Journal of Artificial Intelligence and Metaheuristic (JAIM)

♦ ISSN: 2833-5597 Vol: 07, No. 01, 2024 ♦ PP. 39–44

Article DOI: https://doi.org/10.54216/JAIM.070105

REVIEW ARTICLE

Advances and Challenges in Feature Selection Methods: A Comprehensive Review

Mohamed Ziad Ali[1,⋆] **Abdulrahman Abdullah**[1] **Ahmed Mohamed Zaki**[2] **Faris H. Rizk**[2] **Marwa M. Eid**[3] **Elsayed M. El-Kenway**[1]

[1] Department of Communications and Electronics, Delta Higher Institute of Engineering and Technology (DHIET), Mansoura 35111, Egypt

[2] Computer Science and Intelligent Systems Research Center, Blacksburg 24060, Virginia, USA

[3] Faculty of Artificial Intelligence, Delta University for Science and Technology, Mansoura 35111, Egypt

Emails: CH2100056@dhiet.edu.eg; CH2100197@dhiet.edu.eg; azaki@jcsis.org; farisrizk@jcsis.org; mmm@ieee.org; skenawy@ieee.org

Received: April 23, 2023 Revised: August 28, 2023 Accepted: January 18, 2024 ⋆ Corresponding author

ABSTRACT

This review examines the role of feature selection in data analytics and its growing use in data-intensive fields, including cybersecurity, healthcare, and finance. It surveys foundational concepts and current applications, with particular attention to reducing dimensionality and improving the interpretation of complex relationships in healthcare data. The review covers optimization-based methods, swarm intelligence, and feature-selection approaches for heart-disease diagnosis, intrusion detection, and other classification tasks. It also discusses hybrid Ant Colony Optimization and Grey Wolf Optimizer methods, alongside the Improved Salp Swarm Algorithm (ISSA). The reviewed studies illustrate that feature selection is more than a preprocessing step: it is an important component of machine learning and data science that can improve model efficiency, interpretability, and classification performance. Challenges and open research directions are summarized to support future work.

Keywords: Feature selection ▪ Optimization algorithms ▪ Machine learning ▪ Artificial intelligence ▪ Feature extraction

1. INTRODUCTION

Reliable preprocessing is essential for data-driven analysis. Feature selection is a central preprocessing task: it identifies useful information in large, evolving datasets while removing irrelevant or redundant attributes. It has been applied in financial analysis, biological data processing, intrusion detection, and medical applications. In healthcare, reducing dimensionality can also help researchers interpret relationships among variables and disease outcomes. Earlier work has examined feature selection for intrusion detection and ensemble-based intrusion detection, as well as DNA-microarray classification [1, 2, 3].

The growth of high-dimensional data has motivated methods that group features or exploit high-dimensional model representations, as well as broader reviews of feature selection for complex datasets [4, 5, 6]. In healthcare, optimization-based approaches combine genetic algorithms (GA) and support vector machines (SVMs) to identify attributes relevant to heart-disease diagnosis. Related work has also considered SVM-based stock-trend prediction [7, 8]. Other studies address streaming high-dimensional data and feature selection for IoT traffic intrusion detection [9, 10].

Ant Colony Optimization (ACO) has been applied to text classification, unsupervised feature selection, web-page classification, relevance–redundancy analysis, and network intrusion

© 2024 The Author(s). Published by ASPG. This article is licensed under CC BY 4.0.

detection. Hybrid ant-and-bee approaches have also been proposed for feature selection and classification [11, 12, 13, 14, 15, 16, 17]. Hybrid ACO methods combine filter and wrapper strategies; some use feature-dimension-dependent pheromone updates and memory mechanisms to retain the best candidate subsets. These approaches have been evaluated on microarray data and face-emotion recognition, among other tasks. Bio-inspired feature-selection research also includes systematic reviews, swarm-based optimization, and applications to software effort estimation [18, 19, 20, 21, 22, 23]. The Improved Salp Swarm Algorithm (ISSA), for example, introduces an inertia-weight parameter to improve the original Salp Swarm Algorithm and has been assessed with a *k*-nearest-neighbor classifier on UCI datasets. Together, these studies motivate a comparative review of feature-selection approaches, their application areas, and their limitations.

2. LITERATURE REVIEW

Feature selection identifies informative attributes for a particular task and can reduce computational cost while improving model clarity. Its applications range from cybersecurity and biological data processing to clinical analysis. The studies summarized in Table 1 cover medical applications, ensembles, swarm intelligence, optimization, intrusion detection, and other high-dimensional classification problems.

Medical feature-selection reviews cover DNA-microarray analysis, biomedical signals, and medical imaging, including practical case studies with patient data [24]. Ensemble learning combines several models and has been investigated as a way to improve feature selection; reviews of this area describe its foundations, recent developments, and future directions [25]. In large datasets, the number of attributes can greatly exceed the number of observations. Feature selection addresses this "curse of dimensionality" by reducing redundancy and noise while retaining attributes relevant to the target class. Swarm-intelligence reviews classify and compare methods developed for this challenge [26].

Optimization has also been applied to clinical prediction. A GA–SVM framework selects features for heart-disease diagnosis and is evaluated on the Cleveland heart-disease dataset in MATLAB. The study compares its results with methods including CFS, Relief, filtered subset, information gain, consistency subset, chi-squared, gain ratio, and other attribute-based filters; classifier performance is also examined using receiver-operating-characteristic analysis [27]. A separate wrapper method combines the Grey Wolf Optimizer with two-phase mutation. A sigmoid function maps the continuous search space to binary feature subsets; the first mutation phase removes features, while the second can add informative features. The method uses a *k*-nearest-neighbor classifier with cross-validation and is compared across more than 35 datasets with approaches such as flower, particle-swarm, multi-verse, whale-optimization, and bat algorithms [28]. Broader reviews discuss swarm-intelligence feature selection in classification, noting that representation and search strategies remain important and that comparative coverage of their advantages and limitations is incomplete [29].

In cybersecurity, a filter-and-wrapper intrusion-detection approach combines Feature Grouping based on the Linear Correlation Coefficient (FGLCC) with the Cuttlefish Algorithm (CFA). A decision tree is used for classification, and evaluation on the KDD Cup 99 dataset reports 95.03% accuracy, a 95.23% detection rate, and a 1.65% false-positive rate [30]. For IoT traffic, bijective soft sets and the CorrACC measure are used to select features for machine-learning classifiers. The wrapper-based CorrACC procedure filters candidate features using classification accuracy, and tests with four classifiers on the BoT-IoT dataset report accuracy above 95% [31].

ACO research includes a hybrid wrapper–filter design intended to reduce the computational cost of subset evaluation. The approach uses feature-dimension-dependent pheromone updates for multiobjective selection and a memory mechanism for retaining strong candidate solutions. Evaluations use *k*-nearest-neighbor and multilayer-perceptron classifiers on real-world datasets, including the NIPS 2003 feature-selection challenge, and further assess face-emotion-recognition and microarray applications [32]. The Salp Swarm Algorithm (SSA), a bio-inspired optimizer based on the swarming behavior of salps, has likewise been extended with an inertia-weight parameter intended to improve con-

Table 1. Summary of related feature-selection studies.

Ref.	Focus	Methodology	Key findings
[24]	Medical applications	Review of current techniques	Covers DNA microarrays, medical imaging, and biomedical signal processing, with case studies using patient data.
[25]	Ensemble feature selection	Review of principles and developments	Surveys ensemble learning for feature selection and identifies future research directions.
[26]	Curse of dimensionality	Review of swarm-intelligence techniques	Categorizes and compares contemporary swarm-based feature-selection methods.
[27]	Heart-disease detection	Genetic algorithm with SVM	Uses GA–SVM feature selection and evaluates its performance against existing methods.
[28]	Grey Wolf Optimizer	Wrapper approach with two-phase mutation	Combines the optimizer and mutation strategy for feature selection and compares results across datasets.
[29]	Swarm intelligence for classification	State-of-the-art review	Examines representation and search procedures and discusses practical considerations.
[30]	Intrusion detection	FGLCC feature grouping and Cuttlefish Algorithm	Reports 95.03% accuracy, a 95.23% detection rate, and a 1.65% false-positive rate on KDD Cup 99.
[31]	IoT cybersecurity	Bijective soft sets and CorrACC	Evaluates feature selection with four classifiers on BoT-IoT and reports accuracy above 95%.
[32]	ACO feature selection	Wrapper–filter hybridization	Tests the method on real-world, face-emotion, and microarray datasets.
[33]	Improved Salp Swarm Algorithm	Inertia-weight-enhanced SSA	Compares ISSA with SSA and other swarm methods on 23 UCI datasets.

vergence and solution reliability. In the reported comparison across 23 UCI datasets, ISSA is evaluated with a *k*-nearest-neighbor classifier and improves feature reduction and classification performance relative to basic SSA and other swarm optimizers [33].

3. FEATURE-SELECTION PROCEDURE

Feature selection is an essential preprocessing step in data analysis. Its objective is to identify and retain a useful subset of the original attributes, thereby reducing the effects of high dimensionality, noise, and redundancy while preserving predictive information [34]. A successful selection can also improve the interpretability of the resulting model and reduce computational demands. Figure 1 presents the general transition from an original dataset to a selected subset.



Figure 1. General concept of feature selection.

Feature-selection procedures generate candidate subsets, evaluate their suitability, and continue searching until a stopping condition is met. The resulting subset is then passed to the prediction model. Figure 2 illustrates this general process. Feature-selection approaches are commonly grouped into filter, wrapper, and embedded methods [35].

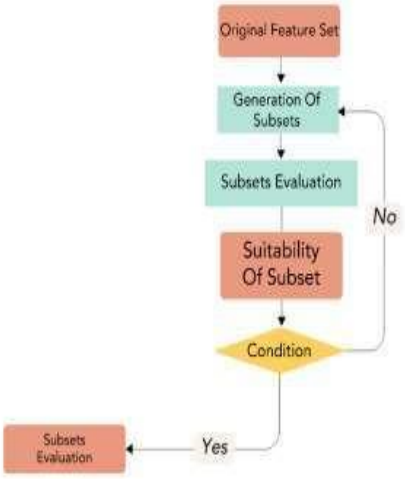


Figure 2. Illustration of the general feature-selection process.

4. FEATURE-SELECTION METHODS

Dimensionality reduction can be performed through feature extraction or feature selection. Feature extraction constructs new features from the original data, whereas feature selection retains a subset of the original attributes. The reviewed methods are grouped into information-theory-based, filter, wrapper, and embedded approaches. Figure 3 shows this categorization.

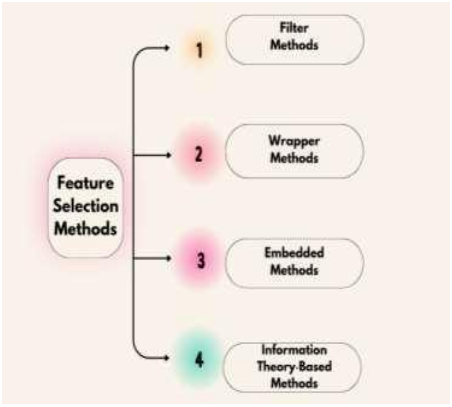


Figure 3. Categorization of feature-selection methods.

4.1 Filter Methods

Filter methods rank variables according to statistical criteria, such as the distance or correlation between a feature and the output. They select high-ranked attributes before model training and are generally computationally efficient and less prone to overfitting. However, they may ignore dependencies among features. The reviewed work describes their use in prediction from technical indicators [36].

4.2 Wrapper Methods

Wrapper methods evaluate candidate feature subsets using a learning algorithm. Approaches such as recursive feature elimination, sequential selection, and genetic algorithms can improve task-specific performance, but repeated training and cross-validation make them computationally expensive, particularly on large datasets [37].

4.3 Embedded Methods

Embedded methods perform feature selection during model training. By incorporating selection into the learning process, they combine aspects of filter and wrapper approaches and can be more computationally efficient than wrappers. Because they model feature interactions as part of training, they can also capture dependencies that simple filter methods may miss [36].

4.4 Information-Theory-Based Methods

Information-theory-based feature selection evaluates the information provided by candidate attributes. Distributed implementations on platforms such as Apache Spark have been proposed to improve the speed and scalability of established methods. Experiments on real-world datasets report benefits over sequential processing, including for datasets with very high dimensionality or many observations [38].

5. FEATURE-EXTRACTION TECHNIQUES AND APPLICATIONS

Feature extraction reduces the number of input dimensions by generating new features that represent information in the original set. The reviewed studies describe statistical and optimization-based extraction strategies. Feature-selection methods are also applied across domains where data are high-dimensional and the choice of informative attributes affects model performance.

5.1 Application in Neuroimaging

In neuroimaging, feature selection can address the large number of measurements relative to the number of samples. Sparse-regularization approaches combine optimization and classification to constrain feature coefficients; coefficients driven to zero effectively remove the corresponding attributes from the classifier. Evaluations on three functional magnetic resonance imaging (fMRI) datasets report improved classification accuracy, particularly when the initial feature set is large, and highlight the role of sparse regularization in functional localization and generalization [39].

5.2 Application to Cancer Diagnosis

Cancer diagnosis studies use sensitivity and specificity to assess classification performance. A multiobjective GA-based method selects genes from microarray datasets, with a support vector machine used for classification and 10-fold cross-validation applied to randomly generated subsets. This approach seeks to select informative genes while balancing diagnostic performance; comparative evaluations consider alternatives with multiple objectives [40].

6. RESULTS AND DISCUSSION

The reviewed literature demonstrates the breadth of feature selection across cybersecurity, healthcare, neuroimaging, cancer diagnosis, and large-scale data analysis. The methods examined include statistical filters, wrapper searches, embedded selection, information-theoretic approaches, and metaheuristic optimization. Across these areas, feature selection can reduce dimensionality and computational cost while supporting model accuracy and interpretability.

The reported case studies also show that performance depends on the dataset, classifier, and search strategy. Optimization and swarm-intelligence methods provide flexible ways to explore large feature spaces, while hybrid approaches combine selection strategies to balance search cost and predictive performance. At the same time, dataset bias, computational scalability, and the design of search and representation procedures remain important considerations. Further comparative evaluation is needed to clarify the strengths and limitations of methods under consistent experimental settings.

7. CONCLUSION

This review has surveyed feature-selection concepts, methods, and applications in data science and machine learning. It has covered filter, wrapper, embedded, and information-theory-based approaches, as well as optimization and swarm-intelligence techniques such as ACO and ISSA. The cited applications demonstrate how selecting informative attributes can address high dimensionality in healthcare, cybersecu-rity, neuroimaging, and cancer diagnosis. Feature selection can contribute to more efficient and interpretable predictive models, but its effectiveness depends on the data and evalu-ation strategy. Future work should further assess scalability, method limitations, and robust feature-selection designs for complex real-world datasets.

REFERENCES

[1] F. Amiri, M. Rezaei Yousefi, C. Lucas, A. Shakery, and N. Yazdani. Mutual information-based feature selection for intrusion detection systems. *Journal of Network and Computer Applications*, 34(4):1184–1199, 2011.

[2] E. Jaw and X. Wang. Feature selection and ensemble-based intrusion detection system: An efficient and comprehensive approach. *Symmetry*, 13(10), 2021.

[3] Esra'a Alhenawi, Rizik Al-Sayyed, Amjad Hudaib, and Seyedali Mirjalili. Feature selection methods on gene expression microarray data for cancer classification: A systematic review. *Computers in Biology and Medicine*, 140:105051, 2022.

[4] M. García-Torres, F. Gómez-Vela, B. Melián-Batista, and J. M. Moreno-Vega. High-dimensional feature selection via feature grouping: A variable neighborhood search approach. *Information Sciences*, 326:102–118, 2016.

[5] V. Bolón-Canedo, N. Sánchez-Maroño, and A. Alonso-Betanzos. Feature selection for high-dimensional data. *Progress in Artificial Intelligence*, 5(2):65–75, 2016.

[6] G. Taskin, H. Kaya, and L. Bruzzone. Feature selection based on high dimensional model representation for hyperspectral images. *IEEE Transactions on Image Processing*, 26(6):2918–2928, 2017.

[7] I. Aljarah, A. M. Al-Zoubi, H. Faris, M. A. Hassonah, S. Mirjalili, and H. Saadeh. Simultaneous feature selection and support vector machine optimization using the grasshopper optimization algorithm. *Cognitive Computation*, 10(3):478–495, 2018.

[8] M.-C. Lee. Using support vector machine with a hybrid feature selection method to the stock trend prediction. *Expert Systems with Applications*, 36(8):10896–10904, 2009.

[9] E. A. K. Zaman, A. Mohamed, and A. Ahmad. Feature selection for online streaming high-dimensional data: A state-of-the-art review. *Applied Soft Computing*, 127:109355, 2022.

[10] Joffrey L. Leevy, John Hancock, Taghi M. Khoshgoftaar, and Jared M. Peterson. Iot information theft prediction using ensemble feature selection. *Journal of Big Data*, 9:6, 2022.

[11] M. H. Aghdam, N. Ghasem-Aghaee, and M. E. Basiri. Text feature selection using ant colony optimization. *Expert Systems with Applications*, 36(3):6843–6853, 2009.

[12] B. Z. Dadaneh, H. Y. Markid, and A. Zakerolhosseini. Unsupervised probabilistic feature selection using ant colony optimization. *Expert Systems with Applications*, 53:27–42, 2016.

[13] S. Tabakhi, P. Moradi, and F. Akhlaghian. An unsupervised feature selection algorithm based on ant colony optimization. *Engineering Applications of Artificial Intelligence*, 32:112–123, 2014.

[14] An ant colony optimization based feature selection for web page classification. *The Scientific World Journal*, page 649260, 2014.

[15] S. Tabakhi and P. Moradi. Relevance-redundancy feature selection based on ant colony optimization. *Pattern Recognition*, 48(9):2798–2811, 2015.

[16] H.-H. Gao, H.-H. Yang, and X.-Y. Wang. Ant colony optimization based network intrusion feature selection and detection. In *2005 International Conference on Machine Learning and Cybernetics*, volume 6, pages 3871–3875, 2005.

[17] P. Shunmugapriya and S. Kanmani. A hybrid algorithm using ant and bee colony optimization for feature selection and classification (AC-ABC hybrid). *Swarm and Evolutionary Computation*, 36:27–36, 2017.

[18] Montha Petwan and Ku Ruhana Ku-Mahamud. A review on bio-inspired optimization method for supervised feature selection. *International Journal of Advanced Computer Science and Applications*, 13(5):122–132, 2022.

[19] Kishore Balasubramanian and N. P. Ananthamoorthy. Correlation-based feature selection using bio-inspired algorithms and optimized kelm classifier for glaucoma diagnosis. *Applied Soft Computing*, 128:109432, 2022.

[20] M. Zivkovic, C. Stoean, A. Chhabra, N. Budimirovic, A. Petrovic, and N. Bacanin. Novel improved salp swarm algorithm: An application for feature selection. *Sensors*, 22(5):1711, 2022.

[21] V. R. Balasaraswathi, M. Sugumaran, and Y. Hamid. Feature selection techniques for intrusion detection using non-bio-inspired and bio-inspired optimization algorithms. *Journal of Communications and Information Networks*, 2(4):107–119, 2017.

[22] Mohammad Aizat Basir, Yuhanis Yusof, and Mohamed Saifullah Hussin. Optimization of attribute selection model using bio-inspired algorithms. *Journal of Information and Communication Technology*, 18(1):35–55, 2019.

[23] A. Ali and C. Gravino. Using bio-inspired feature selection algorithms in software effort estimation: A systematic literature review. In *2019 45th Euromicro Conference on Software Engineering and Advanced Applications*, pages 220–227, 2019.

[24] B. Remeseiro and V. Bolón-Canedo. A review of feature selection methods in medical applications. *Computers in Biology and Medicine*, 112:103375, 2019.

[25] V. Bolón-Canedo and A. Alonso-Betanzos. Ensembles for feature selection: A review and future trends. *Information Fusion*, 52:1–12, 2019.

[26] M. Rostami, K. Berahmand, E. Nasiri, and S. Forouzandeh. Review of swarm intelligence-based feature selection methods. *Engineering Applications of Artificial Intelligence*, 100:104210, 2021.

[27] C. B. Gokulnath and S. P. Shantharajah. An optimized feature selection based on genetic approach and support vector machine for heart disease. *Cluster Computing*, 22(6):14777–14787, 2019.

[28] M. Abdel-Basset, D. El-Shahat, I. El-henawy, V. H. C. de Albuquerque, and S. Mirjalili. A new fusion of grey wolf optimizer algorithm with a two-phase mutation for feature selection. *Expert Systems with Applications*, 139:112824, 2020.

[29] B. H. Nguyen, B. Xue, and M. Zhang. A survey on swarm intelligence approaches to feature selection in data mining. *Swarm and Evolutionary Computation*, 54:100663, 2020.

[30] S. Mohammadi, H. Mirvaziri, M. Ghazizadeh-Ahsaee, and H. Karimipour. Cyber intrusion detection by combined feature selection algorithm. *Journal of Information Security and Applications*, 44:80–88, 2019.

[31] M. Shafiq, Z. Tian, A. K. Bashir, X. Du, and M. Guizani. IoT malicious traffic identification using wrapper-based feature selection mechanisms. *Computers & Security*, 94:101863, 2020.

[32] M. Ghosh, R. Guha, R. Sarkar, and A. Abraham. A wrapper-filter feature selection technique based on ant colony optimization. *Neural Computing and Applications*, 32(12):7839–7857, 2020.

[33] Ah. E. Hegazy, M. A. Makhlouf, and Gh. S. El-Tawel. Improved salp swarm algorithm for feature selection. *Journal of King Saud University - Computer and Information Sciences*, 32(3):335–344, 2020.

[34] D. Khafaga, E.-S. El-kenawy, F. Alrowais, S. Kumar, A. Ibrahim, and A. Abdelhamid. Novel optimized feature selection using metaheuristics applied to physical benchmark datasets. *Computers, Materials & Continua*, 74(2):4027–4041, 2022.

[35] U. Stanczyk. Feature evaluation by filter, wrapper, and embedded approaches. In U. Stanczyk and L. C. Jain, editors, *Feature Selection for Data and Pattern Recognition*, pages 29–44. Springer, 2015.

[36] Ameer Alhaq and Eman Salih Al-Shamery. Enhancing prediction of NASDAQ stock market based on technical indicators. *Journal of Engineering and Applied Sciences*, 13:4630–4636, 2018.

[37] G. Chen and J. Chen. A novel wrapper method for feature selection and its applications. *Neurocomputing*, 159:219–226, 2015.

[38] S. Ramirez-Gallego, H. Mourino-Talin, D. Martinez-Rego, V. Bolon-Canedo, J. M. Benitez, A. Alonso-Betanzos, and F. Herrera. An information theory-based feature selection framework for big data under apache spark. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 48(9):1441–1453, 2018.

[39] K. Kampa, S. Mehta, C. A. Chou, W. A. Chaovalitwongse, and T. J. Grabowski. Sparse optimization in feature selection: Application in neuroimaging. *Journal of Global Optimization*, 59(2):439–457, 2014.

[40] J. Garcia-Nieto, E. Alba, L. Jourdan, and E. Talbi. Sensitivity and specificity based multiobjective approach for feature selection: Application to cancer diagnosis. *Information Processing Letters*, 109(16):887–896, 2009.