Al-Noor Journal of Engineering Management and Computer Science

ISSN: 3079-0689 (Online)

https://njemcs.edu.iq/index.php/njemcs/

A Review of Feature Selection Techniques for Medical Datasets

Baidaa Mutasher Rashed [1], Zahoor M. Aydam[2]

[1]Department of Computers, College of Computer Science and Mathematics, University of Thi-Qar, Nasiriyah, Thi-Qar, Iraq.
[2] Department of Department of Information Technology, College of Computer Science and Mathematics, University of Thi-Qar, Thi-Qar, 64001, Iraq

ARTICLE INFO

Article history:
Received 15-06-2026
Revised 15-06-2026
Accepted 23-06-2026
Available online 24-06-2026

Keywords:
Feature selection
Filter methods
Wrapper methods
Embedded methods

ABSTRACT

Feature selection is an important part of machine learning, especially in medical research with high-dimensional datasets. Selecting relevant features can improve the accuracy, interpretability, and efficiency of prediction models. There are three feature selection methods, i.e., filter, wrapper, and embedding methods. In this paper, we analyze these algorithms applied to medical datasets; the basic idea of the most popular algorithms in each category is described, and advantages, efficiency, and disadvantages are discussed.

The paper also describes the problem of complexity, the number of samples, and class imbalance in medical datasets. Finally, the analysis offers some clues for the most effective feature selection strategies and possible paths for future research in the medical field. This review is expected to serve as a useful guide for researchers and practitioners who are interested in choosing the best features to optimize the machine learning models of medical applications.

1. Introduction

Machine Learning (ML) is a powerful tool for pattern recognition and prediction across many scientific fields where large amounts of data are available. ML models have revolutionized healthcare and medical research by helping them understand what diseases people have, how they change, how to develop new pharmaceuticals, and how to tailor therapies for each patient. However, medical datasets have major problems that make it more difficult to effectively employ ML algorithms. Examples of such problems are noise, duplication, useless features, and high dimensionality [1]. ML models with too many features can cause many problems, especially when they include unnecessary or redundant features. This can cause increased computational complexity, longer training times, less generalizability of the model, and less interpretability. Genomic studies can study thousands of genes, but only a subset of those will give a true picture of a given disease [2]. Electronic health records (EHRs) contain a wealth of information about patients, but not all of this information is relevant to a specific diagnostic or prognostic task [3]. Feature selection solves these issues by being an integral part of the ML process that narrows down the original dataset to the most relevant and usable features. Feature selection is used to enhance the predictive accuracy of models,

Corresponding author E-mail address: baidaaalsafy@utq.edu.iq
https://doi.org/10.71229/93rpkv96

lower computing costs (since there are fewer features), make the model more straightforward to explain (by removing unnecessary variables), and lower the risk of overfitting (particularly when the sample size is small)[4].

In the case of medical datasets, the choice of the right features becomes even more important. Interpretability is important because a doctor should understand why a diagnosis or prediction was made before making a clinical decision. Feature selection can be useful to identify useful biomarkers and clinical indicators, which can lead to new medical therapies and better care of patients. Thus, it is important to develop basic, reliable models that can work on new patient data [5].

2. Feature Selection Algorithms

There are three basic types of feature selection techniques: Filter methods, Wrapper methods, and Embedded methods [6]. All categories have different ways of selecting the most important features and thus have different trade-offs in cost, performance, and generalizability. Figure 1 shows this classification and gives several common algorithms for each kind.

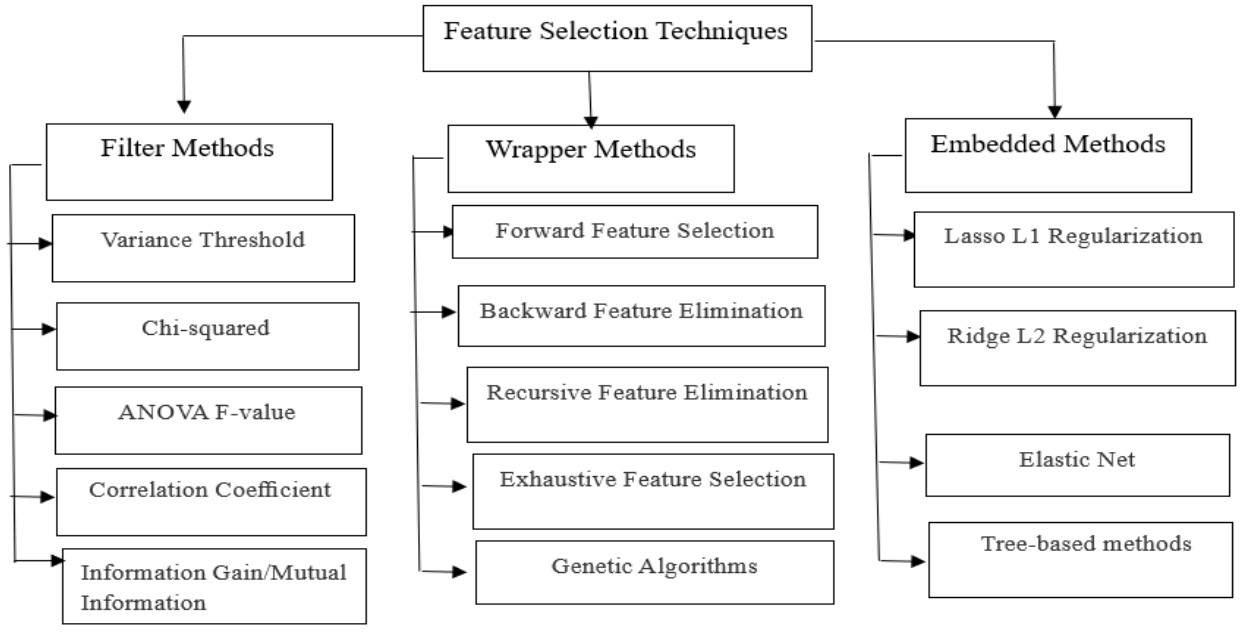

Figure 1. Classification of Feature Selection Techniques

2.1 Filter Methods

Filter methods are the easiest and quickest way to choose features. They work on their own, without using any ML model [7]. They check how relevant features are based only on their properties, including how correlated they are with the target variable or how much volatility there is in the feature itself. The preprocessing step of these methods is typically used to rank or score features, and a subset of the top-ranked features is subsequently selected for model training [8]. The predictive ability of each feature is checked independently in filter techniques. After assigning a statistical score to each feature, they can eliminate low-scoring features and retain high-scoring ones by using a threshold or a predetermined number of features. Not only is the process quick and easy to scale, but it also doesn't need building a ML model. This is particularly useful for multi-dimensional medical research datasets [9].

They are perfect for big datasets with many features because of their fast and inexpensive processing. Their ability to identify and remove similar or related properties makes them useful for purging unnecessary traits [10].

One major problem is that they can't fully convey the interplay and complexity of relationships between qualities. In the event that feature combinations outperform individual features in terms of prediction accuracy, their feature isolation evaluations

may lead to feature sets that are below par. Because feature selection is not integral to model training, it is possible that the features chosen will not provide the best results for a given ML model [11].

In the context of medical diagnosis, filter methods are frequently implemented as an initial screening phase, particularly in high-throughput data sets such as genomics, to rapidly reduce the dimensionality prior to the application of more computationally intensive methods [12].

2.1.1 Variance Threshold

One of the simplest filter approaches is the Variance Threshold method. In this method, features with variances below a specific threshold are removed. The basic premise is that characteristics that exhibit extremely low variation, meaning they nearly always have the same value across all samples, are highly unlikely to provide useful information for a predictive model. Although simple, this approach disregards the target variable and any relationships between features [13, 14].The algorithm computes the population variance σ^2 for each attribute independently, utilizing the following equation:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}$$

Where x_i refers to the values of the sample individual for a feature, μ refers to the mean of feature values, n refers to the overall number of samples.

2.1.2 Chi-squared Test

The Chi-squared statistic test is often used to choose which features to utilize in classification issues, especially when the features and the target variable are both categorical. This test shows the independence of a feature from the target variable [15]. A greater Chi-squared value means that the feature is more meaningful because it shows that the two variables are more related. If a feature has a low Chi-squared score, it is thought to be independent of the target and can be deleted [16]. The equation for Chi-Square is:

$$X^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Where O_i indicates the observed values, E_i indicates the expected values.

2.1.3 ANOVA F-value

The F-value is a statistical test used to determine if there are big differences between the means of two or more groups in an Analysis of Variance (ANOVA) [17]. In feature selection, it looks at the linear relationship between a categorical target variable and a numerical feature. A high F-value means that a feature is statistically significant in telling the different classes of the target variable apart. This method works well for input variables that are numbers and output variables that are categories [18]. The F-value is computed by taking the ratio of Mean Square Between (MSB) and Mean Square Error/Within (MSE):

$$F = \frac{MSB}{MSE}$$

2.1.4 Correlation Coefficient

Methods based on correlation look at how each attribute is related to the target variable. The Pearson correlation coefficient is a common way to find out how two continuous variables are related in a straight line [19]. Features that are closely related to the target variable are thought to be more important. Correlation can also be used to find and get rid of unnecessary features by removing one of two that are quite similar to each other. Spearman correlation is utilized for monotonic connections and nonlinearities [20]. The equation of the sample for the Correlation Coefficient *r* is:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\ \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Where x_i, y_i represents the points of the individual sample, $\bar{x}, \bar{y}$ represents the means of the sample, and n refers to the observation numbers.

2.1.5 Information Gain/Mutual Information

Information Gain (IG) and Mutual Information (MI) are ideas that come from the

study of information. Information Gain quantifies the decrease in entropy (or uncertainty) of the target variable upon acquisition of a specific feature. Mutual knowledge measures how much knowledge you get about one random variable by looking at the other. Both approaches are strong because they can find nonlinear connections between features and the target variable. This makes these tools useful for many different types of data [21, 22].

MI quantifies the reduction in uncertainty of Y given knowledge of

$$MI(X;Y) = H(Y) - H(Y/X)$$

IG computes the entropy decrease by using the following equation

$$IG(S,X) = H(S) - H(S/X)$$

where S is the entire data set (or node), and the conditional entropy is the weighted average of the subsets generated by splitting on feature X.

2.2 Wrapper Methods

Wrapper approaches check subsets of features by using a specific ML model to train and test each one. The goal of this iterative procedure is to find the feature subset that works best with the specified learning algorithm. Wrapper methods are usually more accurate than filter methods since they take into account feature interactions and the given model, but are significantly more computationally expensive [23].

The basic idea of wrapper techniques is to explore various combinations of features using a search algorithm. For each set of data, it trains an ML model and then evaluates it using various metrics, such as accuracy, F1-score, and AUC. The subgroup with the best model fit is then selected [24].

Wrapper techniques often improve prediction performance for the chosen model, but are more expensive to implement due to their iterative nature. Wrapper approaches work with the learning algorithm to look at sets of features and find complicated interactions between them.

Wrapper methods frequently surpass filter approaches in predictive accuracy due to their consideration of the interdependence and interaction of information [25].

The main problem with wrapper techniques is their high cost and inefficiency. When there are a large number of features, it can be extremely time-consuming to train and evaluate a model for each potential feature subset. The possibility of overfitting the model and dataset used for feature selection increases the likelihood of subpar generalization to new data. The feature subset that is chosen has a strong relationship with the ML model that is chosen. Different models may have different ideal subsets [26].

In medical diagnostics, wrapper approaches are frequently favoured when superior predicted accuracy is essential, and computer resources permit their application. These factors are especially useful when it is believed that the interactions between clinical or genetic elements significantly influence the manifestation of a disease [27].

2.2.1 Forward Feature Selection (FFS)

FFS is a search method that starts with no features and works its way up. It adds the feature that, when combined with the features that have previously been chosen, makes the model work better (for example, by increasing its accuracy or F1-score). This procedure goes on until no more improvements can be seen or a set number of features are reached [28].

2.2.2 Backward Feature Elimination (BFE)

BFE is the opposite of FFS. It starts with all the features and eliminates the one that hurts the model's performance the least at each step. This procedure goes on until there are no more features that can be deleted without a big drop in performance, or the right number of features is left [29].

2.2.3 Recursive Feature Elimination (RFE)

RFE repeatedly trains a model while removing unimportant features. It requires a feature-weighting estimator, similar to linear model coefficients or tree-based model feature importance [30]. Every time the model is trained, the features with the lowest

significance are removed. The technique is repeated on the smaller set until the right number of characteristics are found. RFE is strong and can work with many different ML models [31].

2.2.4 Exhaustive Feature Selection (EFS)

EFS is a brute-force method that checks all combinations of features to find the best one. It guarantees the best subset for a given model. But it is too hard to use for datasets with few features. [32].

2.2.5 Genetic Algorithms (GA)

GAs are metaheuristic search algorithms based on natural selection. Each possible subset of features is considered as a single chromosome in a larger population. The algorithm iteratively evolves this population by selection, crossover, and mutation processes [33]. The objective is to identify optimal or near-optimal feature subsets that optimize a fitness function (e.g., accuracy of a model). GAs are computationally expensive but can explore a large search space and deal with complex features [34].

2.3 Embedded Methods

Embedded approaches work by selecting features during the model generation process on their own. The system will learn to favor features that improve the accuracy of forecasts. In the training of the model, features are selected by assigning weights or coefficients to features and then removing those that do not contribute significantly [35].

Embedded approaches are computationally more efficient than wrapper strategies since feature selection is part of the model training. This is because they chose not to train the model continuously on different subsets. Furthermore, they surpass filter techniques in terms of accuracy by integrating feature interactions into the learning algorithm. Embedded approaches implicitly account for dependencies and interactions between attributes by identifying them during training. This process is wholly dependent on the learning algorithm [36]. Comparable to wrapper methods, the optimal set of features may vary depending on the model employed. The features that were chosen may not be ideal for different types of models [37].

The ability to efficiently manage high-dimensional data while still capturing complex relationships between features is the reason embedded methods are becoming more popular in medical diagnosis. For example, LASSO has been effectively implemented in genomic research to identify critical genetic markers that are linked to disease, and Random Forests have been employed to identify critical clinical variables in a variety of diagnostic tasks [38].

2.3.1 Lasso (L1 Regularization)

Lasso (Least Absolute Shrinkage and Selection Operator) is a linear model that uses L1 regularization. It adds a penalty to the loss function that is equal to the absolute value of the size of the coefficients. This penalty makes the model sparser, which means it can shrink some feature coefficients all the way to zero, which is like automatic feature selection. Lasso is very useful when dealing with datasets with a lot of features that are not useful [39]. The typical equation for the Lasso algorithm comprises the Mean Squared Error (MSE) in conjunction with the L1 penalty term:

$$L(\beta) = \frac{1}{2n}\sum_{i=1}^{n} y_i - \hat{y}_i + \lambda \sum_{j=1}^{p} |\beta_j|$$

Where $L(\beta)$ refers to the objective function, y_i indicate the value of the actual target, $\hat{y}_i$ indicate the predicate value, β_j refers to coefficients allocated to the j feature, the penalty's strength is determined by λ, and n refers to the number of training samples.

2.3.2 Ridge (L2 Regularization)

Ridge regression adds a penalty term proportional to the squared size of the coefficients to the loss function. This is called L2 regularization. It tends to push coefficients towards zero, but not necessarily to zero. The primary purpose of the Ridge regression is to avoid overfitting and multicollinearity by shrinking the less significant features. It does not perform explicit feature selection, such as

Lasso [40]. The cost function for Ridge Regression is described as:

$$J(\theta) = \frac{1}{2n}(X\theta - y)^T(X\theta - y) + \frac{\lambda}{2}\theta^T\theta$$

Where X indicates the array of input features, y indicates the target values, θ refer to the coefficients feature, λ is the regularization hyperparameter that determines the balance between fitting the data and decreasing the weights, n indicates the number of training examples, and m refers to features number.

2.3.3 Elastic Net

Elastic Net is a regularization method that combines the L1 (Lasso) and L2 (Ridge) penalties. This is particularly helpful when there are groups of features that are very close together. Lasso generally selects a single feature from a group and ignores the others. Ridge keeps all of them. ItType equation here. can select groups of correlated attributes at the same time. It is a trade-off between the sparsity and the group effect. It requires tuning two regularization parameters [41]. This algorithm aims to minimise a cost function that is a combination of the normal Residual Sum of Squares (RSS) plus a penalty term that reflects the L1 and L2 norms:

$$J(\beta) = RSS + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=1}^{p} \beta_j^2$$

Where $J(\beta)$ is the overall cost function, RSS represents the residual sum of squares, β_j refer to the feature coefficients, $\lambda_1$and λ_2 refer to the tuning parameters that regulate the intensity of the L1 and L2 penalties correspondingly

2.3.4 Tree-based Methods

The system is trained with a variety of tree-based ML approaches, including Decision Trees, Random Forests, and Gradient Boosting Machines (e.g., XGBoost and LightGBM). The result of these algorithms is a set of rules for decision-making. The rules are based on factors that best partition the data and produce the fewest errors. The importance of a feature is measured by the degree of reduction of the impurity in all the trees in the group [42]. Those features that are often used to split or lead to a large impurity reduction are the more relevant ones. These methods can capture intricate non-linear interactions and are robust to outliers [43].

3. Comparison of Feature Selection Techniques

Table 1 presents a comprehensive overview of the essential characteristics, benefits, and drawbacks of the presented feature selection strategies.

Table 1: Comparison analysis of Feature Selection Techniques.

Category	Algorithm	Advantages	Disadvantages
Filter Methods	Variance Threshold	-Computationally efficient -Simple to implement -Independent of the learning algorithm	-Ignores feature dependencies -May select redundant features -Does not optimize for model performance
	Chi-squared (χ^2)	-Effective for categorical features - Computationally inexpensive	-Only considers individual feature relevance -Sensitive to data sparsity -Not suitable for numerical features
	ANOVA F-value	-Effective for numerical features and categorical targets -Identifies statistically significant features	-Assumes normality and equal variances -Only considers individual feature relevance
	Correlation Coefficient (e.g., Pearson, Spearman)	-Simple to understand and implement -Identifies linear/monotonic relationships	-Sensitive to outliers -Only captures pairwise relationships
	Information Gain/Mutual Information	- Captures non-linear relationships -Effective for both numerical and categorical data	-Computationally more intensive than simpler filter methods -Can be biased towards features with more values

Wrapper Methods	Forward Feature Selection	- Considers feature interactions - Often leads to better model performance	-Computationally expensive -Prone to local optima - Can be slow for high-dimensional data
	Backward Feature Elimination	-Considers feature interactions -Often leads to better model performance	-Computationally expensive -Prone to local optima -Can be slow for high-dimensional data
	Recursive Feature Elimination (RFE)	-Considers feature interactions -Works well with various models -Provides feature ranking	-Computationally intensive -Performance depends on the chosen estimator -Sensitive to noisy features
	Exhaustive Feature Selection (EFS)	-Suitable for big databases -Works well with various models	-Not suitable for small datasets -Can be complex to implement
	Genetic Algorithms (GA)	-Can explore a large search space -Finds optimal or near-optimal feature subsets -Handles complex interactions	-Very computationally expensive -Requires careful parameter tuning -Results can vary between runs.
Embedded Methods	Lasso (L1 Regularization)	-Performs feature selection and regularization simultaneously -Handles multicollinearity	-Can arbitrarily select one feature from a group of highly correlated features -Not suitable for all model types
	Ridge (L2 Regularization)	-Prevents overfitting -Reduces coefficient magnitudes	-Does not perform true feature selection (coefficients shrink but rarely become zero) -Less interpretable than Lasso for feature selection
	Elastic Net	-Combines L1 and L2 regularization -Handles correlated features better than Lasso -Prevents overfitting	-Requires tuning of two regularization parameters -Can be more complex to implement
	Tree-based methods (e.g., Random Forest, Gradient Boosting)	-Naturally perform feature selection -Capture non-linear relationships -Robust to outliers	-Feature importance scores can be biased towards high-cardinality features-Less interpretable than simpler models -Less interpretable than simpler models

4. Results and Discussion

When working with sensitive and complex medical data, it's very important to use the correct feature selection approach. The three methods discussed in this paper were filter, wrapper, and embedding. They all work in different ways and have their own advantages and drawbacks. This section will discuss the need for reliable, clear, and relevant models, as well as the results achieved by using these methods on medical datasets.

The efficacy of ML models in medical diagnosis is significantly affected by the feature selection methodology employed. The effects are complicated and affect the models' ability to be understood, their speed of computation, their ability to be used in different situations, and their ability to make accurate predictions. The effectiveness of a feature selection technique is affected by the medical dataset's characteristics, the kind of disease being diagnosed, and the specific ML algorithm being used [44].

Impact on prediction accuracy: Studies have shown that the right feature selection can greatly improve prediction accuracy measures such as sensitivity, specificity, recall, F1-score, and AUC. Models can focus on what is important, i.e., to remove unnecessary and duplicate attributes, and in this way improve the classification accuracy and establish more exact decision limits. For example, in genomic data, there are many features (e.g., SNPs) but few samples. So, the feature selection is very important. This is done to prevent overfitting and to obtain accurate predictions [4]. Wrapper and embedding methods are usually more accurate than filter methods, as they consider the interaction between features. This becomes very clear in cases where complex relationships between diagnostically significant elements

exist. However, wrapper methods have higher computational costs, which might be problematic for very large datasets [45].

Impact on the model's explainability: The importance of feature selection for better predictive accuracy and interpretability of the models (the latter being of utmost importance in the therapeutic setting) has been highlighted. Feature selection can help physicians and researchers find the most important biomarkers, clinical data, or genetic factors related to a disease. This is because feature selection is useful in selecting a subset of features that are highly relevant. With this knowledge, researchers would be able to better understand diseases, develop new treatments, and make judgments on strong evidence. For example, LASSO regularization causes the most important predictors to pop out by shrinking the coefficients of the unimportant features to zero. It scores features based on how much they contribute to the model's performance, similar to Random Forest Importance, and it gives a clear picture of the relative importance of each feature. More complex models are more accurate, but less complex models with less features can sometimes give more useful information in terms of interpretability [46].

Impact on computing efficiency: Training may require additional time and memory because of the challenges associated with managing high-dimensional medical datasets. Feature selection instantly lowers the number of dimensions in the data, which fixes these problems. Filter methods are great for initial data reduction because they don't need a lot of computing power. This is especially true when working with very large datasets or doing exploratory analysis. They may rapidly take away a few things, which makes it easier to look at them again later. Even though they cost more to compute than filters, wrappers and embedding methods nonetheless make training models on the complete, unselected feature set far more efficient. Because ML models have fewer dimensions, they can be trained and predicted faster and use less memory. Because of this, it is easier to use ML models in real-life clinical settings [47].

Impact on overfitting: Overfitting occurs frequently in machine learning, particularly when models are trained on data with an excessive number of dimensions and insufficient sample size, which in turn impacts generalizability.

One approach to improving a model's stability is to eliminate extraneous data and noise. Because of this, the model will not fit the data too closely. Feature selection helps build models that perform better with novel, unlabeled patient data by zeroing in on the true signals in the data. For medical applications, this is crucial, as models must apply to a large population. Rigid cross-validation methods are essential for determining whether feature selection improves generalizability [48].

Considerations while dealing with health records: Problems such as class imbalance, missing data, and complicated non-linear relationships are more common in medical datasets. These characteristics should be thought about when deciding how to pick features. When dealing with outliers or missing data, some filter approaches may not work, and when dealing with unbalanced data, some wrapper methods may not work either. Improving diagnostic efficacy and handling the complexities of medical data (e.g., using a filter for initial reduction and a container for refining) are two goals of the present research into hybrid techniques [49] .

In light of the unique diagnostic task at hand and the data at hand, the primary objective is to strike a compromise between the predicted degree of accuracy, the degree to which it is understandable, and the degree to which it is computationally viable [50].

Medical datasets often have an imbalanced class distribution (e.g. the number of health controls is large compared to the number of disease cases) which can strongly influence the choice of features. Some of the techniques are biased towards the majority class, which may result in the selection of features that do not discriminate well against the minority class. This has to be considered in the evaluation phase. This can be done using performance metrics like F1-score or AUC-ROC, which are less sensitive to class imbalance. Also, use

feature selection along with resampling techniques.

In conclusion, there is usually no definitive answer to feature selection in medical datasets. In this process, properties of the datasets (e.g., dimensionality, sample size, data types, class balance), analysis goals (e.g., prediction accuracy vs. interpretability), and available computational resources are often taken into careful consideration. There is a growing body of work on hybrid approaches. These take advantage of portions of other categories (e.g., an initial filter step with a wrapper or embedding method). The idea is to exploit the good things of different techniques and minimize the bad things. The progress in the collection of medical data and the changing environment of feature selection algorithms are unlocking the potential for more advanced and efficient tools for precision medicine and healthcare analytics.

5. Conclusions

Feature selection plays a vital role in the ML pipeline, particularly in the challenging and high-risk area of medical data analytics. We perform a systematic review on feature selection strategies, classifying them into filters, wrappers, and embedded methods; discussing the operating principles, advantages, and disadvantages of the most important algorithms within each category; and providing a basic understanding of how to use them.

The analysis in this paper demonstrates that there is no one best feature selection method; the selection of a method depends on the characteristics of the medical dataset and the goals of the analytical task. Filter methods are computationally efficient and are particularly useful to reduce dimensionality in high-dimensional cases.

Wrapper methods are computationally expensive but often lead to better predictions by choosing the best feature subsets for a learning algorithm. They shine when it really counts to get things right. Embedded methods, which incorporate feature selection directly into the process of training the model, offer a good balance between the cost of processing and the power of prediction. This is a realistic way to get you to the middle ground.

Feature selection in medical datasets is a challenging task that requires expertise due to the challenges of high dimensionality, class imbalance, small sample sizes, and the requirement for interpretability. The interpretability of some features is especially relevant in therapeutic settings, since it can lead to a better understanding, validation of results, and discovery of new biomarkers or etiological agents of diseases. Future research in this field will probably be mainly devoted to the development of hybrid methods, combining properties of different kinds. This means methods will be more robust and flexible, able to handle the subtleties and complexities of real-world medical data. Furthermore, the combination of AI and feature selection will keep on gaining more strength. The objectives are to identify significant characteristics and justify their selection.

References

[1] Rane, J., R.A. Chaudhari, and N.L. Rane, Data Analysis and Information Processing Frameworks for Ethical Artificial Intelligence Implementation: Machine-Learning Algorithm Validation in Clinical Research Settings. Ethical Considerations and Bias Detection in Artificial Intelligence/Machine Learning Applications, 2025: p. 192.

[2] Cheng, X., A comprehensive study of feature selection techniques in machine learning models. Available at SSRN 5154947, 2024.

[3] Hancox, Z., et al., A systematic review of networks for prognostic prediction of health outcomes and diagnostic prediction of health conditions within Electronic Health Records. Artificial Intelligence in Medicine, 2024. 158: p. 102999.

[4] Büyükkeçeci, M. and M.C. Okur, A comprehensive review of feature selection and feature selection stability in machine learning. Gazi University Journal of Science, 2023. 36(4): p. 1506-1520.

[5] Naheed, N., et al., Importance of features selection, attributes selection, challenges and future directions for medical imaging data: a review. Computer Modeling in Engineering & Sciences, 2020. 125(1): p. 314-344.

[6] Bashir, S., et al., A novel feature selection method for classification of medical data using filters, wrappers, and embedded approaches. Complexity, 2022. 2022(1): p. 8190814.

[7] Chen, C.W., et al., Ensemble feature selection in medical datasets: Combining filter, wrapper, and embedded feature selection results. expert systems, 2020. 37(5): p. e12553.
[8] Mallidi, S.K.R. and R.R. Ramisetty, Optimizing intrusion detection for IoT: a systematic review of machine learning and deep learning approaches with feature selection and data balancing. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2025. 15(2): p. e70008.
[9] Sadeghian, Z., et al., A review of feature selection methods based on meta-heuristic algorithms. Journal of Experimental & Theoretical Artificial Intelligence, 2025. 37(1): p. 1-51.
[10] Cherrington, M., et al. Feature selection: filter methods performance challenges. in 2019 International Conference on Computer and Information Sciences (ICCIS). 2019. IEEE.
[11] Sosa-Cabrera, G., et al., Feature selection: A perspective on inter-attribute cooperation. International Journal of Data Science and Analytics, 2024. 17(2): p. 139-151.
[12] Theng, D. and K.K. Bhoyar, Feature selection techniques for machine learning: a survey of more than two decades of research. Knowledge and Information Systems, 2024. 66(3): p. 1575-1637.
[13] Natarajan, K., D. Baskaran, and S. Kamalanathan, An adaptive ensemble feature selection technique for model-agnostic diabetes prediction. Scientific Reports, 2025. 15(1): p. 6907.
[14] Fida, M.A.F.A., T. Ahmad, and M. Ntahobari. Variance threshold as early screening to Boruta feature selection for intrusion detection system. in 2021 13th International Conference on Information & Communication Technology and System (ICTS). 2021. IEEE.
[15] Abdo, A., R. Mostafa, and L. Abdel-Hamid, An optimized hybrid approach for feature selection based on chi-square and particle swarm optimization algorithms. Data, 2024. 9(2): p. 20.
[16] Haghighat, H., Machine Learning Techniques and Chi-square Feature Selection for Diagnostic Classification Model of Autism Spectrum Disorder Using fMRI Data.
[17] Nasiri, H. and S.A. Alavi, A Novel Framework Based on Deep Learning and ANOVA Feature Selection Method for Diagnosis of COVID-19 Cases from Chest X-Ray Images. Computational intelligence and neuroscience, 2022. 2022(1): p. 4694567.
[18] Ahmad, B., J. Chen, and H. Chen, Feature selection strategies for optimized heart disease diagnosis using ML and DL models. arXiv preprint arXiv:2503.16577, 2025.
[19] Zhou, H., X. Wang, and R. Zhu, Feature selection based on mutual information with correlation coefficient. Applied intelligence, 2022. 52(5): p. 5457-5474.
[20] Jiang, J., X. Zhang, and Z. Yuan, Feature selection for classification with Spearman's rank correlation coefficient-based self-information in divergence-based fuzzy rough sets. Expert Systems with Applications, 2024. 249: p. 123633.
[21] Efe, Y. and L. Demir, The impact of feature selection models on the accuracy of tree-based classification algorithms: Heart disease case. Procedia Computer Science, 2025. 253: p. 757-764.
[22] Rehman, M., R. Kalakoti, and H. Bahşi. Comprehensive feature selection for machine learning-based intrusion detection in healthcare IoMT networks. in 11th International Conference on Information Systems Security and Privacy. 2025. SCITEPRESS-Science and Technology Publications.
[23] Behera, S.R., B. Pati, and S. Parida, A Meta-heuristic Hybrid Wrapper Method based on Feature Selection for Classification of Biological Samples. Computación y Sistemas, 2025. 29(2).
[24] Baranauskas, J.A. and M.C. Monard, Experimental feature selection using the wrapper approach. WIT Transactions on Information and Communication Technologies, 2025. 22.
[25] Liyew, C.M., et al., A review of feature selection methods for actual evapotranspiration prediction. Artificial Intelligence Review, 2025. 58(10): p. 292.
[26] Ali, M.Z., et al., Advances and challenges in feature selection methods: a comprehensive review. J. Artif. Intell. Metaheuristics, 2024. 7(1): p. 67-77.
[27] Shakir, H.M., A.K. Oleiwi, and H.A. Mejbel Al-Madhee, A New Structure based on Filter-Wrapper Feature Selection and Optimized Hybrid Classification for Coronary Artery Disease Diagnosis. International Journal of Intelligent Engineering & Systems, 2025. 18(7).
[28] Okwuosa, C.N. and J.-W. Hur. Enhancing Induction Motor Reliability Through Advanced Feature Selection and Diagnostic Models in Low-Load Conditions. in 2025 International Conference on Artificial Intelligence in Information and Communication (ICAIIC). 2025. IEEE.
[29] Soladoye, A.A., et al., Enhancing Alzheimer's Disease Prediction Using Random Forest: A Novel Framework Combining Backward Feature Elimination and Ant Colony Optimization. Current Research in Translational Medicine, 2025: p. 103526.

[30] Bulut, O., et al., Benchmarking Variants of Recursive Feature Elimination: Insights from Predictive Tasks in Education and Healthcare. Information, 2025. 16(6): p. 476.
[31] Priyatno, A.M. and T. Widiyaningtyas, A systematic literature review: recursive feature elimination algorithms. JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer), 2024. 9(2): p. 196-207.
[32] Saad, A.H., N.A.A. Hamzah, and W.M.D.W. Zaki. Exhaustive Feature Selection Using Wrapper method for Artery-Vein Classification in Retinal Fundus Image. in 2025 21st IEEE International Colloquium on Signal Processing & Its Applications (CSPA). 2025. IEEE.
[33] Ali, W. and F. Saeed, Hybrid filter and genetic algorithm-based feature selection for improving cancer classification in high-dimensional microarray data. Processes, 2023. 11(2): p. 562.
[34] Fang, Y., et al., A feature selection based on genetic algorithm for intrusion detection of industrial control systems. Computers & Security, 2024. 139: p. 103675.
[35] Carrasco, M., et al., Embedded feature selection for robust probability learning machines. Pattern Recognition, 2025. 159: p. 111157.
[36] Lei, C., et al., Comparisons of filter, wrapper, and embedded feature selection for rockfall susceptibility prediction and mapping. Natural Hazards, 2025. 121(2): p. 1911-1943.
[37] Vallabhaneni, P., et al., Comparative Study of Feature Selection Algorithms for Heart Disease Prediction, in Real-World Applications and Implementations of IoT. 2025, Springer. p. 231-244.
[38] Royhan, W. and A. Amalia. Feature Selection Using Ensemble Lasso Regression, Random Forest and Recursive Feature Elimination Methods in Breast Cancer Classification. in 2025 International Conference on Computer Sciences, Engineering, and Technology Innovation (ICoCSETI). 2025. IEEE.
[39] Awasthi, N. and P.R. Gautam, Android ransomware network traffic detection using decision tree and L1 LASSO regularization feature selection, in Intelligent Computing and Communication Techniques. 2025, CRC Press. p. 729-737.
[40] Hamada, M., et al. Evaluation of Recursive Feature Elimination and LASSO Regularization-based optimized feature selection approaches for cervical cancer prediction. in 2021 IEEE 14th International symposium on embedded multicore/many-core systems-on-chip (MCSoC). 2021. IEEE.
[41] Amini, F. and G. Hu, A two-layer feature selection method using Genetic Algorithm and Elastic Net. Expert Systems with Applications, 2021. 166: p. 114072.
[42] Demir, S. and E.K. Sahin, An investigation of feature selection methods for soil liquefaction prediction based on tree-based ensemble algorithms using AdaBoost, gradient boosting, and XGBoost. Neural Computing and Applications, 2023. 35(4): p. 3173-3190.
[43] Yıldız, A.Y. and A. Kalayci, Gradient boosting decision trees on medical diagnosis over tabular data. arXiv preprint arXiv:2410.03705, 2024.
[44] Noroozi, Z., A. Orooji, and L. Erfannia, Analyzing the impact of feature selection methods on machine learning algorithms for heart disease prediction. Scientific reports, 2023. 13(1): p. 22588.
[45] Ahmad, H.F., et al., Investigating health-related features and their impact on the prediction of diabetes using machine learning. Applied Sciences, 2021. 11(3): p. 1173.
[46] Islam, M.A., et al., Precision healthcare: A deep dive into machine learning algorithms and feature selection strategies for accurate heart disease prediction. Computers in Biology and Medicine, 2024. 176: p. 108432.
[47] Pathan, M.S., et al., Analyzing the impact of feature selection on the accuracy of heart disease prediction. Healthcare Analytics, 2022. 2: p. 100060.
[48] Olawade, D.B., et al., Comparative analysis of machine learning models for coronary artery disease prediction with optimized feature selection. International Journal of Cardiology, 2025: p. 133443.
[49] Oreski, D., S. Oreski, and B. Klicek, Effects of dataset characteristics on the performance of feature selection techniques. Applied Soft Computing, 2017. 52: p. 109-119.
[50] Shilaskar, S. and A. Ghatol, Feature selection for medical diagnosis: Evaluation for cardiovascular diseases. Expert systems with applications, 2013. 40(10): p. 4146-4153.