Journal of Optimization in Soft Computing (JOSC)

Vol. 2, Issue 4, pp: (16-20), Winter-2024
Journal homepage: https://sanad.iau.ir/journal/josc

Paper Type (Research paper)

A Review of Feature Selection

Jafar Abdollahi[1], Babak Nouri-Moghaddam[1], Naser Mikaeilvand[2], Sajjad Jahanbakhsh Gudakahriz[3], Ailin Khosravani[1], Abbas Mirzaei[1]

1. Department of Computer Engineering, Ardabil Branch, Islamic Azad University, Ardabil, Iran
2. Department of Computer Engineering, Central Tehran Branch, Islamic Azad University, Tehran, Iran
3. Department of Computer Engineering, Germi Branch, Islamic Azad University, Germi, Iran

Article Info

Article History:
Received: 2024/11/28
Revised: 2024/12/29
Accepted: 2025/03/13

DOI:

Keywords:
Data Mining; Medical Applications; Dimension Reduction; Feature Selection

**Corresponding Author's Email Address:*

Abstract

Feature selection is a preprocessing technique that identifies the salient features of a given scenario. It has been used in the past for a wide range of problems, including intrusion detection systems, financial problems, and the analysis of biological data. Feature selection has been especially useful in medical applications, where it may help identify the underlying reasons for an illness in addition to reducing dimensionality. We provide some basic concepts of medical applications and the necessary background information on feature selection. We review the most recent feature selection methods developed for and applied to medical problems, covering a broad spectrum of applications including medical imaging, DNA microarray data analysis, and biomedical signal processing. A case study of two medical applications utilizing actual patient data is used to demonstrate the usefulness of applying feature selection techniques to medical challenges and to highlight how these methods function in practical scenarios.

1. Introduction

Feature selection is one way to reduce dimensionality; in this strategy, only significant traits are retained while superfluous and redundant ones are discarded. Two ways that a reduction in input dimensionality might improve performance are either decreasing learning time and model complexity or increasing generalization capabilities and classification accuracy. Using the right features might improve problem understanding and reduce measurement expenses. In certain situations, the impact of feature selection may be substantial; for example, in microarray data analysis, just two of the 7129 features may be used to improve classification performance [1].

There are two kinds of feature selection models:

- Supervised Models: The technique that selects features based on the output label class is known as supervised feature selection.
- Unsupervised Models: An approach that selects features without requiring knowledge of the output label class is known as unsupervised feature selection.

In many applications, it has been necessary to combine pattern recognition algorithms with FS techniques, since many of them were not designed to handle large amounts of irrelevant data at first. Preventing overfitting and enhancing model performance—more specifically, prediction performance in supervised classification and improved cluster detection in clustering—are the primary objectives of feature selection. Other objectives include (a) producing faster and more efficient models and (b) gaining a deeper comprehension of the underlying processes that generated the data. Nevertheless, the advantages of feature selection strategies are not without a price, as the search for a subset of pertinent characteristics raises the bar for modeling complexity. We must now determine the model's optimal parameters for the optimal feature subset in addition to optimizing its parameters for the full feature subset, since there is no guarantee that the model's ideal parameters for the entire feature set will also be optimal for the optimal feature subset. Thus, identifying the optimal subset of pertinent attributes expands the scope of the search within the model hypothesis space. Every feature selection technique uses a different technique to include this search in the extra space of feature subsets when choosing a model [2, 4, 5].

Filter approaches assess the significance of the features by concentrating on the intrinsic properties of the data. Generally, features are ranked according to their relevance, and those with lower scores are ignored. This selection of attributes is then given as input to the classification algorithm. Because of the advantages of filter approaches—which include their simplicity and speed in computation, their independence from the classification algorithm, and their ability to scale to extremely high-dimensional datasets—only one feature selection process is needed before multiple classifiers can be evaluated [2,].

Unlike filter techniques, which tackle the problem of finding a suitable feature subset independently of the model selection phase, wrapper approaches incorporate the model hypothesis search into the feature subset search. In this scenario, various feature subsets are generated and evaluated in the space of possible feature subsets utilizing a predefined search method [2].

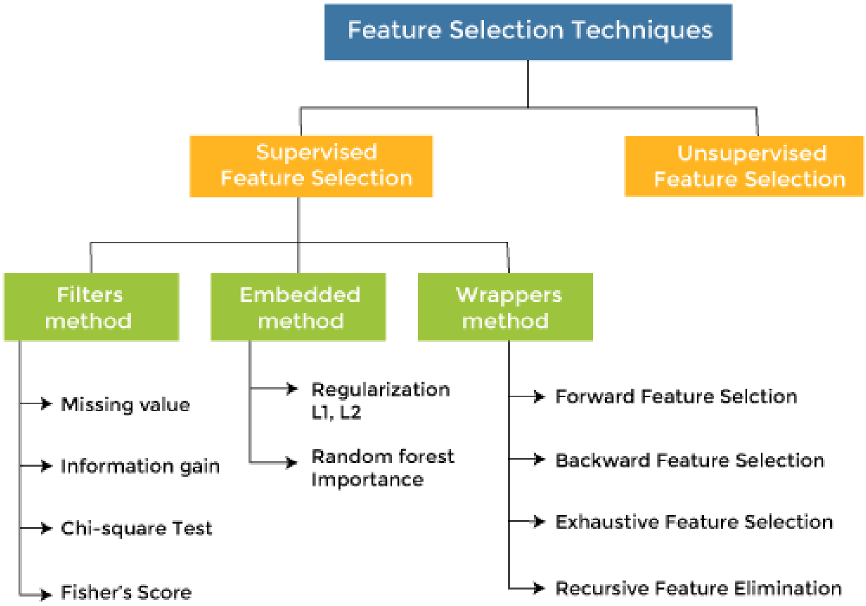

Figure 1: Overview of Feature Selection Strategies [2]

2. Choosing Feature

Medical image and healthcare analysis, including diabetes [15, 24, 32], breast cancer [16, 25], healthcare system [17], forecasting [18], stock market [19], stroke [20], COVID-19 [21], types of epidemic [22], medicinal plants [23], heart [26], lung cancer [27, 30], social networks [28], prediction of diphenhydramine [29], and bupro, have all benefited from the successful application of artificial intelligence, which includes machine learning and deep learning. We present a comprehensive review of feature selection methods applied in medicine over the past five years, some developed on the fly to tackle specific problems. Specifically, feature selection has been applied in three main medical fields: biomedical signal processing, DNA microarray data, and medical imaging. We then go on to discuss current advancements in each of these fields. Then, we discuss how feature selection is applied to two actual medical image analysis situations and show the benefits that follow from doing so [1,39, 44]. The following is a summary of feature choices.

- Feature Selection: Select a subset of input features from the dataset.
- Unsupervised: Do not use the target variable (e.g. remove redundant variables).
- Correlation
- Supervised: Use the target variable (e.g. remove irrelevant variables).
- Wrapper: Search for well-performing subsets of features.
- RFE
- Filter: Select subsets of features based on their relationship with the target.
- Statistical Methods
- Feature Importance Methods

- Intrinsic: Algorithms that conduct automated feature selection during training.
- Decision Trees
- Dimensionality Reduction: Project input data into a lower-dimensional feature space.

The figure above offers an overview of the hierarchy of feature selection strategies.

A. Primary Concepts

In line with how they combine the selection algorithm and the model development, the feature selection strategies are frequently categorized into three forms.

B. Filter Method

Filtering strategy for picking features: Methods of the filter type pick variables without attention to the model. They are just reliant on universal qualities like the correlation with the expected variable. Filtering strategies reduce the least fascinating aspects. The following variables will be added to a regression model or classification scheme used to classify or forecast data. These approaches offer good computational efficiency and are resistant to overfitting. When filter algorithms do not take into consideration the relationships between variables, duplicated variables are typically picked. However, more complicated features, like the Fast Correlation Based Filter (FCBF) algorithm, aim to decrease this problem by removing variables that are highly linked with one another.

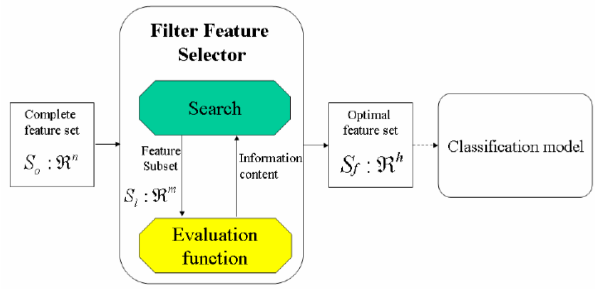

Figure 2: The Hierarchy of Feature Selection Methods [29]

C. Wrapper Method

Wrapper approach for feature selection: Wrapper techniques, in contrast to filter operations, look at subsets of variables, which makes it possible to find any possible interactions between variables. Overfitting becomes more likely when there are insufficient observations, and computation time grows dramatically as there are more variables [41, 42, 43].

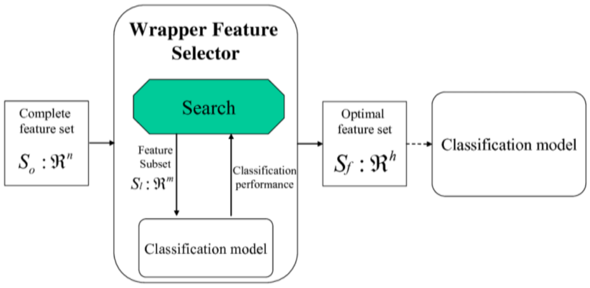

Figure 2: Wrapper Method in Feature Selection [43]

D. Embedded Methodology

Choosing features using the embedded method: Embedded strategies have recently been created with the goal of combining the advantages of the two previous approaches. A learning algorithm, such as the FRMT technique, uses its own variable selection mechanism to carry out feature selection and classification simultaneously [40, 45].

3. Finding

In many bioinformatics applications, feature selection algorithms are needed. Dedicated bioinformatics applications have yielded a wide range of recently proposed methods to supplement the large body of previously developed methods in the fields of data mining and machine learning [2, 46]. In recent years, there has been a notable increase in the use of feature selection approaches in medical datasets. The challenging task in feature selection is to identify the perfect subset of relevant and non-redundant qualities that will provide an optimal solution without adding to the complexity of the modeling process. Therefore, it's critical to draw attention to recent advancements in this area and educate practitioners on feature selection strategies that have worked well with medical data sets. The findings demonstrate that most feature selection methods now in use are based on univariate ranking, which overlooks the stability of the selection algorithms, interactions between variables, and the requirement for additional features to attain very high accuracy. Less attributes may still lead to maximum classification accuracy, but more work has to be done in this area [3, 14, 33-38, 47].

Tables1: Summary of Feature Selection Methods [47]

Filter methods	Wrapper methods	Embedded methods
Generic set of methods which do not incorporate a **specific machine learning algorithm.**	Evaluates on a **specific machine learning algorithm** to find optimal features.	Embeds (fix) features during **model building process.** Feature selection is done by observing each iteration of model training phase.
Much **faster** compared to Wrapper methods in terms of time complexity	**High computation time** for a dataset with many features	Sits **between Filter methods and Wrapper methods** in terms of time complexity
Less prone to **over-fitting**	High chances of **over-fitting** because it involves training of machine learning models with different combination of features	Generally used to reduce **over-fitting** by **penalizing** the coefficients of a model being too large.
Examples – **Correlation, Chi-Square test, ANOVA, Information gain** etc.	Examples - **Forward Selection, Backward elimination, Stepwise selection** etc.	Examples - **LASSO, Elastic Net, Ridge Regression** etc.

4. Conclusion

Feature selection is a fundamental technique for enhancing machine learning models by reducing dimensionality, improving accuracy, and optimizing computational efficiency. This review has highlighted the significance of feature selection in various medical applications, including biomedical signal processing, DNA microarray analysis, and medical imaging. While existing methods provide effective solutions for reducing irrelevant and redundant features, many challenges remain in achieving an optimal subset of features that balances performance and computational cost. Recent advances have shown that hybrid models combining filter, wrapper, and embedded approaches yield promising results. However, issues such as model stability, feature interaction, and scalability need further exploration. Future research should focus on developing more robust feature selection techniques tailored for complex medical datasets, ensuring better diagnostic accuracy and predictive performance in healthcare applications.

References

[1] Remeseiro, B., & Bolon-Canedo, V. (2019). A review of feature selection methods in medical applications. Computers in biology and medicine, 112, 103375.

[2] Saeys, Y., Inza, I., & Larranaga, P. (2007). A review of feature selection techniques in bioinformatics. bioinformatics, 23(19), 2507-2517.

[3] Mwadulo, M. W. (2016). A review on feature selection methods for classification tasks.

[4] Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., & Liu, H. (2017). Feature selection: A data perspective. ACM computing surveys (CSUR), 50(6), 1-45.

[5] Duan, H., & Mirzaei, A. (2023). Adaptive Rate Maximization and Hierarchical Resource Management for Underlay Spectrum Sharing NOMA HetNets with Hybrid Power Supplies. Mobile Networks and Applications, 1-17.

[6] Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. Computers & Electrical Engineering, 40(1), 16-28.

[7] Dash, M., & Liu, H. (1997). Feature selection for classification. Intelligent data analysis, 1(1-4), 131-156.

[8] Somarin, A. M., Barari, M., & Zarrabi, H. (2018). Big data based self-optimization networking in next generation mobile networks. Wireless Personal Communications, 101(3), 1499-1518.

[9] Liu, H., & Motoda, H. (Eds.). (2007). Computational methods of feature selection. CRC press.

[10] Koller, D., & Sahami, M. (1996, July). Toward optimal feature selection. In ICML (Vol. 96, No. 28, p. 292).

[11] Venkatesh, B., & Anuradha, J. (2019). A review of feature selection and its methods. Cybernetics and information technologies, 19(1), 3-26.

[12] Cai, J., Luo, J., Wang, S., & Yang, S. (2018). Feature selection in machine learning: A new perspective. Neurocomputing, 300, 70-79.

[13] Zhao, Z., Morstatter, F., Sharma, S., Alelyani, S., Anand, A., & Liu, H. (2010). Advancing feature selection research. ASU feature selection repository, 1-28.

[14] Saeys, Y., Abeel, T., & Van de Peer, Y. (2008). Robust feature selection using ensemble feature selection techniques. In Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2008, Antwerp, Belgium, September 15-19, 2008, Proceedings, Part II 19 (pp. 313-325). Springer Berlin Heidelberg.

[15] Abdollahi, J., Moghaddam, B. N., & Parvar, M. E. (2019). Improving diabetes diagnosis in smart health using genetic-based Ensemble learning algorithm. Approach to IoT Infrastructure. Future Gen Distrib Systems J, 1, 23-30.

[16] Abdollahi, J., Keshandehghan, A., Gardaneh, M., Panahi, Y., & Gardaneh, M. (2020). Accurate detection of breast cancer metastasis using a hybrid model of artificial intelligence algorithm. Archives of Breast Cancer, 22-28.

[17] Nematollahi, M., Ghaffari, A., & Mirzaei, A. (2024). Task and resource allocation in the internet of things based on an improved version of the moth-flame optimization algorithm. Cluster Computing, 27(2), 1775-1797.

[18] Abdollahi, J., Irani, A. J., & Nouri-Moghaddam, B. (2021). Modeling and forecasting Spread of COVID-19 epidemic in Iran until Sep 22, 2021, based on deep learning. arXiv preprint arXiv:2103.08178.

[19] Abdollahi, J., & Mahmoudi, L. Investigation of artificial intelligence in stock market prediction studies. In 10th International Conference on Innovation and Research in Engineering Science.

[20] Narimani, Y., Zeinali, E., & Mirzaei, A. (2022). QoS-aware resource allocation and fault tolerant operation in hybrid SDN using stochastic network calculus. Physical Communication, 53, 101709.

[21] Abdollahi, J. (2020). A review of Deep learning methods in the study, prediction and

management of COVID-19. In 10th International Conference on Innovation and Research in Engineering Science.

[22] Abdollahi, J., & Mahmoudi, L. (2022, February). An Artificial Intelligence System for Detecting the Types of the Epidemic from X-rays: Artificial Intelligence System for Detecting the Types of the Epidemic from X-rays. In 2022 27th International Computer Conference, Computer Society of Iran (CSICC) (pp. 1-6). IEEE.

[23] Abdollahi, J. (2022, February). Identification of medicinal plants in ardabil using deep learning: identification of medicinal plants using deep learning. In 2022 27th International Computer Conference, Computer Society of Iran (CSICC) (pp. 1-6). IEEE.

[24] Abdollahi, J., & Nouri-Moghaddam, B. (2022). Hybrid stacked ensemble combined with genetic algorithms for diabetes prediction. Iran Journal of Computer Science, 5(3), 205-220.

[25] Abdollahi, J., Davari, N., Panahi, Y., & Gardaneh, M. (2022). Detection of Metastatic Breast Cancer from Whole-Slide Pathology Images Using an Ensemble Deep-Learning Method: Detection of Breast Cancer using Deep-Learning. Archives of Breast Cancer, 364-376.

[26] Jahanbakhsh Gudakahriz, S., Momtaz, V., Nouri-Moghadam, B., Mirzaei, A., & Vajed Khiavi, M. (2025). Link life time and energy-aware stable routing for MANETs. International Journal of Nonlinear Analysis and Applications.

[27] Javadzadeh Barzaki, M. A., Negaresh, M., Abdollahi, J., Mohammadi, M., Ghobadi, H., Mohammadzadeh, B., & Amani, F. (2022, July). USING DEEP LEARNING NETWORKS FOR CLASSIFICATION OF LUNG CANCER NODULES IN CT IMAGES. In Iranian Congress of Radiology (Vol. 37, No. 2, pp. 34-34). Iranian Society of Radiology.

[28] Khavandi, H., Moghadam, B. N., Abdollahi, J., & Branch, A. (2023). Maximizing the Impact on Social Networks using the Combination of PSO and GA Algorithms. Future Generation in Distributed Systems, 5, 1-13.

[28] Mehrpour, O., Saeedi, F., Abdollahi, J., Amirabadizadeh, A., & Goss, F. (2023). The value of machine learning for prognosis prediction of diphenhydramine exposure: National analysis of 50,000 patients in the United States. Journal of Research in Medical Sciences, 28(1), 49.

[29] Rad, K. J., & Mirzaei, A. (2022). Hierarchical capacity management and load balancing for HetNets using multi-layer optimisation methods. International Journal of Ad Hoc and Ubiquitous Computing, 41(1), 44-57.

[30] Mehrpour, O., Saeedi, F., Vohra, V., Abdollahi, J., Shirazi, F. M., & Goss, F. (2023). The role of decision tree and machine learning models for outcome prediction of bupropion exposure: A nationwide analysis of more than 14 000 patients in the United States. Basic & Clinical Pharmacology & Toxicology, 133(1), 98-110.

[31] Nemati, Z., Mohammadi, A., Bayat, A., & Mirzaei, A. (2024). The impact of financial ratio reduction on supervised methods' ability to detect financial statement fraud. Karafan Quarterly Scientific Journal.

[32] Tajidini, F., & Kheiri, M. J. (2023). Recent advancement in Disease Diagnostic using machine learning: Systematic survey of decades, comparisons, and challenges. arXiv preprint arXiv:2308.01319.

[33] Zargar, H. H., Zargar, S. H., Mehri, R., & Tajidini, F. (2023). Using VGG16 Algorithms for classification of lung cancer in CT scans Image. arXiv preprint arXiv:2305.18367.

[34] Mirzaei, A., Barari, M., & Zarrabi, H. (2019). Efficient resource management for non-orthogonal multiple access: A novel approach towards green hetnets. Intelligent Data Analysis, 23(2), 425-447.

[35] Nemati, Z., Mohammadi, A., Bayat, A., & Mirzaei, A. (2023). Financial Ratios and Efficient Classification Algorithms for Fraud Risk Detection in Financial Statements. International Journal of Industrial Mathematics.

[36] Tajidini, F., & Mehri, R. A survey of using Deep learning algorithms for the Covid-19 (SARS-CoV-2) pandemic: A review.

[37] Tajidini, F., & Piri, M. Machine Learning Methods for prediction of Diabetes: A Narrative.

[38] Mirzaei, A., & Najafi Souha, A. (2021). Towards optimal configuration in MEC Neural networks: deep learning-based optimal resource allocation. Wireless Personal Communications, 121(1), 221-243.

[39] Nematollahi, M., Ghaffari, A., & Mirzaei, A. (2024). Task offloading in Internet of Things based on the improved multi-objective aquila optimizer. Signal, Image and Video Processing, 18(1), 545-552.

[40] Nemati, Z., Mohammadi, A., Bayat, A., & Mirzaei, A. (2024). Metaheuristic and Data Mining Algorithms-based Feature Selection Approach for Anomaly Detection. IETE Journal of Research, 1-15.

[41] Nematia, Z., Mohammadia, A., Bayata, A., & Mirzaeib, A. (2024). Predicting fraud in financial statements using supervised methods: An analytical comparison. International Journal of Nonlinear Analysis and Applications, 15(8), 259-272.

[42] Jahandideh, Y., & Mirzaei, A. (2021). Allocating Duplicate Copies for IoT Data in Cloud Computing based on Harmony Search Algorithm. IETE Journal of Research, 1-14.

[43] Mikaeilvand, N., Ojaroudi, M., & Ghadimi, N. (2015). Band-Notched Small Slot Antenna Based on Time-Domain Reflectometry Modeling for UWB Applications. The Applied Computational Electromagnetics Society Journal (ACES), 682-687.

[44] Mikaeilvand, N. (2011). On solvability of fuzzy system of linear matrix equations. J Appl Sci Res, 7(2), 141-153.

[45] Allahviranloo, T., & Mikaeilvand, N. (2011). Non zero solutions of the fully fuzzy linear systems. Appl. Comput. Math, 10(2), 271-282.

[46] Nemati, Z., Mohammadi, A., Bayat, A., & Mirzaei, A. (2024). Fraud Risk Prediction in Financial Statements through Comparative Analysis of Genetic Algorithm, Grey Wolf Optimization, and Particle Swarm Optimization. Iranian Journal of Finance, 8(1), 98-130.