Proceedings of CECNet 2022
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA220582

A Stability Evaluation of Feature Ranking Algorithms on Breast Cancer Data Analysis

Shaode YU [a], Bingjie LI [a], Boji LIU [a], Mingxue JIN [a], Junjie WU [b], and Hang YU [c,1]

[a] *School of Information and Communication Engineering, Communication University of China, Chaoyang, Beijing, China*

[b] *Department of Radiation Oncology, University of Texas Southwestern Medical Center, Dallas, Texas, United States*

[c] *School of Aerospace Science and Technology, Xidian University, Xi'an, Shanxi, China*

Abstract. Stability of feature preference is a most vital yet rarely explored characteristics of feature ranking algorithms. In this study, 23 feature rankers are evaluated on 4 breast cancer datasets (BCDR-F03, WDBC, GSE10810 and GSE15852) using an advanced stability estimator (S), and 3 rankers are identified showing good stability ($S \geq 0.55$) consistently on the four datasets. It suggests that data sufficiency is crucial for the construction of feature importance measure, since more rankers are stable on medical imaging datsets (BCDR-F03 and WDBC) than on gene expression datasets (GSE10810 and GSE15852), and high-dimensional small-sample-size datasets are big challenges of stability estimation. In our future work, more attention should be paid to the topics of developing stable feature ranking algorithms and stability estimators to well tackle different sizes of medical datasets.

Keywords. Stability, feature ranking, breast cancer, data analysis, matFR

1. Introduction

Feature ranking (FR) becomes increasingly important in the fields of precision medicine due to the dramatic growth of feature dimension [1]. As one of FR's most crucial characteristics, stability quantifies how different training sets affect its feature preferences [2].

Four studies have evaluated the stability of feature selection algorithms. For breast cancer (BC) risk prediction, 6 algorithms are analyzed using correlation coefficient and Jaccard index [3]. For colorectal cancer risk forecasting, 6 methods are assessed with two similarity-based estimators [4]. On high-dimensional datasets, the stability of 5 methods is estimated via correlation coefficients and adapted Tanimoto distance [5]. Besides, relative weighted consistency, partially adjusted average Tanimoto index and some other correlation based similarity measures are employed [6].

This study focuses on FR stability and differs itself from three points. First, 23 algorithms are evaluated that surpasses previous studies. Second, an advanced estimator [2]

[1] Corresponding Author: Hang Yu, Xidian University, Shanxi, China; E-mail: hyu@xidian.edu.cn.

is used to observe the dynamic change of stability. Third, stable algorithms are identified on 4 breast cancer datasets. The study could enrich our understanding of FR stability on diverse cancer data analysis.

2. Materials and Methods

2.1. Data collection

Four BC datasets shown in Table 1 are collected for FR stability analysis. BCDR-F03 includes 406 breast lesions (230 benign and 176 malignant)[2] and 17 features are provided [7]. To avoid one lesion with multiple records [8], the first feature record is used and 406 records are analyzed. WDBC contains 357 benign and 212 malignant instances[3]. For a digitized fine needle aspiration (FNA) image, based on multiple delineation, the mean, standard error and largest values of each feature are collected, and 30 features are formed [9]. GSE10810 provides 31 tumor and 27 control samples of valid specimens [10], and 18,382 genes are collected[4]. GSE15852 involves 43 tumor samples and 43 control samples of Malaysian women[5], and 22,283 gene points are detected [11].

Table 1. Summary of the datasets used in this study

	benign (train/test)	malignant (train/test)	feature number (p)	data source
BCDR-F03	230 (141/89)	176 (141/35)	17	MAM
WDBC	357 (170/187)	212 (170/42)	30	FNA
GSE10810	27 (22/5)	31 (22/9)	18382	gene
GSE15852	43 (34/9)	43 (34/9)	22283	gene

2.2. Experiment design on stability estimation

Figure 1 shows the experiment design. In each iteration, a dataset $\{(X,y)\}$ is divided into two subsets and one is for training $\{(X^{train},y^{train})\}$, and each method yields a feature rank in terms of feature importance. Here, vector $< f_{i,1},...,f_{i,k},...,f_{i,p} >$ is the output of the i^{th} running of p features of an algorithm, and $f_{i,k}$ is the order of the k^{th} feature. In this study, $N = 100$, and 100 iterations of each algorithm are conducted.

The outcome is the stability value (S) when top-m features are selected. An algorithm generates a rank in descending feature importance order. When the number (m) is defined, it yields a subset of features. In this study, $N = 30$, and m ranges from 3 to 9 features. When $m = 3$ and $S \geq 0.55$, the algorithm is assumed to be stable.

[2] http://bcdr.inegi.up.pt
[3] https://archive.ics.uci.edu/ml/datasets/
[4] https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE10810
[5] https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE15852

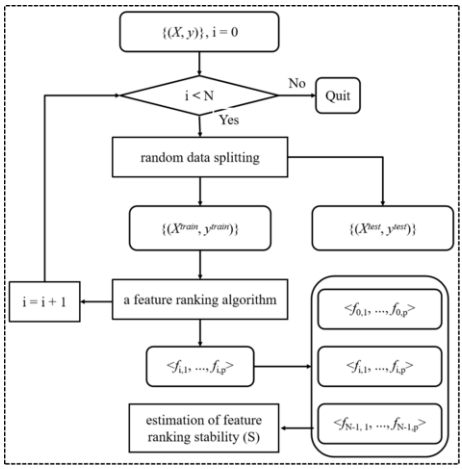

Figure 1. The procedure of stability estimation. (The figure can be enlarged for viewing.)

2.3. Feature ranking algorithms

Twenty-three methods in matFR[6] are evaluated, since the others exceed time expectation (≥ 0.5 hour per iteration) on GSE15852. The core ideas of algorithms are based on absolute values of *t*-test [12], relative entropy [13], Bhattacharyya distance [14], area between the empirical receiver operating characteristic curve and random classifier slope [15], absolute values of Mann-Whitney test [16], ReliefF [17], least absolute shrinkage and selection operator [18], correlation analysis [19], generalized Fisher score (GFS) [20], Gini score [21], Kruskal-Wallis test [22], pairwise feature proximity (PWFP) [23], min-max local structure information [24], local learning-based clustering [25], eigenvector centrality [26], probabilistic latent graph-based measure space [27], concave minimization and SVM [28], convergence properties of the power series of matrices [29], Laplacian score [30], $L_{2,0}$-norm equality constraints (LNEC) [31], adaptive structure learning [32], robust spectral learning of the spectrum information of the graph Laplacian [33], and $L_{2,1}$-norm minimization on processes of both label learning and feature learning [34].

2.4. Stability estimator

The estimator recasts the stability measure as a random variable estimation with explicitly embedded parameters[7]. After sampling distribution is identified, confidence intervals are estimated and hypothesis tests are performed. It allows for reliable comparison of stability across different procedures. Notably, *S* value above 0.75 represents excellent agreement and between 0.40 and 0.75 indicates intermediate to good agreement.

3. Results

Estimated stability is shown in Tables 2 and 3. The values with $S \geq 0.55$ when $m = 3$ are highlighted in red, indicating the algorithms are stable. Table 2 shows most algorithms

[6]https://github.com/NicoYuCN/matFR
[7]https://github.com/nogueirs/JMLR2018

achieve excellent stability. The S values of [12, 22, 23, 26, 30, 33, 34] are larger than 0.75 on both datasets. On contrast, [19] and [24] on BCDR-F03, and [19] and [28] on WDBC, are highly sensitive to data perturbations.

On gene expression datasets, Table 3 indicates few algorithms with good stability. On GSE10810, [14, 20, 23, 28, 31] are with $0.58 \leq S \leq 0.78$, and on GSE15852, [20, 22, 23, 26, 29–31] have values between 0.56 and 0.85. Notably, [13] and [23] achieve $S \geq 0.70$ on GSE10810 and [23] and [27] are with $S \geq 0.80$ on GSE15852, with good agreement when feature subsets change.

In summary, 19, 20, 5, and 7 algorithms show good stability ($S \geq 0.55$ when $m = 3$) on BCDR-F03, WDBC, GSE10810 and GSE15852, respectively. Further observation reveals that [13, 22, 26, 29, 30] are robust on 3 datasets, and 3 feature rankers (GFS [20], PWFP [23] and LNEC [31]) are consistently stable on all the datasets.

4. Discussion

The stability of 23 FR algorithms is investigated using an advanced estimator on 4 BC datasets. Stability is central in massive applications. Since higher stability increases user confidence in complex data analysis, a user prefers an algorithm that yields stable feature ranks even though perturbations exist in training data [2, 5].

Three algorithms show good stability consistently on the datasets. Initially, 19 methods cannot handle GSE15852 effectively. Besides high-performance hardware, these algorithms need massive time to process gene datasets (GSE10810 and GSE15850). Secondly, most of the remaining algorithms achieve stable feature ranks on BCDR-F03 and WDBC, while substantially fewer algorithms are stable on GSE10810 and GSE15850. It is found that more than 18 samples describe a feature on BCDR-F03 and on WDBC, while on gene datasets, samples are far from sufficient to express a feature. This might suggest that data sufficiency is vital to the construction of measure spaces before accurate estimation of feature importance [6]. Close observation finds that GFS [20], PWFP [23] and LNEC [31] are consistently robust on all the four datasets.

There are several reasons that the three algorithms generate stable feature ranks efficiently. To GFS [20], it first finds a subset of features jointly to filter out redundant and unrelated variables. In a reduced feature space, data subsets are optimized in regularized discriminant analysis. Second, in the data space spanned by selected features, the distances between samples in different classes will be expanded as large as possible, and that between samples in the same class will be reduced as small as possible. Finally, the feature ranking problem is formed as a multiple kernel learning problem in each iteration, and thus, time cost decreases and the computing is efficient. To PWFP [23], instead of looking at the samples in groups, it evaluate feature efficiency based on pairwise fashion, *i.e.*, a pair of samples is considered at a time. In particular, the features bringing the sample pairs closer or putting the pairs far away is selected as a good choice for feature ranking. And to LNEC [31], feature selection and data partition are considered in a joint manner that increases interdependence among data samples, cluster labels and selected features. Using $L_{2,0}$-norm equality constraints of dependence guided terms, learned cluster labels are used to fill the information gap between data samples and selected features, and alternating direction method of multipliers is designed to solve the constrained minimization problem iteratively and efficiently.

Table 2. Stability of algorithms on medical image datasets

		top-3	top-4	top-5	top-6	top-7	top-8	top-9
BCDR-F03	[12]	1.00	0.85	0.77	0.76	0.77	0.78	0.74
	[13]	0.85	0.81	0.80	0.69	0.66	0.68	0.64
	[14]	0.87	0.82	0.80	0.68	0.66	0.67	0.63
	[15]	0.80	0.79	0.89	0.79	0.73	0.71	0.71
	[16]	0.73	0.76	0.84	0.80	0.82	0.93	0.84
	[17]	0.46	0.39	0.37	0.39	0.39	0.40	0.39
	[18]	0.74	0.59	0.54	0.50	0.46	0.43	0.47
	[19]	0.06	0.08	0.07	0.10	0.11	0.14	0.14
	[20]	0.67	0.74	0.77	0.83	0.84	0.80	0.86
	[21]	1.00	0.82	0.77	0.73	0.63	0.59	0.59
	[22]	0.81	0.81	1.00	1.00	1.00	1.00	1.00
	[23]	1.00	0.89	0.90	0.91	0.92	0.92	0.99
	[24]	0.24	0.27	0.35	0.36	0.35	0.36	0.38
	[25]	0.72	0.89	0.85	0.91	1.00	0.92	0.91
	[26]	0.83	1.00	0.88	0.75	0.70	0.73	0.73
	[27]	0.52	0.62	0.63	0.62	0.62	0.54	0.48
	[28]	0.73	0.82	0.79	0.84	0.75	0.68	0.71
	[29]	0.79	1.00	0.87	0.78	0.78	0.79	0.83
	[30]	1.00	1.00	0.93	0.89	1.00	0.99	0.92
	[31]	0.77	0.80	0.94	0.85	0.99	0.88	0.88
	[32]	0.79	0.80	0.82	0.78	0.82	0.79	0.69
	[33]	0.88	0.90	0.93	0.94	0.87	0.91	0.89
	[34]	1.00	0.85	1.00	1.00	1.00	0.96	0.89
WDBC	[12]	0.81	1.00	1.00	0.95	0.92	0.89	0.92
	[13]	0.74	0.76	0.78	0.86	0.88	0.86	0.85
	[14]	0.56	0.71	0.84	1.00	0.88	0.95	0.94
	[15]	0.75	0.87	1.00	0.94	0.90	0.89	0.97
	[16]	0.89	1.00	0.88	0.86	0.80	0.86	0.91
	[17]	0.59	0.59	0.56	0.53	0.53	0.51	0.50
	[18]	0.67	0.63	0.57	0.54	0.50	0.47	0.47
	[19]	0.08	0.12	0.14	0.17	0.21	0.22	0.25
	[20]	0.81	0.98	0.97	0.91	0.92	0.95	0.97
	[21]	0.61	0.78	1.00	0.88	0.82	0.85	0.88
	[22]	1.00	0.97	1.00	0.89	1.00	0.94	1.00
	[23]	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	[24]	0.82	0.95	0.96	1.00	0.91	1.00	0.98
	[25]	0.86	0.86	0.90	0.83	0.80	0.78	0.80
	[26]	0.96	1.00	0.88	0.99	1.00	0.96	0.98
	[27]	0.90	0.97	0.91	0.85	0.87	0.95	0.93
	[28]	0.19	0.22	0.23	0.23	0.23	0.23	0.25
	[29]	0.98	1.00	0.88	0.98	1.00	0.96	0.97
	[30]	0.98	0.87	1.00	0.98	1.00	0.98	0.95
	[31]	1.00	1.00	0.90	0.90	0.96	1.00	1.00
	[32]	0.47	0.57	0.66	0.72	0.77	0.81	0.83
	[33]	0.98	0.87	1.00	1.00	1.00	0.98	0.92
	[34]	0.80	0.86	0.77	0.75	0.73	0.80	0.82

Table 3. Stability of algorithms on gene expression datasets

		top-3	top-4	top-5	top-6	top-7	top-8	top-9
GSE10810	[12]	0.44	0.55	0.54	0.52	0.52	0.51	0.50
	[13]	0.78	0.87	0.81	0.77	0.73	0.71	0.72
	[14]	0.44	0.52	0.52	0.49	0.49	0.50	0.50
	[15]	0.52	0.51	0.46	0.44	0.43	0.44	0.47
	[16]	0.21	0.26	0.25	0.25	0.25	0.25	0.25
	[17]	0.43	0.42	0.41	0.41	0.41	0.41	0.42
	[18]	0.15	0.20	0.20	0.20	0.21	0.23	0.24
	[19]	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	[20]	0.58	0.50	0.49	0.48	0.48	0.47	0.49
	[21]	0.52	0.51	0.46	0.44	0.43	0.44	0.47
	[22]	0.31	0.40	0.39	0.39	0.37	0.39	0.39
	[23]	0.71	0.77	0.82	0.85	0.86	0.84	0.84
	[24]	0.11	0.12	0.13	0.16	0.18	0.21	0.22
	[25]	0.00	0.00	0.01	0.01	0.01	0.01	0.01
	[26]	0.46	0.51	0.59	0.66	0.73	0.78	0.80
	[27]	0.00	0.00	0.00	0.00	0.00	0.00	0.01
	[28]	0.63	0.56	0.52	0.50	0.46	0.45	0.44
	[29]	0.46	0.50	0.59	0.66	0.71	0.78	0.80
	[30]	0.25	0.28	0.30	0.30	0.33	0.35	0.36
	[31]	0.73	0.64	0.64	0.65	0.67	0.66	0.66
	[32]	0.05	0.05	0.07	0.09	0.09	0.09	0.09
	[33]	0.51	0.52	0.50	0.48	0.51	0.54	0.57
	[34]	0.17	0.18	0.18	0.21	0.22	0.23	0.24
GSE15852	[12]	0.40	0.49	0.51	0.52	0.59	0.61	0.61
	[13]	0.51	0.51	0.56	0.60	0.61	0.60	0.62
	[14]	0.47	0.51	0.57	0.59	0.58	0.57	0.57
	[15]	0.28	0.33	0.38	0.40	0.40	0.42	0.44
	[16]	0.48	0.52	0.59	0.65	0.70	0.73	0.76
	[17]	0.31	0.41	0.46	0.52	0.55	0.58	0.60
	[18]	0.16	0.21	0.21	0.20	0.20	0.21	0.21
	[19]	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	[20]	0.57	0.56	0.57	0.61	0.66	0.71	0.75
	[21]	0.21	0.24	0.26	0.30	0.32	0.35	0.38
	[22]	0.59	0.62	0.62	0.67	0.73	0.78	0.81
	[23]	0.84	0.88	0.88	0.86	0.88	0.88	0.89
	[24]	0.24	0.26	0.24	0.26	0.27	0.28	0.30
	[25]	0.53	0.46	0.44	0.42	0.43	0.42	0.43
	[26]	0.87	0.97	0.89	0.88	0.85	0.87	0.86
	[27]	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	[28]	0.25	0.28	0.27	0.29	0.31	0.31	0.32
	[29]	0.85	0.96	0.89	0.89	0.86	0.86	0.88
	[30]	0.62	0.65	0.66	0.74	0.80	0.88	0.87
	[31]	0.56	0.54	0.56	0.57	0.59	0.60	0.61
	[32]	0.05	0.05	0.06	0.07	0.06	0.07	0.07
	[33]	0.54	0.58	0.56	0.55	0.55	0.55	0.55
	[34]	0.16	0.18	0.22	0.24	0.26	0.27	0.29

Several limitations exist in this study. First, how the change of training sizes impacts on the stability estimation is interesting. However, due to insufficient samples of gene datasets, the training sample sizes of each dataset is fixed. Second, using one estimator to assess stability seems not convincing, even if the estimator possesses the properties of a good stability measure. One desirable approach is to use more stability estimators for a comprehensive evaluation [2]. In addition, besides handcrafted features, deeply learnt features will be in our future work to improve network robustness and generalization capacity [35]. In this kind of settings, deep networks perform as feature extractors [36, 37]. When using dropout [38] to determine which nodes are activated or selected, a subset feature selection method is formed, and its stability can be measured.

5. Conclusions

This study investigates the stability of 23 FR algorithms on four BC datasets using an advanced estimator, and three algorithms are identified as consistently exhibiting good stability on all the datasets. Stability is crucial for many decision-making applications. In our future work, experiments will be conducted by involving more algorithms, estimators and datasets to recognize stable algorithms for data analysis.

References

[1] Z. Zhang, X. Liang, W. Qin, S. Yu, and Y. Xie, "matFR: a MATLAB toolbox for feature ranking," Bioinformatics, vol. 36(19), pp. 4968–4969, 2020.

[2] S. Nogueira, K. Sechidis, and G. Brown, "On the stability of feature selection algorithms," Journal of Machine Learning Research, vol. 18(1), pp. 6345–6398, 2017.

[3] N. López, M. García-Ordás, F. Vitelli-Storelli, P. Fernández-Navarro, C. Palazuelos, and R. Alaiz-Rodríguez, "Evaluation of feature selection techniques for breast cancer risk prediction," International Journal of Environmental Research and Public Health. vol. 18(20), pp. 10670, 2021.

[4] N. Cueto-López, M. García-Ordás, V. Dávila-Batista, V. Moreno, N. Aragonés, and R. Alaiz-Rodríguez, "A comparative study on feature selection for a risk prediction model for colorectal cancer," Computer methods and programs in biomedicine. vol. 177, pp. 219–229, 2019.

[5] A. Kalousis, J. Prados, and M. Hilario, "Stability of feature selection algorithms: a study on high-dimensional spaces," Knowledge and information systems. vol. 12(1), pp. 95–116, 2007.

[6] D. Dernoncourt, B. Hanczar, and J. Zucker. "Analysis of feature selection stability on high dimension and small sample data," Computational statistics & data analysis. vol. 71, pp. 681–693, 2014.

[7] J. Arevalo, F. González, R. Ramos-Pollán, J. Oliveira, and M. Lopez. "Representation learning for mammography mass lesion classification with convolutional neural networks," Computer methods and programs in biomedicine. vol. 127, pp. 248–257, 2016.

[8] S. Yu, L. Liu, Z. Wang, G. Dai, and Y. Xie. "Transferring deep neural networks for the differentiation of mammographic breast lesions," Science China Technological Sciences. vol. 62(3), pp. 441–447, 2019.

[9] W. Street, W. Wolberg, R. Ramos-Pollán, and O. Mangasarian. "Nuclear feature extraction for breast tumor diagnosis," Biomedical image processing and biomedical visualization. pp. 861–870, 1993.

[10] V. Pedraza, J. Gomez-Capilla, G. Escaramis, C. Gomez, P. Torné, J. Rivera, A. Gil, P. Araque, N. Olea, and X. Estivill, "Gene expression signatures in breast cancer distinguish phenotype characteristics, histologic subtypes, and tumor invasiveness," Cancer: Interdisciplinary International Journal of the American Cancer Society, vol. 116(2), pp. 486–496, 2010.

[11] I. Ni, Z. Zakaria, R. Muhammad, N. Abdullah, N. Ibrahim, N. Emran, N. Abdullah, and S. Hussain, "Gene expression patterns distinguish breast carcinomas from normal breast tissues: the Malaysian context," Pathology-Research and Practice, vol. 206(4), pp. 223–228, 2010.

[12] N. Cressie and H. Whitford, "How to use the two sample t-test," Biometrical Journal, vol. 28(2), pp. 131–148, 1986.

[13] T. Cover and J. Thomas, "Entropy, relative entropy and mutual information," Elements of information theory, vol. 2(1), pp. 12–13, 1991.
[14] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," IEEE transactions on communication technology, vol. 15(1), pp. 52–60, 1967.
[15] F. Hsieh and B. Turnbull, "Nonparametric and semiparametric estimation of the receiver operating characteristic curve," The annals of statistics, vol. 24(1), pp. 25–40, 1996.
[16] N. Nachar, "The Mann-Whitney U: A test for assessing whether two independent samples come from the same distribution," Tutorials in quantitative Methods for Psychology, vol. 4(1), pp. 13–20, 2008.
[17] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," Machine learning, vol. 53(1), pp. 23–69, 2003.
[18] R. Tibshirani, "Regression shrinkage and selection via the lasso," Journal of the Royal Statistical Society: Series B (Methodological), vol. 58(1), pp. 267–288, 2003.
[19] G. Roffo, "Feature selection library (MATLAB toolbox)," arXiv preprint, arXiv:1607.01327, 2016.
[20] Q. Gu, H. Li, and J. Han, "Generalized fisher score for feature selection," arXiv preprint, arXiv:1202.3725, 2012.
[21] J. Uitdehaag and G. Zaman, "A theoretical entropy score as a single value to express inhibitor selectivity," BMC bioinformatics, vol. 12(1), pp. 1–11, 2003.
[22] P. McKight and J. Najab, "Kruskal-wallis test," The corsini encyclopedia of psychology, pp. 1–1, 2010.
[23] S. Happy, R. Mohanty, and A. Routray, "An effective feature selection method based on pair-wise feature proximity for high dimensional low sample size data," European signal processing conference, pp. 1574–1578, 2017.
[24] W. Hu, K. Choi, Y. Gu, and S. Wang, "Minimum–maximum local structure information for feature selection," Pattern Recognition Letters, vol. 34(5), pp. 527–535, 2017.
[25] H. Zeng and Y. Cheung, "Feature selection and kernel learning for local learning-based clustering," IEEE transactions on pattern analysis and machine intelligence, vol. 33(8), pp. 1532–1547, 2010.
[26] G. Roffo and S. Melzi, "Features selection via eigenvector centrality," Proceedings of new frontiers in mining complex patterns, 2016.
[27] G. Roffo, S. Melzi, U. Castellani, and A. Vinciarelli. "Infinite latent feature selection: A probabilistic latent graph-based ranking approach," IEEE international conference on computer vision, pp. 1398–1406, 2017.
[28] P. Bradley and O. Mangasarian, "Feature selection via concave minimization and support vector machines," IEEE international conference on machine learning, vol. 98, pp. 82–90, 1998.
[29] G. Roffo, S. Melzi, and M. Cristani, "Infinite feature selection," IEEE International Conference on Computer Vision, pp. 4202–4210, 2015.
[30] X. He, D. Cai, and P. Niyogi, "Laplacian score for feature selection," Advances in neural information processing systems, pp. 18, 2005.
[31] J. Guo and W. Zhu, "Dependence guided unsupervised feature selection," AAAI Conference on Artificial Intelligence, vol. 32(1), 2018.
[32] L. Du and Y. Shen, "Unsupervised feature selection with adaptive structure learning," ACM SIGKDD international conference on knowledge discovery and data mining, pp. 209–218, 2015.
[33] L. Shi, L. Du, and Y. Shen, "Robust spectral learning for unsupervised feature selection," IEEE International Conference on Data Mining, pp. 977–982, 2014.
[34] M. Qian and C. Zhai, "Robust unsupervised feature selection," International joint conference on artificial intelligence, 2013.
[35] S. Yu, M. Chen, E. Zhang, J. Wu, H. Yu, Z. Yang, L. Ma, X. Gu, and W. Lu, "Robustness study of noisy annotation in deep learning based medical image segmentation," Physics in Medicine & Biology, vol. 65(17), pp. 175007, 2020.
[36] T. Xiao, L. Liu, K. Li, W. Qin, S. Yu, and Z. Li. "Comparison of transferred deep neural networks in ultrasonic breast masses discrimination," BioMed research international, 2018.
[37] L. Zou, S. Yu, T. Meng, Z. Zhang, X. Liang, and Y. Xie, "A technical review of convolutional neural network-based mammographic breast cancer diagnosis," Computational and mathematical methods in medicine, 2019.
[38] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," Journal of machine learning research, vol. 15(1), pp. 1929–1958, 2014.