized correntropy to measure the sparsity penalty in the feature learning algorithm. We apply generalized correntropy to the row vector $\mathbf{w}^i$ of $\mathbf{W}$ to achieve the adaptive weighted sparse constraint, and the problem (6) can be rewritten as

$$\min_{\mathbf{W}} J(\mathbf{W}) = \hat{L}_{GC}(\mathbf{Y}, \mathbf{XW}) + \lambda_1 \sum_{i=1}^{d} (1 - \exp(-s\|\mathbf{w}^i\|_2)) + \lambda_2 \|\mathbf{W}\|_{1,2}^2, \tag{7}$$

where $\hat{L}_{GC}(\mathbf{Y}, \mathbf{XW}) = \sum_{i=1}^{n} (1 - exp(-s\|\mathbf{y}^i - \mathbf{x}^i\mathbf{W}\|_2^\alpha))$. Since minimizing $\|\mathbf{w}^i\|_2$ is equivalent to maximizing $\exp(-s\|\mathbf{w}^i\|_2)$, we add a negative sign to this term. Through the GCIE $\ell_{2,1}$ regularization term in (7), each feature is expected to be enforced with a sparsity constraint of a different weight according to its sparsity level of metabolic information in different brain regions.

Optimization: Since the GCIE $\ell_{2,1}$ is a non-smooth regularization term, the final problem (7) can be optimized by the stochastic gradient method with appropriate initialization of $\mathbf{W}$. To this end, We use the closed-form solution (5) to

initialize $\mathbf{W}$ reasonably, and the error-driven sample weight has reached the optimum based on half-quadratic analysis [13]. Once we obtained the solution to the problem (7), the importance of feature i is proportional to $\|\mathbf{w}_i\|_2$. We then rank the importance of features according to $(\|\mathbf{w}_i\|_2, \cdots, \|\mathbf{w}_d\|_2)$, and select the top m ranked features from the sorted order for further classification.

3 Experimental Results and Conclusion

Experimental Settings: The parameters α and λ_1 are are set to 1, while β and λ_2 are searched form $\{0.5, 1, 2, 5\}$ and $\{0.1, 0.5, 1\}$, respectively. We use Adam optimizer and the learning rate is 0.001. To evaluate the performance of classification, we employ a support vector machine as the basic classifier, where the kernel is set as the radial basis function (RBF) and parameter C is set to 1. We average the 3-fold cross-validation results.

Results and Discussion: We compare the classification accuracy of the proposed REASFS with several SOTA baselines, including two filter methods: maximal information coefficient (MIC) [5], Gini [23], and four sparse coding-based methods: multi-task feature learning via $\ell_{2,1}$ norm [6,12], discriminative feature selection via $\ell_{2,0}$ norm [21], feature selection via $\ell_{1,2}$ norm [10] and exclusive $\ell_{2,1}$ [9]. The proposed REASFS is expected to have better robustness and flexibility. It can be seen from Fig. 2 that the sparse coding-based methods achieve better performance than filter methods under most conditions, where "0%" represents no noise contamination. The highest accuracy of our REASFS demonstrates the effectiveness and flexibility of the proposed GCIE $\ell_{2,1}$.

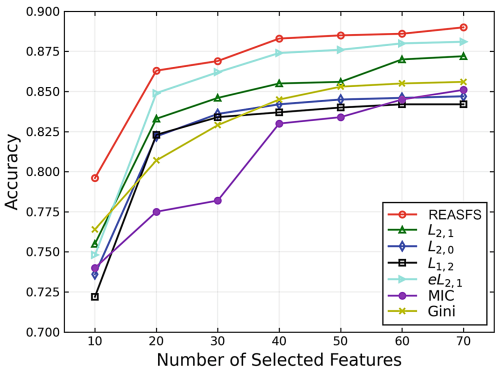
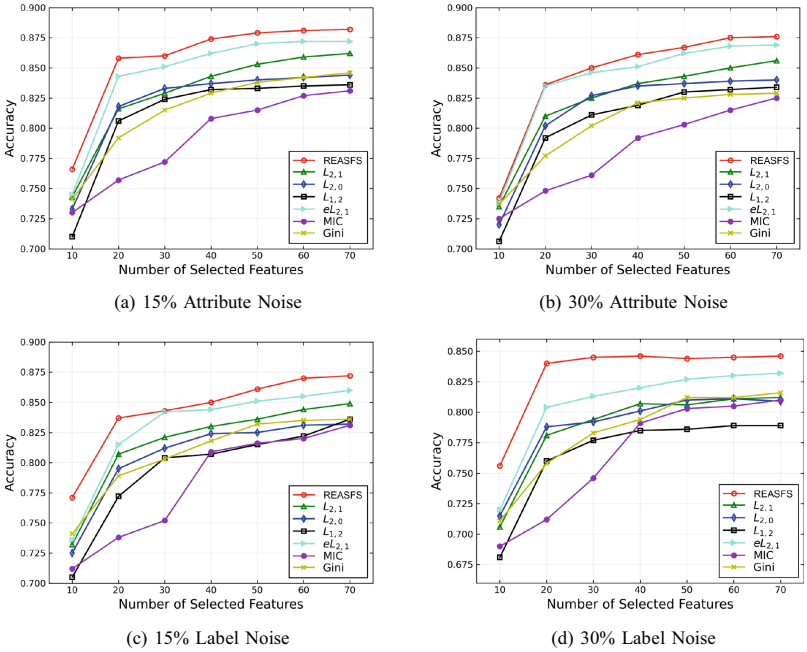


Fig. 2. REASFS versus SOTA baselines on NPSLE dataset (0% noise).

Generally speaking, the probability of samples being contaminated by random noise is equal. Therefore, we randomly select features from the training set and replace the selected features with pulse noise. The number of noisy attributes

is denoted by the ratio between the numbers of selected features and total features, such as 15% and 30%. The classification performances of the NPSLE dataset contaminated by attribute noise are shown in Fig. 3(a) and Fig. 3(b), where one clearly perceives that our REASFS achieves the highest accuracy under all conditions. Besides, it is unreasonable to apply the same level of sparse regularization to noise features and uncontaminated features, and our GCIE $\ell_{2,1}$ can adaptively increase the sparse level of noise features to remove redundant information, and vice versa. For label noise, we randomly select samples from the training set and replace classification labels of the selected samples with opposite values, *i.e.*, $0 \rightarrow 1$ and $1 \rightarrow 0$. The results are shown in Fig. 3(c) and Fig. 3(d), where the proposed REASFS is superior to other baselines. It can be seen from Fig. 3 that our REASFS achieves the highest accuracy in different noisy environments, which demonstrates the robustness of generalized correntropic loss.



(a) 15% Attribute Noise (b) 30% Attribute Noise

(c) 15% Label Noise (d) 30% Label Noise

Fig. 3. REASFS versus SOTA baselines on NPSLE dataset with different noise.

For non-invasive biomarkers, our method shows that some metabolic features contribute greatly to the early diagnosis of NPSLE, i.e., NAAG, mI/Cr+PCr, and Glu+Gln/Cr+PCr in RPCG; Cr+PCr, NAA+NAAG, NAA+NAAG/Cr+PCr, mI/Cr+PCr and Glu+Gln in LPCG; NAA, NAAG, and Cho+PCh in LDT; PCr, Cr+PCr, Cho+PCh, Cho+PCh/Cr+PCr and Glu+Gln/Cr+PCr in RLN; MI/Cr+PCr, Cho+PCh and Cho+PCh/Cr+PCr in LLN; NAA+NAAG/Cr+PCr and Cho+PCh in RI; Cho+PCh/Cr+PCr and

Glu+Gln/Cr+PCr in RPWM; And PCr, NAAG and NAA+NAAG/Cr+PCr in LPWM. Moreover, we use isometric feature mapping (ISOMAP) [19] to analyze these metabolic features and find that this feature subset is essentially a low-dimensional manifold. Meanwhile, by combining the proposed REASFS and ISOMAP, we can achieve 99% accuracy in the early diagnosis of NPSLE. In metabolite analysis, some studies have shown that the decrease in NAA concentration is related to chronic inflammation, damage, and tumors in the brain [18]. In the normal white matter area, different degrees of NPSLE disease is accompanied by different degrees of NAA decline, but structural MRI is not abnormal, suggesting that NAA may indicate the progress of NPSLE. We also found that Glu+Gln/Cr+PCr in RI decreased, which indicates that the excitatory neurotransmitter Glu in the brain of patients with NPSLE may have lower activity. To sum up, the proposed method provides a shortcut for revealing the pathological mechanism of NPSLE and early detection.

Conclusion: In this paper, we develop REASFS, a robust flexible feature selection that can identify metabolic biomarkers and detect NPSLE at its early stage from noisy ^{1}H-MRS data. The main advantage of our approach is its robust generalized correntropic loss and a novel GCIE $\ell_{2,1}$ regularization, which jointly utilizes the row sparsity and exclusive sparsity to adaptively accommodate various sparsity levels and preserve informative features. The experimental results show that compared with previous methods, REASFS plays a very important role in the biomarker discovery and early diagnosis of NPSLE. Finally, we analyze metabolic features and point out their clinical significance.

Acknowledgements. This work was supported by a grant of the Innovation and Technology Fund - Guangdong-Hong Kong Technology Cooperation Funding Scheme (No. GHP/051/20GD), the Project of Strategic Importance in The Hong Kong Polytechnic University (No. 1-ZE2Q), the 2022 Guangdong Basic and Applied Basic Research Foundation (No. 2022A1515011590), the National Natural Science Foundation of China (No. 61902232), and the 2020 Li Ka Shing Foundation Cross-Disciplinary Research Grant (No. 2020LKSFG05D).

References

1. Chen, B., Xing, L., Zhao, H., Zheng, N., Prı, J.C., et al.: Generalized correntropy for robust adaptive filtering. IEEE Trans. Signal Process. **64**(13), 3376–3387 (2016)
2. He, R., Zheng, W.S., Hu, B.G.: Maximum correntropy criterion for robust face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **33**(8), 1561–1576 (2010)
3. Jeltsch-David, H., Muller, S.: Neuropsychiatric systemic lupus erythematosus: pathogenesis and biomarkers. Nat. Rev. Neurol. **10**(10), 579–596 (2014)
4. Kingsmore, K.M., Lipsky, P.E.: Recent advances in the use of machine learning and artificial intelligence to improve diagnosis, predict flares, and enrich clinical trials in lupus. Curr. Opin. Rheumatol. **34**(6), 374–381 (2022)
5. Kinney, J.B., Atwal, G.S.: Equitability, mutual information, and the maximal information coefficient. Proc. Natl. Acad. Sci. **111**(9), 3354–3359 (2014)

6. Liu, J., Ji, S., Ye, J.: Multi-task feature learning via efficient l2, 1-norm minimization. In: Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, pp. 339–348 (2009)
7. Luo, X., et al.: Multi-lesion radiomics model for discrimination of relapsing-remitting multiple sclerosis and neuropsychiatric systemic lupus erythematosus. Eur. Radiol. **32**(8), 5700–5710 (2022)
8. Mackay, M., Tang, C.C., Vo, A.: Advanced neuroimaging in neuropsychiatric systemic lupus erythematosus. Curr. Opin. Neurol. **33**(3), 353 (2020)
9. Ming, D., Ding, C.: Robust flexible feature selection via exclusive l21 regularization. In: Proceedings of the 28th International Joint Conference on Artificial Intelligence, pp. 3158–3164 (2019)
10. Ming, D., Ding, C., Nie, F.: A probabilistic derivation of LASSO and L12-norm feature selections. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 4586–4593 (2019)
11. Monahan, R.C., et al.: Mortality in patients with systemic lupus erythematosus and neuropsychiatric involvement: a retrospective analysis from a tertiary referral center in the Netherlands. Lupus **29**(14), 1892–1901 (2020)
12. Nie, F., Huang, H., Cai, X., Ding, C.: Efficient and robust feature selection via joint $\ell 2$, 1-norms minimization. In: Advances in Neural Information Processing Systems, vol. 23 (2010)
13. Nikolova, M., Ng, M.K.: Analysis of half-quadratic minimization methods for signal and image recovery. SIAM J. Sci. Comput. **27**(3), 937–966 (2005)
14. Quan, T., Yuan, Y., Song, Y., Zhou, T., Qin, J.: Fuzzy structural broad learning for breast cancer classification. In: 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), pp. 1–4. IEEE (2022)
15. Ruiz-Rodado, V., Brender, J.R., Cherukuri, M.K., Gilbert, M.R., Larion, M.: Magnetic resonance spectroscopy for the study of CNS malignancies. Prog. Nucl. Magn. Reson. Spectrosc. **122**, 23–41 (2021)
16. Simos, N.J., et al.: Quantitative identification of functional connectivity disturbances in neuropsychiatric lupus based on resting-state fMRI: a robust machine learning approach. Brain Sci. **10**(11), 777 (2020)
17. Tamires Lapa, A., et al.: Reduction of cerebral and corpus callosum volumes in childhood-onset systemic lupus erythematosus: a volumetric magnetic resonance imaging analysis. Arthritis Rheumatol. **68**(9), 2193–2199 (2016)
18. Tannous, J., et al.: Altered neurochemistry in the anterior white matter of bipolar children and adolescents: a multivoxel 1h MRS study. Mol. Psychiatry **26**(8), 4117–4126 (2021)
19. Tenenbaum, J.B., Silva, V.D., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. Science **290**(5500), 2319–2323 (2000)
20. Tibshirani, R.: Regression shrinkage and selection via the LASSO. J. Roy. Stat. Soc.: Ser. B (Methodol.) **58**(1), 267–288 (1996)
21. Wang, Z., Nie, F., Tian, L., Wang, R., Li, X.: Discriminative feature selection via a structured sparse subspace learning module. In: IJCAI, pp. 3009–3015 (2020)
22. Yuan, Y., Quan, T., Song, Y., Guan, J., Zhou, T., Wu, R.: Noise-immune extreme ensemble learning for early diagnosis of neuropsychiatric systemic lupus erythematosus. IEEE J. Biomed. Health Inform. **26**(7), 3495–3506 (2022)
23. Zhang, S., Dang, X., Nguyen, D., Wilkins, D., Chen, Y.: Estimating feature-label dependence using Gini distance statistics. IEEE Trans. Pattern Anal. Mach. Intell. **43**(6), 1947–1963 (2019)
24. Zhuo, Z., et al.: Different patterns of cerebral perfusion in SLE patients with and without neuropsychiatric manifestations. Hum. Brain Mapp. **41**(3), 755–766 (2020)