Al-Rafidain Journal of Engineering Sciences Vol. 3, Issue 2, 2025:343-359

Al-Rafidain Journal of Engineering Sciences

Journal homepage https://rjes.iq/index.php/rjes

ISSN 3005-3153 (Online)

A Review of Feature Selection Techniques for Cancer Diagnosis and Prediction with Machine Learning and Deep Learning: Current Trends and Future Directions

Aythem Khairi Kareem[1], Manar AL-Mahdawi[2] and Ahmed Adil Nafea[3]

[1]Department of Heet Education, General Directorate of Education in Anbar, Ministry of Education, Heet, 31007 Anbar, Iraq.
[2] Department of Medical Physics, College of Science, AL-Nahrain University, Jadriya, Baghdad, 10072, Iraq.
[3] Department of Artificial Intelligence, College of Computer Science and IT, University of Anbar, Ramadi, Iraq.

ARTICLE INFO

Article history:

Received	09 August 2025
Revised	10 August 2025
Accepted	28 August 2025
Available online	15 September 2025

Keywords:
Cancer Diagnosis
Clinical Decision Support System
Deep Learning
Feature Selection
Machine Learning

ABSTRACT

In the past, cancer diagnosis (CD) and prognosis were based largely on clinical intuition and empirical reasoning whereas now data-driven approaches, especially deep learning (DL) and machine-learning (ML) models have paved the way to predict both disease progression and likely treatment, contingent on their risk or benefit to the individual patient. Nevertheless, the high-dimensionality and heterogeneity of biomedical data suffer from potential overfitting, computational complexity, and reduced model interpretability. Feature selection (FS) techniques have gained increasing attention in also resolving these arisen problems by selecting relevant and informative attributes from a large volume of data. This study is to offer a systematic review of conventional and recent existing nature-inspired FS techniques that are classified in filter, wrapper, embedded, hybrid and ensemble techniques, and to discuss their applications in cancer-related research. We contrast strengths and weaknesses of each technique and discuss which are appropriate for high-dimensional data. Moreover, the paper summary recent works with different cancer types (including breast, lung, prostate, ovarian, and colorectal) to demonstrate how prediction performance is affected by different FS techniques and classifiers. The review also discusses current limitations such as small sample size, noisy or incomplete data, scalability and reproducibility. Finally, we consider future directions focusing on the importance of biologically-inspired FS strategies, appropriate benchmarking approaches, and the design of robust, scalable and interpretable models for clinical application.

1. Introduction

Cancer Disease (CD) stays one of the world’s major general health hazards, given its high mortality and morbidity rates [1]. Different factors influence the development of CD, containing intrinsic factors and extrinsic factors. Numerous patients encounter metastases and recurrences during or after the medical process in spite of the significant advancements made in cancer medicines, such as immunotherapy, targeted therapy, radiotherapy, chemotherapy, and surgery [2]. Metastasis and recurrence result in treatment failure, underscoring the issues posed via treatment resistance and exhibiting the requirement to improve the effectiveness of cancer treatments [3]. The money rate after treatment is closely connected to the duration of CD development. Relevant analyses have demonstrated that patients' survival rates increase with earlier CD detection [4]. CD diagnosis is now being simplified. Usually, ultrasound, magnetic resonance imaging, computed tomography, and additional imaging tests are completed first, then the doctor

Corresponding author E-mail address: manar.thaer@nahrainuniv.edu.iq

https://doi.org/10.61268/q78nda33

delivers a primary clinical diagnosis. CD diagnosis strategy is primarily based on doctors’ subjective judgment; therefore, the diagnostic effects are mainly limited by the doctors’ experience and skill. Inexperienced doctors are inclined to misdiagnose, and even doctors with long-term applicable experience have problems making entirely accurate and consistent judgments when encountering different complicated diseases. Most of the practical information in the image is concealed in tiny gray differences, which are challenging for humans to determine [5].

Developing a robust and accurate approach is vital for predicting CD. Multifarious ML techniques have been invented for clinical applications, including Decision Tree (DT), Support Vector Machines, Naive Bayesian, Random Forest (RF), ensemble models, and Neural Networks (NNs). Using ML techniques, different clinical characteristics of cancer patients can be fused with their survival rates [6]. ML techniques also have the advantages of decreasing medical practitioners' workloads and the risk of human mistakes. The increased ML performance has made it a motivating and exciting mechanism for healthcare providers [7].

Currently, with the beginning of developed technologies, a significant amount of high-dimensional dataset has been delivered in medicine, specifically in fields associated with cancer care and treatment strategies. The high dimensions of the dataset and quality-related issues, such as misleading features, useless, duplicate, missing, and irrelevant data, make it more challenging to achieve discernment from data. In addition, if there are numerous variables and the sample size is small, problems connected to overfitting may occur. High-dimensional data reduces the model's efficiency. A straightforward forecast model with optimized features achieves high performance, corresponding to a full-featured model. Thus, high-dimensional and raw data should be preprocessed to make them more suitable for additional analysis. This technique is called FS [8].

The main objectives of this review paper are to supply a comprehensive overview and vital analysis of FS techniques used in cancer prediction through ML techniques. As cancer-related datasets often have high dimensional data, such as gene expression profiles and other omics characteristics, selecting the more relevant and informative attributes is important for building interpretable, efficient, and accurate predictive models. This review aims to classify and evaluate various FS techniques, including filter, wrapper, embedded, and hybrid techniques, based on their application in cancer research. This review highlights the performances, limitations, and advantages in various anal; we aspire to guide researchers in determining proper techniques for cancer prediction studies and to determine gaps and future trends for enhancing FS strategies in this domain.

2. Feature Selection

Datasets typically contain redundant and irrelevant features that adversely impact the ML process. Thus, features that contribute to the ML process should be detected utilizing FS techniques [9]. Utilizing assistance from domain expertise, common sense, and domain knowledge may allow for the detection and elimination of redundant features. Nevertheless, a more methodical approach is required for high dimensional data sets. Since irrelevant and redundant features have no significant influence on ML, tossing them from the learning strategy will allow for creating simple models, bypass the dimensionality curse, decrease overfitting, and improve learning speed [8].

2.1 Feature Selection Techniques

The FS techniques can be split into three methodological types, including wrappers, filters, and embedded techniques [10]. These techniques differ in terms of 1) the FS characteristic being integrated or separated as a learning algorithm part, 2) evaluation criteria, 3) computational complexities, and 4) the possibility of detecting interactions and redundancies between features. Table 1 provides a more exhaustive outline of the disadvantages and advantages of the three FS techniques [11][12].

Table 1: Exhaustive overview of the disadvantages and advantages of the FS techniques

FS techniques	Advantages	Disadvantage
Filters	- Good generalizability. -Appropriate for high dimensional data. -Fast execution. -Independent of learning model.	- Fail to manage the redundancy issue. - Interactions between features are ignored.
Wrappers	- Identify feature interactions of higher order. -Take into account the interaction between features. -Better performance attainability.	-The learning algorithm is created from scratch for each subset. -Prone to over-fit. -Very expensive in terms of execution times
Embedded	-Identify feature dependencies. -Take into account the interaction between features. -Accurate. -Faster than wrappers.	-Classifier-dependent selection. -Learning algorithm-specific.

In addition to these three main techniques, there are also hybrid, and ensemble approaches that combine two or more of these techniques to leverage their complementary strengths and improve FS performance. Figure 1 displays the types of the FS techniques.

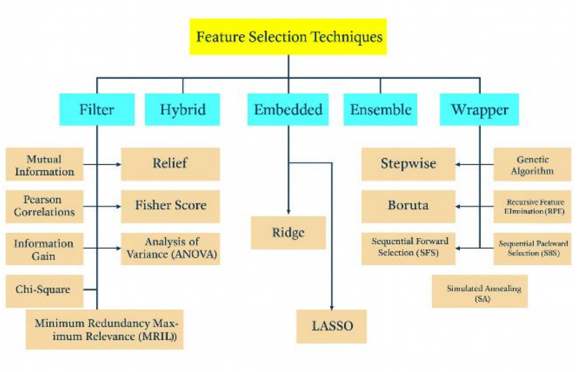

Figure 1. Types of FS techniques.

FS techniques can be described in detail as follows:

2.1.1 Filter Techniques

Filters evaluate feature relevance depend on the main characteristics of the data. The filtering process includes two key steps. To generate the classification model, features are first ranked separately based on a special criterion measurement such as entropy, distance, and Pearson correlation. Second, it determines the best-ranked features employing a threshold value. The rest of the features are considered to be uninformative and unnecessary. Filter techniques are fast, which makes them more appropriate for high-dimensional data. Since the associations between the independent features are not deemed, redundant features may be permitted to be selected. The following sub-sections supply an in-depth analysis of five well-known filter techniques [13].

2.1.1.1 Mutual Information (MI)

MI technique indicates how much data is swapped between the class label and feature. It is an estimate of the influence of one random variable on another [14]. The attribute with the most communicated information within the category target is the most suitable, since it may actually determine associates of one category from those of another. It is mathematically expressed in Equation 1 as follows [11]:

$$L(X,Y) = \sum_{j=1}^{n}\sum_{i=1}^{n} q\left(X_{(j)},Y_{(i)}\right).\,log\left(\frac{q\left(X_{(j)},Y_{(i)}\right)}{q\left(X_{(j)}\right).q\left(Y_{(i)}\right)}\right) \ldots (1)$$

Where, $L(X,Y)$ is the probability of X and Y, $q\left(X_{(j)}\right)$ and $q\left(Y_{(i)}\right)$ are the marginal probability of X and Y, $log\left(\frac{q\left(X_{(j)},Y_{(i)}\right)}{q\left(X_{(j)}\right).q\left(Y_{(i)}\right)}\right)$ measures how much the actual joint distribution deviates. MI is zero when Y and X are statistically independent $\left(q\left(X_{(j)},Y_{(i)}\right) = q\left(X_{(j)}\right).q\left(Y_{(i)}\right)\right)$.

2.1.1.2 Relief

Relief technique is an instance-based technique. It attempts to encounter the nearest neighbor from the same category (nearest hit) and the opposing category for each training sample (nearest miss) [15]. The contrast in each attribute's score is the difference between them. The Relief quantifies individual features with a score R_i, which can be utilized to rank features. This technique can be positively sufficient for high-dimensional data since the weighting R_i of each attribute is iteratively altered for each established case. If the feature weight is from the feature in nearby samples of the exact category instead of nearby samples of the other category, the feature weight reduces; otherwise, it increases. The formula of feature's weight is computed in Equation 2 [11]:

$$R_i = R_{i-}(x_i - nearHit_i)^2 + (x_i - nearMiss_i)^2 \dots (2)$$

The technique's limitations include the fact that it is only applicable to binary classification issues and does not address redundancy. A comprehensive interpretation of Relief named Relief-A is presented for cracking the missing data problem in order to crack the earlier represented issues with Relief. Relief-F is executed to solve multi-class problems.

2.1.1.3 Pearson correlation coefficients (PCC)

Feature selection algorithms Among the filter-based feature selection techniques, PCC is popularly used. It is a parametric technique proposed by Karl Pearson, to examine the degree of the linear association between two variables [16]. PCC works very well with data that are normally distributed. It outputs a CC value with values between -1 and 1, where a value of 1 represents a perfect relationship. A positive correlation coefficient (CC) indicates direct relationship (gain in one variable corresponds to a gain in the other as well), while a negative correlation coefficient shows the inverse relationship, where one of them grows as the other diminishes. A CC equal to zero indicates no linear relationship among the variables. The relationship between two variables is calculated through the correlation coefficient, r and t, respectively by the formula in Equation 3 [17]:

$$C = \frac{\sum_{j=1}^{m}(r_j - \bar{r})(t_j - \bar{t})}{\sqrt{\sum_{j=1}^{m}(r_j - \bar{r})^2(t_j - \bar{t})^2}} \dots (3)$$

where r and t represent the mean of the values of the r and t variable, r_j and t_j represent the value of the r and z variable in an instance.

2.1.1.4 Fisher score (F-score)

The F-score is a well-known supervised FS technique. It considers each feature individually according to the Fisher criterion and ranks them to construct a potentially suboptimal subset [18]. his technique can be used for separation of two real data sets. In the best-case scenario, good descriptors generate same feature for samples of the same class and different feature for samples that have different classes. The F-score for a feature i is obtained via Equation 4 which allude to the count of positive and negative instances [17].

$$F(i) = \frac{(xa_i^+ - xa_i)^2 + (xa_i^- - xa_i)^2}{\frac{1}{nplus - 1}\sum_{j=1}^{n+}\left(x_{ji}^+ - xa_i^+\right)^2 + \frac{1}{mneg - 1}\sum_{j=1}^{n-}\left(x_{ji}^- - xa_i^-\right)^2} \dots (4)$$

Where, xa_i, xa_i^+ and xa_i^- are the average of j^{th} feature of the whole, positive and negative instance and x_{ji}^+ and x_{ji}^- are the j^{th} feature of the i^{th} positive and negative instance respectively.

2.1.1.5 Information Gain (IG)

The IG filter-based FS technique is mainly employed to gauge the significance of features in classifying issues in DT models. It can likewise be extensively employed in other ML models. IG is based on the idea of entropy from information theory, describing the uncertainty or impurity in a group of instances. Mathematically, the IG between two variables A and B is developed as the difference between the initial entropy of A and the entropy of A after following B. It can be represented Equation 5 as [19]:

$$IG(A|B) = H(A) - H(A|B) \dots (5)$$

Where H(A) is the entropy for variable A and H(A|B) describes the dependent entropy for A given B. To calculate the IG value for a feature, calculate the entropy of the target variable for the total dataset and remove the conditional entropies for every possible value of that feature. Also, the entropy H(A) and dependent entropy H(A|B) are calculated from Equations 6 and 7 [17]:

$$H(A) = -\sum_{a\in A} P(a)\log_2(a) \ldots (6)$$

$$H(A|B)$$
$$= -\sum_{a\in A} P(a)\sum_{b\in B} P(a|b)\log_2(P(a|b) \ldots (7)$$

Therefore, when assessing two variables A and C, a variable B is considered to have a more robust correlation with A than with C if $G(A|B)$ > $IG(C|B)$. Moreover, IG estimates each feature independently and considers its relevance to the target variable.

2.1.1.6 Analysis of Variance (ANOVA)

ANOVA is a grouping of parametric statistical measures and their analysis techniques that specifies if the means of two or more instances of data are drawn from the same distribution [20]. The ANOVA technique is a F-statistic used here as an ANOVA F-test. It is a univariate statistical test where each attribute corresponds to the target attribute to determine whether there is any statistically important association between them. Mainly, ANOVA is employed in such classification issues where the class of input features is numerical, and the target feature is categorical. The Equation 8 can describe the mathematical formula for the ANOVA technique [21].

$$ANOVA_{F\_Value} = \frac{VBG}{VWG} \ldots (8)$$

Where VBG is variance between groups, VWG is variance within groups, which calculated from Equations 9 and 10.

$$VBG = \frac{\sum_{j=1}^{i} j_i(\bar{L}_j - \bar{L})^2}{(L-1)} \ldots (9)$$

$$VWG = \frac{\sum_{j=1}^{L}\sum_{p=1}^{i_j}(L_{jp} - \bar{L}_j)^2}{(M-L)} \ldots (10)$$

Where M is the entire instance size, L is the number of i_j groups, is the number of observations in the i^{th} group, $\bar{L}_j$ is the i^{th} group instance mean, $\bar{L}$ is the entire mean of the data, p is the p^{th} observation in the j^{th} out of L groups.

2.1.1.7 Chi-Square

The chi-square test is employed to estimate the independence between the target variable and categorical features. It assesses how observed data distribution differs from predicted commonnesses under the independence assumption. Attributes with higher chi-square scores are considered more appropriate. This standard is generally used in classification issues where the feature and target variables are categorical [22]. Chi2 is a statistical test that assesses the level of relationship between a target class yj and feature ai by approximating the observed Oi and predicted Ei commonnesses. It is mathematically represented in Equation 11.

$$X_v^2 = \sum_{j=1}^{L}\left[\frac{(E_j - Q_j)^2}{O_j}\right] \ldots (11)$$

Where X is the Chi-square, v is the degrees of freedom, E describes the observed commonnesses, Q describes the predicted commonnesses, and L describes the number of instances in the dataset [23].

2.1.1.8 Minimum Redundancy Maximum Relevance (MRMR)

MRMR is an FS technique that aims to encounter features that are maximally appropriate to the data class and minimally redundant to the other features. Technique accuracy is enhanced by terminating redundant features or features that show a robust correlation with the category, as a development of advancing feature relevance and reducing feature redundancy. The main advantages are accurate and quick analysis compared to other techniques, as well as relevancy optimization. The technique, however, is known to be significantly influenced by parametric measure sensitivity [24].

2.1.2 Wrapper Techniques

Wrapper techniques consider a subset of features with a given ML algorithm to access their efficiency. In contrast to filter approaches, which rank features regardless of any learners, wrapper approaches train and test the model repeatedly with different sets of features to select the set that has the best predictive performance [25]. These approaches evaluate the quality of feature subsets according to prediction performance as a target function. The wrapper technique is usually carried out prior to and after the training when learning is performed. This objective is to maximize the prediction accuracy and the quality of the fit by optimally grouping features for classification problems. An advantage of wrapper techniques is that they can consider interactions and dependencies among features, and choose more informative subsets. Nevertheless, one of the main drawbacks is that the computational burden is heavy as the learning algorithm has to be retrained and tested for each candidate subset. The overall workflow of a simplex wrapper-based feature selection process is shown in Figure 2 Familiar wrapper approaches include PSO, GA, and RFE [26].

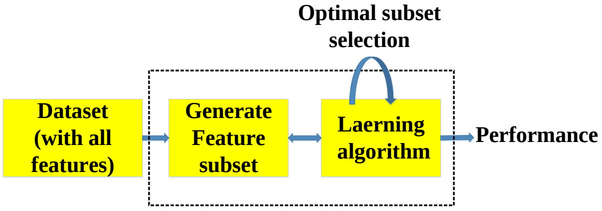

Figure 2. The wrapper technique workflow.

2.1.2.1 Stepwise

Stepwise is a classical wrapper technique in statistics and data science that is frequently used. It uses a greedy algorithm that examines and selects the subsets of variables in terms of contributing to the model performance of each step [27]. Variables are gradually added or removed according to optimization criteria in order to construct an almost optimal subset. One advantage of the stepwise approach is that it is supervised - it assesses the contribution of leads on the outcome which allows meaningful selection of leads based on relative effect. Nonetheless, stepwise selection is not applicable to high-dimensional data, due to a potential in over-fitting and computational expense. In addition, the globally optimal subset cannot be ensured to be found, because it is sensitive to the order of feature inclusion and dependent on the initial solution. This limitation is partially relieved by removing recently added redundant features [28].

2.1.2.2 Boruta

Boruta is the wrapper feature selection technique that centers on the RF classifier. It derives from two core principles. First, it incorporates shadow features which is generated as a mirrored features from the original features, being used as a reference for the assessment of feature importance [29]. If features import score is significantly lower than the top ranked shadow feature then it does not have significant predictive value. Second, Boruta performs this evaluation process repeatedly n times and then in each round uses the binomial distribution to statistically test the significance of the features. Only features which outperform shadow features in all iterations are kept. Although Boruta works very well in terms of identifying all important features, it is computationally expensive since it requires repeated model fitting and ranking. Moreover, it tends to yield models with many features, as it does not favor model sparsity [30].

2.1.2.3 Genetic Algorithm (GA):

GA inspired by evolution in natural systems, are heuristic optimization techniques evolved from the basic principles underlying natural selection and genetics [31]. For the purpose of feature selection, each set of potential features is taken as a “chromosome” of a population. The chromosomes then evolved across successive generations employing operations like crossover (recombination) and mutation for the exploration of the search space by the algorithm.

A fitness function (FF) is a measure employed to estimate the quality of each chromosome at this stage - which normally utilizes the predictive performance of the model with the applied feature selection. There are greater chances that subsets having baser performance are less likely to be selected for reproduction across generations [32]. GA are also beneficial in feature selection for high-dimensional data space where exhaustive search is not feasible. Such techniques can find a set of (near) optimal features with moderate amounts of computation. However, GAs performance may be affected by hyperparameters such as mutation rate, crossover probability, and size of the population, which sometimes needs to be adjusted during an iterative process [30].

2.1.2.4 Recursive Feature Elimination (RFE)

RFE is a valid FS technique for irrelevant feature elimination from the input feature set and identifying the critical features that distinguish between classes [33]. The RFE aims to decrease the complexity of the feature set while keeping high accuracy. It leverages the capacities of RF classifiers to achieve an iterative evaluation of the variable importance, demanding the performance of numerous classification iterations. The iterative strategy consists of multiple key phases: the development of a vital RF classifier, the estimation employing cross-validation techniques, the investigation of feature prerogative metrics, and the transformation of the feature set for succeeding iterations. Every feature in the subset is engaged in the initial classification round. The worst-performing features are then determined and removed from the feature set, training for the subsequent phases. Also, RFE reprises this technique to decrease the likelihood of convergence and dependencies among the input features [34].

2.1.2.5 Sequential forward selection (SFS)

SFS is a wrapper technique in which attributes are consecutively set to empty prospects until the standard is not modified [35]. SFS techniques employed to estimate an initial dimensional feature space from a new dimensional feature subspace are comprised in a set of mean investigation algorithms. The objective is to determine a subset of features that are most appropriate to the goal, resulting in optimal computational implementation while decreasing overfitting by deleting irrelevant data. The SFS functions sufficiently when the optimal subset has a small feature number [12].

2.1.2.6 Sequential backward selection (SBS)

The SBS is a heuristic technique that automatically determines a sub-set of features most relevant to a characteristic issue, assisting in the evolution of a more efficient standard [36]. The SBS technique iteratively eliminates features with the least negligible influence on the performance until the selected number of features is produced. The reduction of unimportant features enables model inference on unrecognized data and enhances computational efficiency. The functional SBS principle begins with all the features, with an original size. The first iteration develops all potential feature subsets. For each subset, one feature is terminated, and the performance of the approach is evaluated. The feature that is skipped from the subset with the most suitable score is terminated. This process persists until the best evaluation score is specified as the last feature set [37].

2.1.2.7 Simulated annealing (SA)

SA is a wrapper-based FS technique that is based on the process of annealing, which is the rearrangement of particles in a material as it cools (with reduced thermal energy) to a more stable state [38]. By analogy, the rank of SA will sort the different feature importance in terms of the classification performance and reshuffling particles for minimum energy. The algorithm starts by randomly choosing a set of features and a specified number of iterations. In each step, a new candidate set is produced by adding or deleting some elements from the previous set. If this change leads to better predict, the new subset is adopted. Otherwise, it may still be accepted according to the decreasing probability, resembling the cooling process. Such randomness helps the algorithm to get out of local minima and take chance of discovering a globally optimal feature subset [30].

2.1.3 Embedded Techniques

Embedded feature selection techniques: they include the selection process in the model training itself, being a compromise between filter and wrapper techniques. In such techniques, the learning algorithm of model construction automatically selects the most informative features. In contrast to Wrapper techniques, which consider different feature subsets during the selection process, embedded techniques select the best features during the learning process, which is in general faster and less overfitting-prone.

Such techniques repeatedly select the most informative features and build a stable-and optimal-classifier. Embedded approaches are good at discovering feature interaction relationships since they are built into the structure of the model. They thus present a sweet compromise between performance and complexity. Figure 3 schematic workflow concept of embedded feature selection procedure [26].

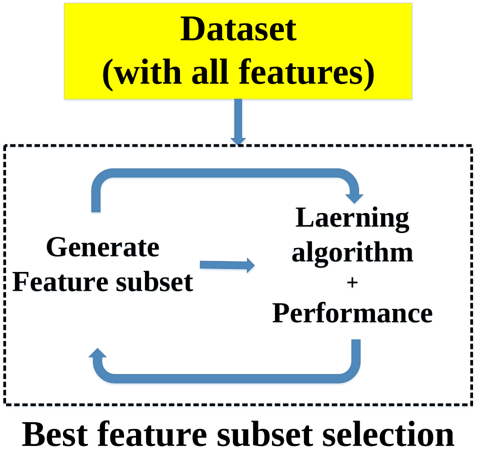

Figure 3. The embedded technique workflow.

2.1.3.1 Least Absolute Shrinkage and Selection Operator (LASSO)

LASSO is an embedded technique of linear regression that integrates both regularization and feature selection. It reduces the sum of squares subject to a constraint on the sum of the absolute values of the model coefficients. This penalty — known as the L1 penalty — leads to a sparse model, meaning that the terms of some coefficients will end up being exactly zero, which in turn can help select important features and make models more interpretable. Mathematically, the LASSO loss of the function is described as: calculated in Equation 12 [28].

$$L1 = \sum_{j=1}^{m}\left(y_i - \sum_{i=1}^{q} x_{ji}\beta_i\right)^2 + \vartheta \sum_{i=1}^{q} |\beta_i| \dots (12)$$

By comparison, Ridge Regression, another regularization approach, applies the L2 penalty punishing the magnitude of the coefficients squared shows in Equation 13.

$$L2 = \sum_{j=1}^{m}\left(y_i - \sum_{i=1}^{q} x_{ji}\beta_i\right)^2 + \vartheta \sum_{i=1}^{q} \beta_i^2 \dots (13)$$

Where m is the number of training instances, q is number of features, y_i is the true target value for the j^{th} training sample, x_{ji} is the value of feature i for the training sample j^{th}, β_i is the coefficient of the i^{th} feature, and ϑ the regularization parameter that controls the strength of the penalty. The main advantage is the lower computation cost since it involves assessing regression parameters subject to a penalty term and creating parsimonious standards. On the other hand, it does not necessarily select the most illuminating features, and periodically, the variables specified are just too few [28].

2.1.3.2 Ridge

Ridge, likewise understood as L2-Regularization, is a regularization technique. The objective process delivers a squared magnitude of the standard parameters as a penalty term. Ridge shrink's technique coefficients near zero but not exactly zero, unlike the LASSO technique, which presents a sparse exhibition of functions. The most significant discrepancy between Ridge and LASSO strategies is that LASSO decreases the coefficients of insignificant variables to zero, effectively compensating them out. Ridge creates non-zero coefficients, which are more helpful in analyzing features [11].

2.1.4 Hybrid Techniques

Hybrid FS techniques Several feature selection techniques have been combined (hybridized) in existing research to take the advantage of two or more techniques at once [24][39]. Usually, the hybrid techniques are formed by the filtering and wrapping techniques combined together to overcome the problems of each technique. This, in turn, they increase the classification accuracy over that of single filter techniques, while keeping down the computational time of wrapper techniques.

Hybrid techniques are also especially useful when confronted with standard data problems, such as missing values and noisy data. They accomplish this by filtering irrelevant features using filter techniques to reduce noise, and then by refining using wrapper techniques or other selection approaches. This two-step approach does not only enhance the model's efficacy as a whole, but reduces the data, which in turn leads to quicker, better results. The typical procedure of a hybrid feature selection approach is shown in Figure 4 [24].

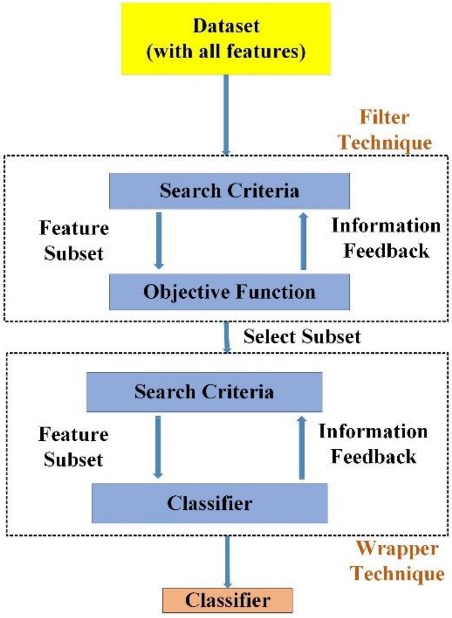

Figure 4. The general process of hybrid technique.

2.1.5 Ensemble Feature Selection (EFS)

EFS integrates the outcomes of multiple FS techniques to create a more accurate and robust subset of features. Rather than counting on a single technique, EFS combines multifarious selection techniques to enhance predictive performance, reduce redundancy, and improve generalization. By merging or aggregating feature subsets from various techniques, EFS apprehends complementary data, mitigates biases of individual techniques, and frequently exceeds standalone techniques in ML tasks. Figure 5 presents the principle of the EFS technique [12].

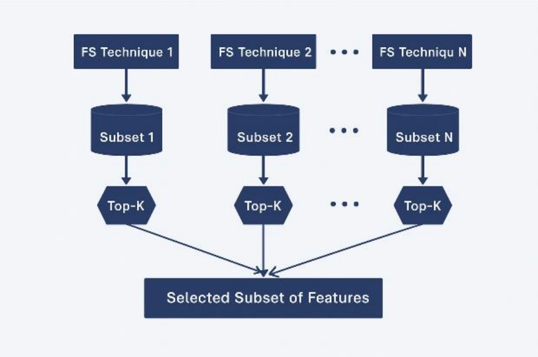

Figure 5. Overview of EFS technique.

2.2 Main Factors to Select FS Techniques

There are several characteristics to contemplate when choosing an FS technique for a special task, such as [30]:

- Dataset size: If the dataset is huge, consider a technique that is scalable and computationally efficient.
- Handle missing values in your dataset: If the dataset has missing values, consider choosing a technique that can effectively handle them.
- Interpretable features: If you consider how the specified features contribute to the model's forecasts, you may want to select a more interpretable technique.
- Selected features to be vital to differences in the modeling assumptions or data: robust technique must be chosen.

3. Overview of Recent Studies FS Techniques for Cancer

Table 2 presented provides a comparative overview of recent studies applying ML and FS techniques for CD and prognosis. Each study focuses on a distinct cancer type such as breast, gastric, lung, ovarian, cervical, colorectal, skin and prostate and explores different datasets, methodologies, and performance outcomes. Key information includes the publication year, cancer type, FS techniques, the type of FS (e.g., filter, wrapper, embedded, ensemble), dataset details, and the best-reported accuracy or related metrics. The structured summary is intended to enable the comparison and evaluation of ML techniques across diverse biomedical contexts, highlighting how FS and classifier combinations impact predictive performance in high-dimensional cancer datasets.

FS techniques have demonstrated substantial effectiveness in cancer prediction by enhancing model accuracy, reducing dimensionality, and mitigating overfitting. Based on the studies summarized in Table 2., a wide range of FS techniques were applied across various cancer types, including filter, wrapper, embedded, hybrid, and Ensemble techniques. For instance, the study [40] used the Variable Importance Measure (VIM) to breast cancer datasets (WDBC and WBC), achieving high accuracies of 97.05% and 97.76%, highlighting the efficacy of variable importance-based filtering. In contrast, [41] used a combination of ANOVA, MI, Extra Trees, and Logistic Regression on the TCGA breast cancer dataset. However, it achieved only 0.86% accuracy due to challenges like class imbalance and the lack of biological feature integration. Recent studies show that hybrid and metaheuristic FS techniques often yield better results. For example, [42] employed PSO across multiple breast cancer datasets and achieved up to 100% accuracy on WDBC, while [43] combined RFE, LASSO, Boruta, PCA, and Relief to reach 99.3%. Similarly,[44] demonstrated exceptional performance (99.98%) across multiple microarray datasets by integrating five filter techniques with Moth-Flame Optimization and Extreme Learning Machine (ELM). However, it suffered from overfitting due to high dimensionality and low sample sizes. In [45], which focused on prostate cancer, a comprehensive suite of FS techniques (Relief-F, L1-Lasso, Boruta, Genetic Algorithm, etc.) was used. Despite the methodological diversity, performance was more modest (AUC = 0.71, F1 = 0.76), illustrating that the number of FS practices alone does not guarantee superior results without careful model tuning and external validation. The [46] and [47] employed RF achieve excellent accuracy (100%) in lung cancer. Generally, the investigation indicates that ensemble and hybrid FS techniques tend to exceed single techniques in cancer prediction tasks.

4. Current Challenges and Research Gaps

In spite of the recent progress of FS technologies developed for cancer analysis, there have many challenges and research gaps that have impeded the generalization, robustness and clinical transformation of these approaches. The discussion of recent research provides a wealth of critical points that have to be covered for stronger and reproducible conclusions. However, several challenges must be handled, such as small sample sizes, computational cost, data imbalance, and lack of external validation remain standard limitations to provide reliable and generalizable clinical models.

4.1 Handling Noisy and Incomplete Data

One of the main issues so far is that the data are often noisy and incomplete, particularly clinical data and microarray data. The works of Lanjewar et al. [47] and Acosta et al. [48] used data that were not resilient to noise in real-world or had no data imputation capabilities. Most models were trained on nice, clean data, which is often quite different from real-world conditions in which you commonly have to deal with missing data and sensor noise. There are few papers that apply strong preprocessing techniques and techniques to deal with missing values. Future work should also address the development of FS techniques traffic flow and pipeline integrated data cleaning to insure clinical relevance in models.

4.2 Small Sample Size and High-Dimensionality

The curse of dimensionality the majority of cancer datasets, especially microarray and transcriptomic datasets, have a large number of features and a small number of samples. Publications like Abraham et al. [49] showed high accuracy, but they also had low sample sizes, which may result in overfitting or low statistical power. showed high accuracy, but they also had low sample sizes, which may result in overfitting or low statistical power. Further directions may include investigating the data augmentation, oversampling, and transfer learning in combating the problem of small data. Dimension reduction together with incorporation of prior biological knowledge is also underexplored.

4.*3 Small Sample Size and High-Dimensionality*

Advanced FS algorithms, such as GA, PSO, and DL-based hybrids (i.e., Lv et al. [50]; Zhu et al. [51]) are highly predictive, but computationally expensive. This scalability problem narrows the scope, in particular in the context of large scale or real-time datasets. Although there is accuracy gain with the two stage (hybrid/ensemble) approaches, there is an urgent demand for efficient, scalable FS frameworks that achieve a tradeoff between performance and speed.

Table 2: comparative overview of recent studies FS techniques for cancer

Study	**Year**	**Cancer Type**	**FS Techniques**	**FS Type**	**Dataset Name**	**Accuracy (Best)**
Sharma et al. [52]	2021	Breast	CFS, SFS, IG	Filter, Wrapper, Filter respectively	WDBC (UCI)	93%
Taghizadeh et al. [41]	2022	Breast	ANOVA, MI, Extra Trees, Logistic Regression	Filter (ANOVA, MI), Embedded (LGR), Ensemble (Extra Trees)	TCGA (Breast cancer transcriptome)	0.86%
Huang et al. [40]	2022	Breast	Variable Importance Measure (VIM)	Filter	WDBC, WBC (UCI)	97.05% / 97.76%
Afrash et al. [53]	2023	Gastric	Boruta, mRMR, LASSO, Relief	Wrapper (Boruta), Filter (mRMR, Relief), Embedded (LASSO)	Ayatollah Talleghani Hospital Iran	89.10%
Varan1 et al. [54]	2023	Prostate	Importance coefficient, Sequential (forward/backward), CC	Filter + Wrapper	PROSTATEx	90%
Lanjewar et al. [47]	2023	Lung	MRMR, P-value test	Filter	Kaggle CT scan	100%
Abraham et al. [49]	2023	Ovarian	Boruta, Lasso	Filter	Cancer Cell Line Encyclopedia	98.5%
Mylona et al. [45]	2024	Prostate	Boruta, mRMRe, ReliefF, RFE, RF-imp, L1-Lasso, CorrSF, HSIC-Lasso, SES, GA	Filter (ReliefF, mRMRe, CorrSF, SES), Wrapper (Boruta, RFE, GA), Embedded (L1-Lasso, HSIC-Lasso), Ensemble (RF-imp)	ProstateNET & ProstateX2 (MRI)	AUC: 0.71, F1: 0.76
Ileberi et al. [55]	2024	Cervical	PSO	Wrapper	Cervical Cancer Risk Factors Dataset (CCRFD)	98%
Sucharita1 et al. [44]	2024	Multiple (Microarray Data: Lung, Brain, Prostate, Colon)	Ensemble of 5 filters + Moth-Flame Optimization (MFO) with ELM	Hybrid (Filter + Wrapper)	4 microarray datasets	99.98%
Tripathy et al. [56]	2024	Multiple (Microarray): Colon, Leukemia, Prostate, Brain tumor, Breast, and Adenoma.	4 Filter Techniques (CBFS, CST, InG, RFS) with ELM Ensemble	Filter + Ensemble	7 microarray datasets	74%

Corresponding author E-mail address: manar.thaer@nahrainuniv.edu.iq

https://doi.org/10.61268/q78nda33

Kazerani [42]	2024	Breast	PSO	Metaheuristic / Wrapper	WDBC, WPBC, Coimbra	100%
Lv et al. [50]	2024	Colorectal	DRMF-PaRa Framework (integrated deep FS)	Embedded / Hybrid (DL + Fusion)	Multimodal dataset	87.6%
Jasurbek et al. [57]	2024	Skin	GA	Wrapper	ISIC 2017	97.02
Sahu et al. [58]	2025	Breast	PCA + LASSO	Hybrid	Breast Cancer Wisconsin Diagnostic Dataset	99.42%
Acosta- et al. [48]	2025	Ovarian	GA	Wrapper	Ovarian Cancer dataset (Soochow University)	90.54%
Khanna et al. [59]	2025	Lung	Gini Importance-Based Rankin, RFE	Ensemble	VOCs from 7 bio sources	100%
Al-Jamimi et al. [46]	2025	Lung	RFE-SVM	Hybrid	Two unnamed lung datasets	100%
Chhillar et al. [43]	2025	Breast	RFE, LASSO, Boruta, PCA, and Relief.	Filter, Wrapper, Embedded	WDBC, MBCD	99.3%
Zhu et al. [51]	2025	Breast	SHAP-RF-RFE + LightGBM + PSO	Hybrid / Ensemble	WDBC	99.0%
Islam et al. [60]	2025	Thyroid	SelectKBest + Random Oversampling + Ensemble Voting	Filter + Ensemble		99.72%
Sharma et al. [52]	2021	Breast	CFS, SFS, IG	Filter, Wrapper, Filter respectively	WDBC (UCI)	93%
Taghizadeh et al. [41]	2022	Breast	ANOVA, MI, Extra Trees, Logistic Regression	Filter (ANOVA, MI), Embedded (LGR), Ensemble (Extra Trees)	TCGA (Breast cancer transcriptome)	0.86%
Huang et al. [40]	2022	Breast	Variable Importance Measure (VIM)	Filter	WDBC, WBC (UCI)	97.05% / 97.76%

4.4 Reproducibility and Standardization Issues

Both Taghizadeh et al. [41] and Chhillar et al. [43] do not have external validation or standardized performances. Reproducibility is hampered by a lack of (1) standardized evaluation metrics, (2) publicly available data, and (3) released code and models.

Close The demand for community-driven benchmark datasets, the development of standardized evaluation pipelines, and open-source repositories is crucial to improve reproducibility and promote fair comparison among FS techniques.

5. Future Directions

With the increasing reliance of CD and prognosis on complex, high-dimensional data, FS proves to be crucial for the ever-advancing need for accuracy, scalability, and interpretability. Although great progress has been achieved, there still exists a set of potential research directions in the FW techniques which are not well-studied yet. Three of the most important research and development directions are described in the next 3 sub-sections.

5.1 DL and FS Integration

DL dominates, although its combination with FS is relatively rare in medical imaging and omics data analysis. Conventional FS techniques are typically partitioned from DL pipelines, and the feature representations generated by them are not guaranteed to be the most appropriate. Future work will consider embedding FS in DL architectures (e.g., as attention mechanisms, as sparsity-inducing layers, etc., or in autoencoderbased selection policies). Such an integration might reveal hierarchical and non-linear relationships even in complex cancer data, and thus increase predictability as well as interpretability of the model. Furthermore, learning paradigms that jointly optimize FS and DL may result in smaller and more generic models for clinical usage.

5.2 Explainable AI and Interpretability

Many of the ML and DL models are black-box, which is another obstacle for translation into clinical practice. The FS techniques improving model interpretability and explainability are necessary for obtaining the trust of the medical experts. Prospective researches could concentrate on explainable AI (XAI)-based FS techniques, e.g., SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) and integrated gradients for literature for ranking important features and the biological interpretation. Integrating FS with domain knowledge and visual analytics offers potential access to underlying disease mechanisms, enabling contributions to better-informed management and patient-specific risk estimations.

5.3 Federated and Privacy-Preserving FS

Access to rich and diverse datasets is the key to training robust predictive models, but privacy considerations and data ownership laws conspire against sharing data between institutions. Federated learning (FL) is a promising paradigm for cooperative FS and model learning without exchanging raw data. In the future, federated FS algorithms that are able to perform selection over decentralized datasets without having access to the original samples should also be explored. We should be investigating approaches such as secure multiparty computation, differential privacy, and homomorphic encryption to support strenuous and compliant FS workflows across hospitals and between hospitals and research institutions.

5.4 Real-Time and Adaptive FS Systems

Real time and adaptive decision support is more and more crucial in CD, particularly where time is a significant factor such as point-of-care or intraoperative. Finally, current FS approaches frequently based on offline and static pipelines are not well suited for dynamic environments. The next-generation FS systems are also expected to have online learning ability to be able to evolve for streaming data, changing disease patterns, and patient-specific heterogeneity. We believe that there is a calling for research in building minimal, incremental FS algorithms that can learn and grow feature subsets incrementally without explicitly retraining the whole model. Adaptive FS-based frameworks of this type might also enhance models' generalizability to a new clinical context or novel cancer subtypes.

5. Conclusions

FS is still a fundamental component in the construction of accurate, interpretable, and efficient machine and DL models for CD and prediction. In this review, we have highlighted the continuum FS techniques—normal filter and wrapper until advanced hybrid and ensemble approaches—that have been shown to be powerful toward data from diverse cancer types. A number of reports on FS acknowledge that hybrid and ensemble FS approaches tend to yield better performance in comparison to other FS techniques but with significantly increased computational and algorithmic effort. Furthermore difficulties, i.e., small sample sizes, high-dimensional data, class imbalance, and lack of reproducibility, still remain in the domain.

The review highlights the challenges and gaps within the area that deserve immediate attention: (1) linking FS techniques with biologically relevant knowledge to increase clinical utility of models; (2) identifying noise-robust scalable algorithms to generalize better across datasets; and (3) standardization of benchmarking for promoting reproducibility and making the technique comparison process easier. Future work could further explore the development of interpretable FS frameworks, and utilize the recent progress in federated learning, synthetic data generation and multimodal fusion to address the current limitations. Solving these challenges will help to achieve the next generation of FS techniques for early cancer diagnostics, predictive and personalized treatment planning, and on-the-fly clinical decision support systems.

References

[1] M. S. Hossain *et al.*, "Colorectal cancer: a review of carcinogenesis, global epidemiology, current challenges, risk factors, preventive and treatment strategies," *Cancers (Basel)*, vol. 14, no. 7, p. 1732, 2022.

[2] M. Dal-Bo *et al.*, "Intrinsic and extrinsic factors influencing the clinical course of B-cell chronic lymphocytic leukemia: prognostic markers with pathogenetic relevance," *J Transl Med*, vol. 7, pp. 1–14, 2009.

[3] X. Wang *et al.*, "Overcoming cancer treatment resistance: unraveling the role of cancer-associated fibroblasts," *Journal of the National Cancer Center*, 2025.

[4] I. F. Jassam, A. A. Mukhlif, A. A. Nafea, M. A. Tharthar, and A. I. Khudhair, "A Review of Breast Cancer Histological Image Classification: Challenges and Limitations," *Iraqi Journal for Computer Science and Mathematics*, vol. 6, no. 1, p. 1, 2025.

[5] H. Jiang, W. Sun, H. Guo, J. Zeng, X. Xue, and S. Li, "Review of intelligent diagnosis methods for imaging gland cancer based on machine learning," *Virtual Reality & Intelligent Hardware*, vol. 5, no. 4, pp. 293–316, 2023.

[6] S. H. Majeed, A. K. Kareem, A. A. Nafea, and K. M. Ali Alheeti, "Efficient cardiovascular disease classification using supervised machine learning on photoplethysmography data," in *IET Conference Proceedings CP906*, IET, 2024, pp. 431–435.

[7] F. A. Altuhaifa, K. T. Win, and G. Su, "Predicting lung cancer survival based on clinical data using machine learning: A review," *Comput Biol Med*, vol. 165, p. 107338, 2023.

[8] M. Büyükkeçeci and M. C. Okur, "A comprehensive review of feature selection and feature selection stability in machine learning," *Gazi University Journal of Science*, vol. 36, no. 4, pp. 1506–1520, 2023.

[9] A. A. Nafea, A.-M. Manar, K. M. A. Alheeti, M. S. I. Alsumaidaie, and M. M. AL-Ani, "A hybrid method of 1D-CNN and machine learning algorithms for breast cancer detection," *Baghdad Science Journal*, vol. 21, no. 10, p. 3333, 2024.

[10] H. J. Mohammed, A. A. Nafea, H. K. Almulla, S. A. S. Aliesawi, and M. M. Al-Ani, "An effective hybrid model for skin cancer detection using transfer learning," in *2023 16th International Conference on Developments in eSystems Engineering (DeSE)*, IEEE, 2023, pp. 840–845.

[11] Y. Bouchlaghem, Y. Akhiat, and S. Amjad, "Feature selection: a review and comparative study," in *E3S web of conferences*, EDP Sciences, 2022, p. 01046.

[12] A. Jafar and M. Lee, "High-Accuracy COVID-19 Prediction Using Optimized Union Ensemble Feature Selection Approach," *IEEE Access*, 2024.

[13] K. A. ElDahshan, A. A. AlHabshy, and L. T. Mohammed, "Filter and Embedded Feature Selection Methods to Meet Big Data Visualization Challenges.," *Computers, Materials & Continua*, vol. 75, no. 1, 2023.

[14] S. Asghari, H. Nematzadeh, E. Akbari, and H. Motameni, "Mutual information-based filter hybrid feature selection method for medical datasets using feature clustering," *Multimed Tools Appl*, vol. 82, no. 27, pp. 42617–42639, 2023.

[15] V. Ventafridda, M. Tamburini, A. Caraceni, F. De Conno, and F. Naldi, "A validation study of the WHO method for cancer pain relief," *Cancer*, vol. 59, no. 4, pp. 850–856, 1987.

[16] H.-C. Huang, S. Zheng, and Z. Zhao, "Application of Pearson correlation coefficient (PCC) and Kolmogorov-Smirnov distance (KSD) metrics to identify disease-specific biomarker genes," *BMC Bioinformatics*, vol. 11, pp. 1–2, 2010.

[17] R. Gupta, A. K. Yadav, S. K. Jha, and P. K. Pathak, "Composition of feature selection techniques for improving the global horizontal irradiance estimation via machine learning models," *Thermal Science and Engineering Progress*, vol. 48, p. 102394, 2024.

[18] A. Naik, V. Kuppili, and D. R. Edla, "Binary dragonfly algorithm and fisher score based hybrid feature selection adopting a novel fitness function applied to microarray data," in *2019 international conference on applied machine learning (ICAML)*, IEEE, 2019, pp. 40–43.

[19] F. Yang, Z. Xu, H. Wang, L. Sun, M. Zhai, and J. Zhang, "A hybrid feature selection algorithm combining information gain and grouping particle swarm optimization for cancer diagnosis," *PLoS One*, vol. 19, no. 3, p. e0290332, 2024.

[20] M. Çakır, A. Degirmenci, and O. Karal, "Exploring the behavioural factors of cervical cancer using ANOVA and machine learning techniques," in *International Conference on Science, Engineering Management and Information Technology*, Springer, 2022, pp. 249–260.

[21] M. Salman Pathan, A. Nag, M. Mohisn Pathan, and S. Dev, "Analyzing the impact of feature selection on the accuracy of heart disease prediction," *arXiv e-prints*, p. arXiv-2206, 2022.

[22] X. Cheng, "A Comprehensive Study of Feature Selection Techniques in Machine Learning Models," *Insights in Computer, Signals and Systems*, vol. 1, no. 1, pp. 10–70088, 2024.

[23] T. Q. Al-Ghadi, S. Manickam, I. D. M. Widia, E. R. N. Wulandari, and S. Karuppayah, “Leveraging federated learning for DoS attack detection in IoT networks based on ensemble feature selection and deep learning models,” *Cyber Security and Applications*, vol. 3, p. 100098, 2025.

[24] A. Khoirunnisa, D. Adytia, and M. M. Deris, “Comprehensive Review of Hybrid Feature Selection Methods for Microarray-Based Cancer Detection,” *IEEE Access*, 2025.

[25] M. M. M. Mufassirin and R. G. Ragel, “A novel filter-wrapper based feature selection approach for cancer data classification,” in *2018 IEEE International Conference on Information and Automation for Sustainability (ICIAfS)*, IEEE, 2018, pp. 1–6.

[26] W. Kanyongo and A. E. Ezugwu, “Feature selection and importance of predictors of non-communicable diseases medication adherence from machine learning research perspectives,” *Inform Med Unlocked*, vol. 38, p. 101232, 2023.

[27] A. Obulkasim, G. A. Meijer, and M. A. van de Wiel, “Stepwise classification of cancer samples using clinical and molecular data,” *BMC Bioinformatics*, vol. 12, pp. 1–12, 2011.

[28] A. Gramegna and P. Giudici, “Shapley feature selection,” *FinTech*, vol. 1, no. 1, pp. 72–80, 2022.

[29] T. Öznacar and T. Güler, “Prediction of early diagnosis in ovarian cancer patients using machine learning approaches with Boruta and advanced feature selection,” *Life*, vol. 15, no. 4, p. 594, 2025.

[30] D. Theng and K. K. Bhoyar, “Feature selection techniques for machine learning: a survey of more than two decades of research,” *Knowl Inf Syst*, vol. 66, no. 3, pp. 1575–1637, 2024.

[31] S. Shah and A. Kusiak, “Cancer gene search with data-mining and genetic algorithms,” *Comput Biol Med*, vol. 37, no. 2, pp. 251–261, 2007.

[32] A. Zibakhsh and M. S. Abadeh, “Gene selection for cancer tumor detection using a novel memetic algorithm with a multi-view fitness function,” *Eng Appl Artif Intell*, vol. 26, no. 4, pp. 1274–1281, 2013.

[33] C. Peng, X. Wu, W. Yuan, X. Zhang, Y. Zhang, and Y. Li, “MGRFE: Multilayer recursive feature elimination based on an embedded genetic algorithm for cancer classification,” *IEEE/ACM Trans Comput Biol Bioinform*, vol. 18, no. 2, pp. 621–632, 2019.

[34] N. S. Azman *et al.*, “Support vector machine–Recursive feature elimination for feature selection on multi-omics lung cancer data,” *Progress In Microbes & Molecular Biology*, vol. 6, no. 1, 2023.

[35] S. Bouatmane, M. A. Roula, A. Bouridane, and S. Al-Maadeed, “Round-Robin sequential forward selection algorithm for prostate cancer classification and diagnosis using multispectral imagery,” *Mach Vis Appl*, vol. 22, pp. 865–878, 2011.

[36] S. Gunasundari and S. Janakiraman, “A hybrid PSO-SFS-SBS algorithm in feature selection for liver cancer data,” in *Power Electronics and Renewable Energy Systems: Proceedings of ICPERES 2014*, Springer, 2014, pp. 1369–1376.

[37] S. O. Aregbesola, J. Won, S. Kim, and Y.-H. Byun, “Sequential backward feature selection for optimizing permanent strain model of unbound aggregates,” *Case Studies in Construction Materials*, vol. 19, p. e02554, 2023.

[38] B. Haznedar, M. T. Arslan, and A. Kalinli, “Optimizing ANFIS using simulated annealing algorithm for classification of microarray gene expression cancer data,” *Med Biol Eng Comput*, vol. 59, pp. 497–509, 2021.

[39] O. J. Kadhim, A. A. Nafea, S. A. S. Aliesawi, and M. M. Al-Ani, “Ensemble model for prostate cancer detection using MRI images,” in *2023 16th International Conference on Developments in eSystems Engineering (DeSE)*, IEEE, 2023, pp. 492–497.

[40] Z. Huang and D. Chen, “A breast cancer diagnosis method based on VIM feature selection and hierarchical clustering random forest algorithm,” *IEEE Access*, vol. 10, pp. 3284–3293, 2021.

[41] E. Taghizadeh, S. Heydarheydari, A. Saberi, S. JafarpoorNesheli, and S. M. Rezaeijo, “Breast cancer prediction with transcriptome profiling using feature selection and machine learning methods,” *BMC Bioinformatics*, vol. 23, no. 1, p. 410, 2022.

[42] R. Kazerani, “Improving breast cancer diagnosis accuracy by particle swarm optimization feature selection,” *International Journal of Computational Intelligence Systems*, vol. 17, no. 1, p. 44, 2024.

[43] I. Chhillar and A. Singh, “An improved soft voting-based machine learning technique to detect breast cancer utilizing effective feature selection and SMOTE-ENN class balancing,” *Discover Artificial Intelligence*, vol. 5, no. 1, p. 4, 2025.

[44] S. Sucharita, B. Sahu, T. Swarnkar, and S. K. Meher, “Classification of cancer microarray data using a two-step feature selection framework with moth-flame optimization and extreme learning machine,” *Multimed Tools Appl*, vol. 83, no. 7, pp. 21319–21346, 2024.

[45] E. Mylona *et al.*, “Optimizing radiomics for prostate cancer diagnosis: feature selection strategies, machine learning classifiers, and MRI sequences,” *Insights Imaging*, vol. 15, no. 1, p. 265, 2024.

[46] H. A. Al-Jamimi, S. Ayad, and A. El Kheir, “Integrating Advanced Techniques: RFE-SVM Feature Engineering and Nelder-Mead Optimized XGBoost for Accurate Lung Cancer Prediction,” *IEEE Access*, 2025.

[47] M. G. Lanjewar, K. G. Panchbhai, and P. Charanarur, “Lung cancer detection from CT

scans using modified DenseNet with feature selection methods and ML classifiers," *Expert Syst Appl*, vol. 224, p. 119961, 2023.

[48] S. Acosta-Jiménez *et al.*, "Detection of ovarian cancer using a methodology with feature extraction and selection with genetic algorithms and machine learning," *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 14, no. 1, p. 3, 2024.

[49] A. Abraham, H. S. Mohideen, and R. Kayalvizhi, "A tabular variational auto encoder-based hybrid model for imbalanced data classification with feature selection," *IEEE Access*, vol. 11, pp. 122760–122771, 2023.

[50] Z. Lv, R. Yan, Y. Lin, L. Gao, F. Zhang, and Y. Wang, "A disentangled representation-based multimodal fusion framework integrating pathomics and radiomics for KRAS mutation detection in colorectal cancer," *Big Data Mining and Analytics*, 2024.

[51] J. Zhu *et al.*, "An integrated approach of feature selection and machine learning for early detection of breast cancer," *Sci Rep*, vol. 15, no. 1, p. 13015, 2025.

[52] A. Sharma and P. K. Mishra, "Performance analysis of machine learning based optimized feature selection approaches for breast cancer diagnosis," *International Journal of Information Technology*, vol. 14, no. 4, pp. 1949–1960, 2022.

[53] M. R. Afrash, E. Mirbagheri, M. Mashoufi, and H. Kazemi-Arpanahi, "Optimizing prognostic factors of five-year survival in gastric cancer patients using feature selection techniques with machine learning algorithms: a comparative study," *BMC Med Inform Decis Mak*, vol. 23, no. 1, p. 54, 2023.

[54] M. Varan, J. Azimjonov, and B. Maçal, "Enhancing prostate cancer classification by leveraging key radiomics features and using the fine-tuned linear svm algorithm," *Ieee Access*, vol. 11, pp. 88025–88039, 2023.

[55] E. Ileberi and Y. Sun, "Machine Learning-Assisted Cervical Cancer Prediction Using Particle Swarm Optimization for Improved Feature Selection and Prediction," *IEEE Access*, 2024.

[56] J. Tripathy, R. Dash, and B. K. Pattanayak, "Unleashing the power of machine learning in cancer analysis: a novel gene selection and classifier ensemble strategy," *Research on Biomedical Engineering*, vol. 40, no. 1, pp. 125–137, 2024.

[57] N. Jasurbek, D. F. Ivan, O. S. Ajani, and R. Mallipeddi, "Genetic algorithm guided image channel selection for skin lesion segmentation," *IEEE Access*, 2024.

[58] P. K. Sahu and T. Fatma, "Optimized Breast Cancer Classification Using PCA-LASSO Feature Selection and Ensemble Learning Strategies with Optuna Optimization," *IEEE Access*, 2025.

[59] D. Khanna, A. Kumar, and S. A. Bhat, "Volatile Organic Compound for the Prediction of Lung Cancer by using Ensembled Machine Model and Feature Selection," *IEEE Access*, 2025.

[60] M. R. Islam, A. I. Chowdhury, S. Shama, and M. M. H. Lamyea, "Enhanced thyroid disease prediction using ensemble machine learning: a high-accuracy approach with feature selection and class balancing," *Discover Artificial Intelligence*, vol. 5, no. 1, p. 9, 2025.