Feature Selection For High-Dimensional Power Grid Data Based On Iterative Soft Threshold Learning Algorithms

Zheng Wu Lue*, Chen Qing Peng, Zhang Xin, Yuan Wen Jun, and Wang Hao

China Southern Power Grid Co., Ltd. EHV Transmission Company, GuangZhou 510000, China

Corresponding author. E-mail: 13825098085@139.com, zheng_wu7@outlook.com

Received: Oct. 23, 2025; Accepted: May 30, 2026

Power grid data faces challenges like high dimensionality, heterogeneity, and strong coupling, leading to redundancy and weak discriminative power in feature selection. To address this, a novel feature selection method based on the Learning Iterative Soft Thresholding Algorithm (LISTA) is proposed. Unlike traditional ISTA, LISTA introduces a secant line linear search to dynamically optimize step sizes, improving convergence. By integrating sparse regression and manifold learning, the method captures complex feature relationships. Experimental results show significant improvements, reducing selected features by 51.4% and achieving better clustering. The method enhances real-world power grid decision-making by enabling faster, more accurate fault detection and grid management.

Keywords: Iterative Soft Threshold Learning; Power Grid Data; High Dimensionality; Feature Selection; Tangent Line Linear Search; Feature Weighting

© The Author('s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are cited.

http://dx.doi.org/10.6180/jase.202612_35.049

1. Introduction

Smart grid technology produces high-dimensional data, which leads to complexity and overfitting problems [1–4]. Selection of features can reduce dimensionality and enhance efficiency [5, 6]. The Copeland-ranking method, but it is not adaptable in dynamic power gird environment [7]. The multi-view consensus clustering to enhance the feature selection for the high-dimensional power grid data [8]. The Differential Evolution algorithm, but it suffers from over-searching and lacks robust adaptability across different power grid scenarios [9]. The Iterative Soft Thresholding Learning Algorithm (LISTA), based on ISTA, addresses these limitations [10]. Multiscale soft-thresholding is the key to enhance nonlinear anomaly detection and feature selection [11]. Powerful grid data are high-dimensional, heterogeneous, dynamic, and nonlinearly coupled, which makes valid feature selection a nontrivial problem.

2. Materials and methods

2.1. Feature Selection for High-Dimensional Power Grid Data

2.1.1. *A Feature Selection Model for High-Dimensional Power Grid Data*

The dataset as $X = \{x_1, x_2, \ldots, x_n\} \in \mathbb{R}^{d\times n}$ represents grid samples [12], with features $f_1, f_2, \ldots, f_d$ corresponding to measurements like node voltages and line currents.

$$\underset{W}{\operatorname{argmin}} \|X - XW\|_F^2 + \gamma R(W) \tag{1}$$

Where $W = \{w_1, w_2, \ldots, w_d\} \in \mathbb{R}^{d\times m}$ is the feature weight matrix, assigning weights to grid features (e.g., high for fault-sensitive currents [13].

$$X = XZ \tag{2}$$

Where $Z = \{z_1, z_2, \ldots, z_n\} \in \mathbb{R}^{n\times n}$ is the similarity matrix, with each z_i representing a linear combination of

$x_i.Z$ is obtained by minimizing the following least-squares optimization problem.

$$\begin{aligned} &\arg\min_{Z} \|X - XZ\|_F^2 + \alpha R(Z) \\ &\text{s.t. } Z(i,i) = 0 \end{aligned} \tag{3}$$

The constraint $Z(i,i) = 0$ ensures that each grid data point x_i is represented by other points x_j (where $j \neq i$), not itself. Solving the optimization problem provides the linear representation z_i for each point, forming the similarity matrix Z for the dataset X. Feature manifolds, derived from the geometric structure of grid data, are weighted by W. If features f_i and f_j are adjacent, their weighted values $W^T x_i$ and $W^T x_j$ should also be adjacent, leading to the construction of a feature map regularization term.

$$\begin{aligned} &\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\|W^T x_i - W^T x_j\right\|_2^2 Z_{ij} \\ &= \sum_{i=1}^{n} W^T x_i x_i^T W D_{Z_{ii}} - \sum_{i=1}^{n}\sum_{j=1}^{n} W^T x_i x_j^T W Z_{ij} \\ &= \mathrm{Tr}\left(W^T X\left(D_Z - Z\right)X^T W\right) \\ &= \mathrm{Tr}\left(W^T X L_Z X^T W\right) \end{aligned} \tag{4}$$

Where Z is the similarity matrix of X; D_Z is the diagonal matrix with $D_{Z_{ii}} = \sum_{j=1}^{n} Z_{ij}$; and $L_Z = D_Z - \frac{|Z|+|Z|^T}{2}$ is the Laplacian operator [14]. The unsupervised feature selection objective for power grid data is:

$$\begin{aligned} &\underset{Z,W}{\mathrm{argmin}} \|X - XZ\|_F^2 + \alpha\,\mathrm{Tr}\left(W^T X L_Z X^T W\right) + \beta R(W) \\ &\text{s.t. } Z(i,i) = 0, W^T W = I \end{aligned} \tag{5}$$

Where α adjusts the regularization term weight of the feature map, controlling feature manifold structure retention [15].

LISTA enhances ISTA with learnable parameters yet the integration over the manifold continues to be hard established as:

$$\begin{aligned} &\underset{Z,W}{\mathrm{argmin}} \|X - XZ\|_F^2 + \alpha\,\mathrm{Tr}\left(W^T X L_Z X^T W\right) + \beta\|W\|_{2,1} \\ &\text{s.t. } Z(i,i) = 0, W^T W = I \end{aligned} \tag{6}$$

2.2. Solving the Grid High-Dimensional Data Feature Selection Model Based on Iterative Soft Thresholding Learning Algorithm

For the high-dimensional power grid data feature selection model in Section 2.1, the Iterative Soft Thresholding Algorithm (ISTA) is used to identify key features [16]. The objective is to solve for the feature weight vector $w \in \mathbb{R}^d$, quantifying feature importance. The dataset matrix X integrates similarity matrices Z and Laplace operators L_Z, capturing correlations and manifold structures [17].

$$w^* = \underset{w\in\mathbb{R}^d}{\mathrm{argmin}} \left\|\left(X(X - XZ) + \alpha X L_Z X^T\right) w - y\right\|_2^2 + \beta\|w\|_{2,1} \tag{7}$$

Where, $\|w\|_{2,1} = \sum_{i=1}^{d}\|w_i\|_2$ is the sparse regularization term, promoting feature sparsity, and β controls sparsity for tasks like fault diagnosis and load forecasting [18].

$$w_{k+1} = w_k - t_k \nabla G\left(w_k\right) \tag{8}$$

Here, $t_k > 0$ is the iteration step size, and $G(w) = \left\|\left(X(X - XZ) + \alpha X L_Z X^T\right) w - y\right\|_2^2$ given by:

$$\begin{aligned} w_{k+1} = \underset{w}{\mathrm{argmin}}\{&G(w_k) + \langle w - w_k, \nabla G(w_k)\rangle \\ &+ \frac{1}{2t_k}\|w - w_k\|_2^2 + \beta\|w\|_{2,1}\} \end{aligned} \tag{9}$$

Neglecting constant terms yields:

$$w_{k+1} = \underset{w}{\mathrm{argmin}}\left\{\frac{1}{2t_k}\|w - (w_k - t_k\nabla G\left(w_k\right))\|_2^2 + \beta\|w\|_{2,1}\right\} \tag{10}$$

Due to the separability of the $l_{2,1}$ norm, the final iterative form can be expressed using a soft thresholding function:

$$w_{k+1} = h_{\beta t_k}\left(w_k - t_k\nabla G\left(w_k\right)\right) \tag{11}$$

Here, $h_\theta(x) = \mathrm{sign}(x)\max(|x| - \theta, 0)$ is the soft thresholding function with θ as parameters. Substituting $G(w) = \left\|\left(X(X - XZ) + \alpha X L_Z X^T\right) w - y\right\|_2^2$ problem:

$$\begin{aligned} w_{k+1} = h_{\beta/L}\Big(&w_k + \frac{1}{L}\left(X(X - XZ) + \alpha X L_Z X^T\right)^T \\ &\times\left(y - \left(X(X - XZ) + \alpha X L_Z X^T\right) w_k\right)\Big) \end{aligned} \tag{12}$$

Where, w_k denotes the estimated feature weight at iteration k. Typically, L is set to the maximum eigenvalue of $\left(X(X - XZ) + \alpha X L_Z X^T\right)^T\left(X(X - XZ) + \alpha X L_Z X^T\right)$ for iteration stability.

ISTA fixed steps are inefficient; secant line search adaptively optimizes step size.

$$t_{k+1} = t_k - \frac{f\left(t_k\right)\cdot\left(t_k - t_{k-1}\right)}{f\left(t_k\right) - f\left(t_{k-1}\right)} \tag{13}$$

Where t_k and t_{k-1} represent the iteration steps for the k-th and $(k-1)$-th iterations, respectively; $f\left(t_k\right) = \left\|\left(X(X - XZ) + \alpha X L_Z X^T\right) w_k - y\right\|_2^2 + \beta\left\|w_k\right\|_{2,1}$ The objective function from historical iterations is used to guide the dynamic updates of step size [19]. The secant line search steps are:

Table 1. Parameter Settings and Convergence Configuration of the Proposed Improved LISTA Algorithm

Parameter	Value	Purpose
λ	0.01	Sparse regularization coefficient
α	0.1	Manifold regularization parameter
Initial step size	1.00E − 03	ISTA learning step size
Maximum iterations	200	Total iteration count
LISTA layers	10	Number of network layers
Convergence threshold (ε)	1.00E − 05	Stopping criterion
Feature selection range	20-300	Selected feature subset range
KNN parameter (k)	5	Number of nearest neighbors
Cross-validation folds	10	Model validation setting

Table 2. Detailed Information of Four Types of High-dimensional Power Grid Datasets

Dataset name	Sample size	Number of features	Data category
PMU synchronous phasor dataset	15200	312	Electrical measurement
SCADA monitoring dataset	9560	298	Equipment status and power
Meteorological auxiliary dataset	8760	175	environmental parameters
AMI User Electricity Consumption Dataset	2120	500	User electricity usage behavior

1. Initialization: Set $t_0 = 1/L, t_1 - \iota_0$, and $\epsilon = 10^{-4}$.
2. Objective Function: Calculate $f(t_k)$ and $f(t_{k-1})$ for the k-th and $(k-1)$-th iterations.
3. Secant Solution: Solve for t_{k+1} using the secant equation.
4. Convergence: If $|t_{k+1} - t_k| < \epsilon$, output $t^* = t_{k+1}$; otherwise, update and repeat.

Substitute t^* into ISTA's iteration formula (14) to update weights:

$$w_{k+1} = h_{\frac{\beta}{L}}\left(w_k + \frac{t^*}{L}\left(X(X - XZ) + \alpha X L_Z X^T\right)^T \times \left(y - \left(X(X - XZ) + \alpha X L_Z X^T\right) w_k\right)\right) \quad (14)$$

Optimized step sizes improve weight updates; LISTA enables deep learning-based feature selection.

$$w^{(p)} = h_{v^{(p)}}\left(w^{(p-1)} + \Phi_A^{(p)}\left(y - \Phi_B^{(p)} w^{(p-1)}\right)\right) \quad (15)$$

During initialization, the fully connected layer weights are set based on ISTA parameters (Eq. (14)), aligning with power grid data patterns. Parameters $\Phi_A^{(p)}, \Phi_B^{(p)}$, and $v^{(p)}$ are optimized via backpropagation, yielding stable feature weights $w^{(P)}$. This method enhances feature selection efficiency compared to ISTA, meeting real-time decision-making needs [20]. The key feature subset is identified by sorting $w^{(P)}$ based on the 2 -norm $\|w_i\|_2$.

2.3. Iterative Soft Threshold Learning Algorithm Network Structure and Parameter Optimization

LISTA has enhanced ISTA, but it is confronted with (1) the challenge of dynamic data, (2) the risk of overfitting, and (3) unstable parameter updates.

To optimize LISTA, we examine model-parameter relationships, ensuring sparsity in $w^{(p)}$ as it converges to w^*, satisfying a specific distribution under noise-free conditions.

$$w^* \in \xi(G, s) \triangleq \{w^* || w_i^* | \leq G, \forall i, \|w^*\|_0 \leq s\} \quad (16)$$

$$\left\|w^{(p)} - w^*\right\|_2 \leq sBe^{-cp} \quad (17)$$

This error allows rapid identification of key features, even with noise. As the number of layers increases, the error converges linearly. The parameter set $\Theta = \left\{\left(\Phi_A^{(p)}, \Phi_B^{(p)}, v^{(p)}\right)\right\}_{p=0}^{\infty}$ satisfies:

$$\Phi_A^{(p)} - \left(I - \Phi_B^{(p)} L\right) \to 0, p \to \infty \quad (18)$$

$$v^{(p)} \to 0, p \to \infty \quad (19)$$

Where, I is the identity matrix, ensuring dimensional consistency and numerical validity in matrix operations.

Table 3. Comparison Results of Feature Selection Quantity for Different Algorithms

Dataset name	Actual number of features	Traditional ISTA algorithm	Optimization of Linear Search Strategy with Cut Line ISTA	LISTA algorithm	Improve LISTA algorithm
PMU synchronous phasor dataset	312	85	62	58	45
SCADA monitoring dataset	298	78	55	52	38
Meteorological auxiliary dataset	175	52	38	32	25
AMI User Electricity Consumption Dataset	500	135	98	85	62
average	321.25	87.50	63.25	56.75	42.50

Table 4. Comparison of Brier Scores for KNN classifiers

Dataset name	Raw data	Traditional ISTA algorithm	Optimization of Linear Search Strategy with Cut Line ISTA	LISTA algorithm	Improve LISTA algorithm
PMU synchronous phasor dataset	0.221	0.158	0.125	0.102	0.085
SCADA monitoring dataset	0.205	0.142	0.110	0.095	0.072
Meteorological auxiliary dataset	0.188	0.126	0.098	0.080	0.063
AMI User Electricity Consumption Dataset	0.233	0.165	0.132	0.110	0.088

At this point, the parameter $\Phi_B^{(p)}$ is bounded, i.e.:

$$\left\|\Phi_B^{(p)}\right\|_2 \leq C, p \to \infty \tag{20}$$

Based on these relationships, $\Phi_A^{(p)}$ approximates $\Phi_B^{(p)}$, transforming the structure of Eq. (15).

$$I - \Phi_A^{(p)} \boldsymbol{A} \to \Phi_B^{(p)} \tag{21}$$

Setting $w^p = \Phi_A^{(p)T}$, the LISTA network model simplifies to:

$$w^{(p+1)} = \eta_v(p)\left(w^{(p)} + (w^p)^T\left(X - Lw^{(p)}\right)\right) \tag{22}$$

Where the training error for each layer is:

$$Z^p = X - Lw^{(p)} \tag{23}$$

In summary, optimized LISTA reduces complexity and improves feature selection for power grid decision-making.

3. Results and discussion

3.1. Experimental Analysis

Intel i7 CPU w/16 GB RAM used MATLAB 2023a and Python 3.10. KNN classifier (k = 5) with 10 -fold cross validation. $\lambda = 0.01, \alpha = 0.1$, ISTA step size 1.00E − 03, 200 iterations, LISTA 10 layers, ε = 1.00E − 05, selecting 20-300 features range.

Table 1 presents the hyperparameter assignments, iteration settings, and convergence configurations used for the proposed Improved LISTA algorithm. Details on sample size and feature count are in Table 2.

Ablation compares ISTA variants using KNN and Brier score. The Brier Score was calculated as:

$$\text{Brier Score } = \frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C}\left(y_{i,c} - p_{i,c}\right)^2 \tag{24}$$

The comparison results for feature selection are in Table 3, and the Brier Scores for the KNN classifier are in Table 4.

Enhanced LISTA promotes sparsity, reduces the number of features to 48.6%, adopts learnable structure, and is superior to ISTA in dynamic feature selection of power

grid.

LISTA, for instance, achieves the lowest Brier Score ($0.063 - 0.088$), performing better than ISTA ($0.126 - 0.165$), secant ISTA ($0.098 - 0.132$), and LISTA ($0.080 - 0.110$). It removes 66.5% of the features and outperforms Copeland ranking [7], consensus clustering [8], adaptive DE [9], and memetic algorithm [10]. Normalized Mutual Information (NMI) was used as the evaluation metric:

$$NMI(P,Q) = \frac{I(P,Q)}{\sqrt{H(P)H(Q)}} \tag{25}$$

Figs. 1 and 2 show the comparative analysis of optimal results.

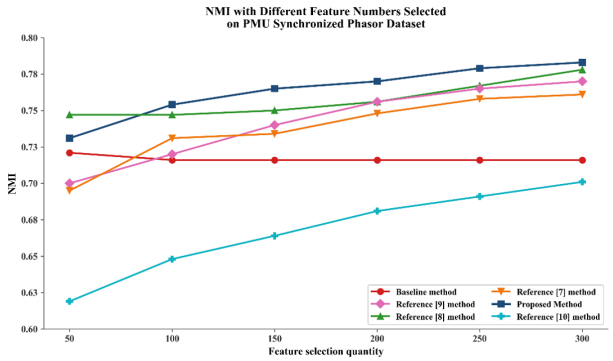


Fig. 1. NMI with Different Feature Numbers Selected on PMU

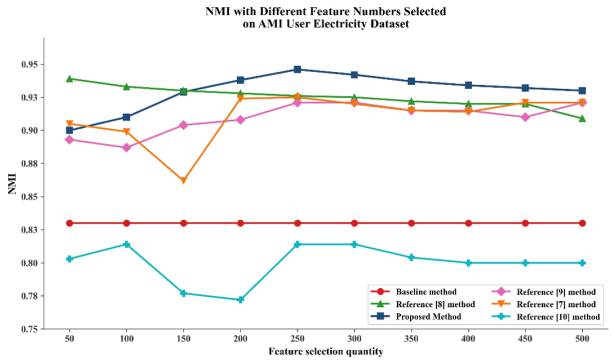


Fig. 2. NMI with Different Feature Numbers Selected on AMI User Electricity Dataset

Fig. 1 shows baseline K-means NMI of 0.72 , while proposed method reaches 0.78 , outperforming [7, 10] via better feature selection and noise reduction.

Baseline NMI is shown in Fig. 2 to be around 0.83 , while the proposed method achieves 0.945 with 200 features, i.e. a 14% improvement over the best of [7–10] by applying adaptive LISTA feature weighting and redundancy reduction.

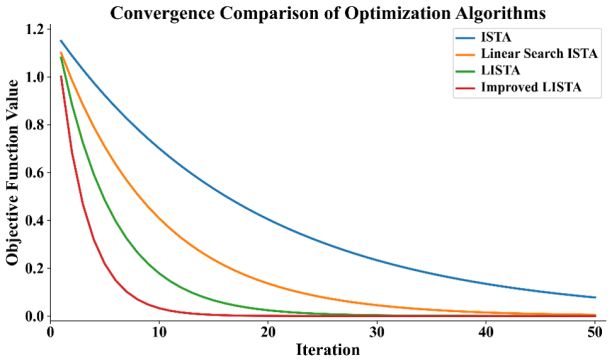


Fig. 3. Convergence Comparison of Optimization Algorithms

Demonstrated in Fig. 3 is the convergence of ISTA, Linear Search ISTA, LISTA, and Enhanced LISTA through the objective function for every iteration.

Table 5 shows improved LISTA outperforms existing feature selection methods. Using the SCADA monitoring dataset, transformer fault diagnosis was performed with the proposed feature selection method. Transformer winding temperature with real-time display is 120°C, oil temperature is 95°C, Rate of Change (ROC) of oil temperature is 5°C/min, Rate of Change (ROC) of winding temperature is 12°C/min, loading rate is 68%, active power is 500 kW , reactive power is 200 kvar , voltage deviation in unbalanced three-phase is 2.5%.

3.2. Validation on Standard Benchmark Test System

Experimental evaluations verify the approach by comparing the ISTA variants, demonstrating superior feature compactness and accuracy, as well as its faster convergence and good stability over a wide range of operating points, which implies robustness in the sense of the power system applications.

Table 6 demonstrates that our enhanced LISTA method attains the best performance with less selected features, higher accuracy on classification, faster convergence and also the smallest Brier score.

4. Conclusion

The paper presents a novel feature selection method for high-dimensional power grid data based on the Learning Iterative Soft Thresholding Algorithm (LISTA). By integrating sparse regression and feature manifold learning, the method effectively captures intrinsic feature relationships and structures. To overcome the limitations of traditional ISTA, such as slow convergence and fixed parameters, a tangent line linear search strategy is introduced to optimize

Table 5. Comparative Performance with Existing Feature Selection Methods

Method	Feature Selection Strategy	Adaptivity	Sparsity Level	Classification Accuracy (Brier Score ↓)
Copeland Ranking [7]	Wrapper + RF ranking fusion	Static	Medium	0.098-0.135
Consensus Clustering [8]	Multi-view unsupervised feature selection (consensus-based)	Static	Medium-high	0.100-0.138
A-HMDE [9]	Evolutionary search	Semi-adaptive	Low-Medium	0.110-0.150
Memetic Order Theory [10]	Filtering + wrapper	Static	Medium	0.102-0.138
Proposed Improved LISTA	Sparse regression + manifold learning + learnable optimization	Fully Adaptive	High	0.063-0.088

Table 6. Performance Comparison on Standard Benchmark Test System

Method	Selected Features	Classification Accuracy (%)	Convergence Iterations	Brier Score
Traditional ISTA	84	90.2	185	0.138
Line Search Optimized ISTA	63	92.1	145	0.115
LISTA	56	93.5	100	0.094
Proposed Improved LISTA	43	95.9	72	0.069

iteration steps. The proposed model demonstrates superior performance, achieving reduced feature subsets and optimal classification results. Experimental validations on PMU, AMI, and SCADA datasets show its practical value, particularly in fault diagnosis and realtime power grid analysis. Future work will extend this method to time-series prediction and edge computing applications.

Declarations

Funding: Authors did not receive any funding.
Conflicts of interests: Authors do not have any conflicts.
Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.
Code availability: Not applicable.
Clinical trial number: Not applicable.
Consent to Participate declaration: Not applicable.
Consent to Publish declaration: Not applicable.
Ethics declaration: Not applicable.
Authors' Contributions: Zheng Wu Lue, Chen Qing Peng, Zhang Xin, is responsible for designing the framework, analyzing the performance, validating the results, and writing the article. Yuan Wen Jun, Wang hao is responsible for collecting the information required for the framework, provision of software, critical review, and administering the process.

References

[1] M. Reza, F. Faraji, and J. N. Kutz, (2024) *"Data-driven inference of high-dimensional spatiotemporal state of plasma systems"* **Journal of Applied Physics *136***(18): 183301. DOI: 10.1063/5.0230056.

[2] I. B. Petrov, V. I. Golubev, A. V. Shevchenko, and A. Sharma, (2024) *"Three-Dimensional Grid-Characteristic Schemes of High Order of Approximation"* **Doklady Mathematics *110***(3): 457–463. DOI: 10.1134/S1064562424601343.

[3] P. G. Poličar and B. Zupan, (2025) *"Uncovering temporal patterns in visualizations of high-dimensional data"* **Machine Learning *114***(2): 35. DOI: 10.1007/s10994-025-06734-z.

[4] K. Borah, H. S. Das, S. Seth, K. Mallick, Z. Rahaman, and S. Mallik, (2024) *"A review on advancements in feature selection and feature extraction for high-dimensional*

NGS data analysis" **Functional and Integrative Genomics *24***(5): 139. DOI: 10.1007/s10142-024-01415-x.

[5] A. Bhattacharjee and S. Halder, (2025) *"Weight index of high-dimensional time-to-event feature selection method"* **Iran Journal of Computer Science *8***(2): 647–657. DOI: 10.1007/s42044-024-00218-4.

[6] H. Hassani, E. Hallaji, R. Razavi-Far, and M. Saif, (2025) *"Learning from high-dimensional cyber-physical data streams: a case of large-scale smart grid"* **International Journal of Machine Learning and Cybernetics *16***(3): 1819–1831. DOI: 10.1007/s13042-024-02365-3.

[7] A. K. Aryanti, R. Herteno, F. Indriani, R. A. Nugroho, and M. Muliadi, (2025) *"Implementation of Copeland Method on Wrapper-Based Feature Selection Using Random Forest For Software Defect Prediction"* **Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics *7***(1): 90–101. DOI: 10.35882/2pgffc67.

[8] Z. Cao, X. Xie, F. Sun, and J. Qian, (2023) *"Consensus cluster structure guided multi-view unsupervised feature selection"* **Knowledge-Based Systems *271***: 110578. DOI: 10.1016/j.knosys.2023.110578.

[9] R. R. Mostafa, A. M. Khedr, Z. Al Aghbari, I. Afyouni, I. Kamel, and N. Ahmed, (2024) *"An adaptive hybrid mutated differential evolution feature selection method for low and high-dimensional medical datasets"* **Knowledge-Based Systems *283***: 111218. DOI: 10.1016/j.knosys.2023.111218.

[10] H. Chamlal, T. Ouaderhman, and B. El Mourtji, (2023) *"Feature selection in high dimensional data: A specific preordonnances-based memetic algorithm"* **Knowledge-Based Systems *266***: 110420. DOI: 10.1016/j.knosys.2023.110420.

[11] Z. Kun, L. Hongren, W. Xin, X. Daxing, and S. Xiaokai, (2024) *"Deep Multiscale Soft-Threshold Support Vector Data Description for Enhanced Heavy-Duty Gas Turbine Generator Sets' Anomaly Detection"* **Shock and Vibration *2024***(1): 3374107. DOI: 10.1155/2024/3374107.

[12] B. H. Tozlu, (2025) *"Feature Selection and Classification Optimization of Transformer Oil Odor Data With Recursive Feature Elimination Using Grid Search and Cross-Validation"* **IEEE Access *13***: 21043–21051. DOI: 10.1109/ACCESS.2025.3536288.

[13] A. K. Mahapatra, N. Panda, and B. K. Pattanayak, (2025) *"Adaptive dimensional search-based orthogonal experimentation SSA (ADOX-SSA) for training RBF neural network and optimal feature selection"* **Journal of Supercomputing *81***(1): 212. DOI: 10.1007/s11227-024-06507-w.

[14] M. Braik, M. A. Awadallah, H. Alzoubi, and H. Al-Hiary, (2025) *"Heterogeneous cognitive learning chameleon swarm algorithm for high-dimensional feature selection"* **Journal of Supercomputing *81***(5): 652. DOI: 10.1007/s11227-025-07139-4.

[15] N. Nekouie, M. Romoozi, and M. Esmaeili, (2023) *"Multimodal feature selection from microarray data based on Dempster–Shafer evidence fusion"* **Journal of Supercomputing *79***(11): 12591–12621. DOI: 10.1007/s11227-023-05138-x.

[16] S. S. Baawi, E. Hakem, A. A. Al-Hamzawi, D. Al-Shammary, A. Ibaida, and A. M. Mahdi, (2025) *"Hilbert similarity convex for efficient EEG feature selections"* **Journal of Supercomputing *81***(5): 736. DOI: 10.1007/s11227-025-07199-6.

[17] M. S. Oh, A. B. Das, T. Kim, D. J. Love, and C. G. Brinton, (2024) *"Minimum Description Feature Selection for Complexity Reduction in Machine Learning-Based Wireless Positioning"* **IEEE Journal on Selected Areas in Communications *42***(9): 2585–2600. DOI: 10.1109/JSAC.2024.3413977.

[18] E. Hancer, B. Xue, and M. Zhang, (2023) *"An evolutionary filter approach to feature selection in classification for both single- and multi-objective scenarios"* **Knowledge-Based Systems *280***: 111008. DOI: 10.1016/j.knosys.2023.111008.

[19] M. Firouznia, P. Ruiu, and G. A. Trunfio, (2023) *"Adaptive cooperative coevolutionary differential evolution for parallel feature selection in high-dimensional datasets"* **Journal of Supercomputing *79***(14): 15215–15244. DOI: 10.1007/s11227-023-05226-y.

[20] L. N. Felber, H. Harbrecht, and M. Schmidlin, (2024) *"Identification of Sparsely Representable Diffusion Parameters in Elliptic Problems"* **SIAM Journal on Imaging Sciences *17***(1): 61–90. DOI: 10.1137/23M1565346.