Article

Feature Selection Method Based on Simultaneous Perturbation Stochastic Approximation Technique Evaluated on Cancer Genome Data Classification

Satya Dev Pasupuleti ⬡ and Simone A. Ludwig *⬡

Department of Computer Science, North Dakota State University, Fargo, ND 58105, USA;
satya.pasupuleti@ndsu.edu
* Correspondence: simone.ludwig@ndsu.edu

Abstract

Cancer classification using high-dimensional genomic data presents significant challenges in feature selection, particularly when dealing with datasets containing tens of thousands of features. This study presents a new application of the Simultaneous Perturbation Stochastic Approximation (SPSA) method for feature selection on large-scale cancer datasets, representing the first investigation of the SPSA-based feature selection technique applied to cancer datasets of this magnitude. Our research extends beyond traditional SPSA applications, which have historically been limited to smaller datasets, by evaluating its effectiveness on datasets containing 35,924 to 44,894 features. Building upon established feature-ranking methodologies, we introduce a comprehensive evaluation framework that examines the impact of varying proportions of top-ranked features (5%, 10%, and 15%) on classification performance. This systematic approach enables the identification of optimal feature subsets most relevant to cancer detection across different selection thresholds. The key contributions of this work include the following: (1) the first application of SPSA-based feature selection to large-scale cancer datasets exceeding 35,000 features, (2) an evaluation methodology examining multiple feature proportion thresholds to optimize classification performance, (3) comprehensive experimental validation through comparison with ten state-of-the-art feature selection and classification methods, and (4) statistical significance testing to quantify the improvements achieved by the SPSA approach over benchmark methods. Our experimental evaluation demonstrates the effectiveness of the feature selection and ranking-based SPSA method in handling high-dimensional cancer data, providing insights into optimal feature selection strategies for genomic classification tasks.

Keywords: high dimensional data; classification models; SPSA; feature selection; machine learning; cancer genomics

1. Introduction

The information technology industry uses the buzzword Big Data for high-dimensional data with a large number of features. Big Data has three characteristics, which are simply called the 3Vs—volume, velocity, and variety. This means that Big Data has a large volume, a huge variety, and changes rapidly [1]. Data are divided into numerous categories based on the data size, and certain datasets up to sizes of 10 terabytes or more can be considered as Big Data. Most industries, such as the internet, biomedicine, and astronomy, have massive data with a great number of features [2]. Scaling large databases is a huge issue in Big Data

systems, as they usually contain lots of redundant and irrelevant data, which consumes computing resources and also contributes to performance reduction. Thus, it is important to reduce the unnecessary features and extract the necessary and valuable features in order to build good models based on this Big Data. The dimensionality reduction will lower the consumption of computer resources and also improve the model's performance [3].

There are algorithms that reduce the dimensionality of the data that can make the learning model more generalized and denser [4]. Dimensionality reduction is divided into two types—feature extraction and feature selection [5]. The feature extraction technique aims to convert high-dimensional data into a low-dimensional space. Features of low-dimensional data are a linear or nonlinear combination of the original features. The feature selection technique selects the best feature subsets from the original features using a certain process. Usually, feature extraction is said to improve the model performance, but they tend to compress and transform the original features, leading to data distortion and affecting the efficiency of data processing [6]. On the other hand, feature selection retains the semantic meaning of the original features and thus has better interpretability. In the feature selection technique, the most relevant features are chosen from the original dataset, whereas feature extraction creates new features by transforming the existing ones. This will reduce the cost of the feature collection [7].

Feature selection is divided into three categories—filter, wrapper, and embedded. The filter feature selection technique assumes that data is completely independent of the classifier algorithm and forms the subset of features according to their measurement of contribution to class attributes [8]. For the wrapper feature technique, the domain knowledge is needed, and a performance metric is employed on the classification algorithm for feature subset evaluation, and based on the results, it searches for an optimal feature subset [9]. The embedded feature selection technique incorporates feature selection into the learning process of the classifier, and then searches for a feature subset by a functional optimization that is designed in advance. The embedded technique thus deletes the features that have a minor influence on the outcome of the model result and only retains good features that are essential for the model result [10].

The rest of the paper is organized as follows: Section 2 discusses related work and highlights the shortcomings. Section 3 introduces and describes our proposed feature selection model as well as the comparison models, and also describes different classification models used in this research. Section 4 explains the experiment setup, statistical analysis, and results. Finally, Section 5 concludes the paper.

2. Related Work

Many researchers in the past have carried out experiments to use feature selection techniques and applied classification methods on reduced-dimensional data to improve the performance of the models. The authors in [11] proposed the integration of Gradient Boosting (GB), Random Forest (RF), Logistic Regression with Lasso Regularization, Logistic Regression with Ridge Regression, and SVM with the K-Means-Clustering-based feature selection method. They applied the proposed model to the Coimbra breast cancer dataset. The authors developed a gene expression-based cancer classification network in [12]. In this network, they used AlexNet-based transfer learning to extract the features and then used a hybrid fuzzy ranking network to rank and select the features and finally used a multi kernel Support Vector Machine for multiclass classification on colon, ovarian, and lymphography cancer data.

Traditional gradient-based methods such as Forward Selection, Backward Elimination, and Stepwise Regression represent classical approaches to feature selection [13,14]. However, these methods often struggle with high-dimensional datasets due to compu-

tational complexity and local optima issues [15]. Derivative-free optimization methods, including Random Search [16] and Bayesian Optimization [17], have gained attention for their ability to handle discrete and non-convex optimization landscapes typical in feature selection problems.

The Boruta method [18] represents a notable wrapper-based approach that utilizes Random Forest as a base classifier to identify relevant features. Boruta addresses the feature selection problem by comparing the importance of original features with randomly permuted shadow features, providing statistical significance testing for feature relevance. While Boruta has demonstrated effectiveness in identifying truly relevant features and handling feature interactions [19,20], the method is computationally heavy, particularly when applied to high-dimensional datasets with tens of thousands of features. The computational burden stems from the need to repeatedly train Random Forest models with augmented feature sets, including shadow features, making it less practical for large-scale genomic datasets [21].

Other ensemble-based methods include Recursive Feature Elimination (RFE) with various base classifiers [22], stability selection [23], and bootstrap-based feature ranking [24]. These methods often provide robust feature selection but at increased computational cost.

In [25], the authors integrated Binary Particle Swarm Optimization (BPSO) and Grey Wolf Optimizer (GWO) algorithm for feature selection on the Breast Cancer Wisconsin dataset. Another approach, introducing a guided PSO approach, was presented in [26]. In [27], the authors used the Krill Herd (KH) optimization algorithm to address problems in feature selection methods. They incorporated adaptive genetic operators to enhance the KH algorithm.

A Genetic Algorithm-based feature selection model (GA-FS) is proposed in [28] and was applied on a breast cancer dataset. The authors combined GA-FS with different classification models and compared the Accuracy before and after GA-FS. Authors in [29] proposed a two-stage feature selection method to classify colon cancer. In the filtering phase, they used ReliefF for feature ranking and then selected the best gene expression subset from 2000 features. Then finally applied the Support Vector Machine classifier for classifying colon cancer.

The authors applied Recursive Feature Selection (RFE) on different classification models to compare the performance of the models with regard to Accuracy, Precision, and F-measures [30]. The authors used five types of feature selection methods to classify gene expression datasets for ovarian, leukemia, and central nervous system (CNS) cancer in [31], and after discovering the minimal feature sets, applied five classifiers for classifying the data. In [32], authors proposed the Gradient Boosting Deep Feature Selection (GBDFS) algorithm to reduce the feature dimension of omics data, and thus, improved the classifier Accuracy of gastric cancer subtype classification.

James Spall introduced the Simultaneous Perturbation Stochastic Approximation (SPSA), which is a pseudo-gradient descent stochastic optimization algorithm in [33]. Initially, Spall introduced the SPSA method into the control area to tune a large number of neurons of a neural network controller with applications in a water treatment plant. In the beginning, SPSA was used in many successful applications in control problems, such as traffic signal control [34], robot arm control [35], etc.

In [36], the authors adopted Spall's SPSA approach for the first time to perform feature selection for a Nearest Neighbor classifier with the Minkowski distance metric for Artificial Nose and Golub Gene datasets. Later in [37], the authors introduced the concept of Binary SPSA (BSPSA). The feature selection problem is treated as a stochastic optimization problem where the features are represented as binary variables. BSPSA is used for feature selection on both small and large datasets. In [38], the authors proposed the Simultaneous Perturba-

tion Stochastic Approximation (SPSA) algorithm that mitigates the slow convergence issue of BSPSA in feature selection and feature ranking. The authors compared SPSA with the four wrapper methods on eight datasets (the largest dataset contains 2400 features) and further applied classification on the datasets using four classifiers' mean classification rates.

In [39], the authors also used SPSA with Barzilai and Borwein (BB) non-monotone gains on various public datasets with Nearest Neighbors Naïve Bayes classifiers as wrappers. They compared the proposed method with full features against seven popular meta-heuristics-based FS algorithms. SPSA-BB converges to a good feature set in about 50 iterations on average, regardless of the number of features (the largest dataset contains 1000 features). The authors in [40] generated subsets using Simultaneous Perturbation Stochastic Approximation (SPSA), migrating birds optimization, and Simulated Annealing algorithms. The subsets generated by the algorithms are evaluated by using correlation-based FS, and the performance of the algorithms is measured using a Decision Tree (C4.5) as the classifier. The computational experiments are conducted on the 15 datasets taken from the UCI machine learning repository. The authors concluded that the SPSA algorithm outperforms other algorithms in terms of Accuracy values, and all algorithms reduce the number of features by more than 50%.

The authors in [41] present SPFSR, a novel stochastic approximation approach for performing simultaneous k-best feature ranking (FR) and feature selection (FS) based on Simultaneous Perturbation Stochastic Approximation (SPSA) with Barzilai and Borwein (BB) non-monotone gains. The proposed method is performed on 47 public datasets, which contain both classification and regression problems, with the mean Accuracy reported from four different classifiers and four different regressors, respectively. The authors concluded that for over 80% of classification experiments and over 85% of regression experiments, SPFSR provided a statistically significant improvement or equivalent performance compared to existing, well-known FR techniques.

As seen by the related work in the paragraphs above, the SPSA method for feature selection has traditionally been applied to smaller datasets. In this research, we investigate its effectiveness on large-scale datasets used for cancer classification. Our approach builds on prior work, particularly [41], which employed feature ranking; however, we extended this by evaluating the impact of using varying proportions of the top-ranked features (5%, 10%, and 15%). Specifically, we apply feature selection and ranking via the SPSA method to datasets containing over 35,000 features (ranging from 35,924 to 44,894), with the goal of identifying features most relevant to cancer detection. To the best of our knowledge, this is the first study to apply the SPSA-based feature selection technique to such large cancer datasets. We conducted a comprehensive experimental evaluation and analysis, including comparisons with state-of-the-art feature selection and classification methods. Additionally, we assessed whether SPSA yields statistically significant improvements over ten benchmark methods.

3. Proposed Approach and Comparison Methods

In this section, we discuss our proposed methodology of feature selection based on the SPSA algorithm. Then, we discuss other popular feature selection models—RelChaNet, ReliefF, Genetic Algorithm, Mutual Information, Simulated Annealing, and Minimum Redundancy Maximum Relevance feature selection types that we are going to use to compare our SPSA feature selection method with. Further, we explain all the classification models we used in this research—Decision Tree, K-Nearest Neighbors, Light Gradient Boosting Machine, Logistic Regression, Support Vector Machine, and Extreme Gradient Boosting.

As illustrated in Figure 1, all ten cancer datasets were first divided into training (80%) and testing (20%) subsets. Feature selection was performed only on the training data

using all seven feature selection methods. From each training set, the top 5%, 10%, and 15% of features were selected, resulting in 30 reduced feature subsets across all datasets. These selected feature subsets were then applied to the corresponding test sets. Next, classification models were trained on the reduced training sets and evaluated on the held-out test sets. Performance was assessed using Accuracy, Precision, Recall, F1 Score, and Balanced Accuracy.
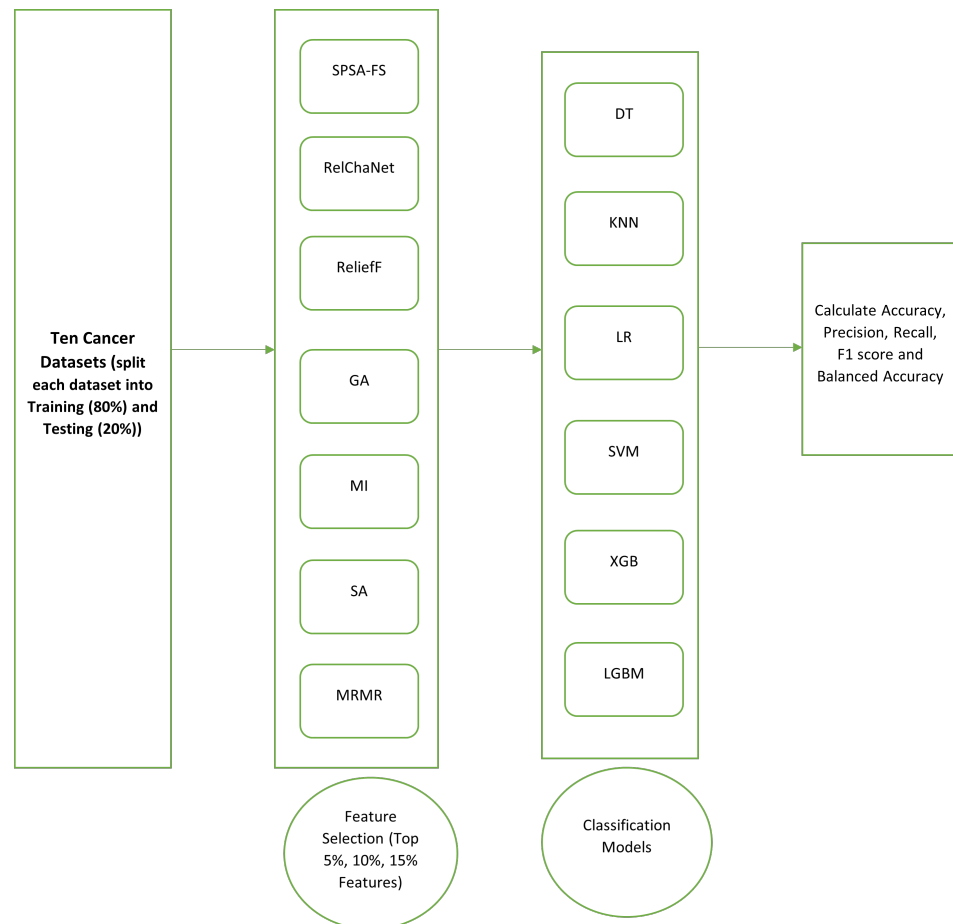
Figure 1. Overview of the proposed methodology workflow.

3.1. Proposed Methodology

3.1.1. Simultaneous Perturbation Stochastic Approximation (SPSA) Algorithm as Feature Selection (FS) Method

Spall introduced the Simultaneous Perturbation Stochastic Approximation (SPSA) [33], which is a pseudo-gradient descent stochastic optimization algorithm. The algorithm first starts with a random solution of a vector, and it gradually moves towards the optimal solution during iterations, where the current solution is perturbed simultaneously by offsets that are random and generated from a specific probability distribution.

Let us say $\mathcal{L} : \mathbb{R}^p \to \mathbb{R}$ is a real-valued objective function. Gradient descent search starts from an arbitrary initial solution and iteratively moves toward a local minimum of the objective function $\mathcal{L}$. At each step, the gradient of the objective function is evaluated, and the algorithm updates the solution in the direction of the negative gradient $-\nabla \mathcal{L}$. The process continues until converging to a local minimum, where the gradient is zero. In the language of machine learning, $\mathcal{L}$ can be called as loss function for the minimization problem. This gradient descent method cannot be applied where the loss function and loss function's gradient are unknown. Therefore, stochastic pseudo-gradient descent algorithms such as

SPSA are used, so that the gradient from noisy loss function measures is approximated and does not need the loss function information.

At each iteration κ, SPSA evaluates three noisy measurements of loss functions $\mathcal{Y}_\kappa^+$ and $\mathcal{Y}_\kappa^-$. $\mathcal{Y}_\kappa^+$ and $\mathcal{Y}_\kappa^-$ are used for gradient approximations, and $\mathcal{Y}(\hat{\mathcal{W}}_{\kappa+1})$ is used to measure the performance of the next iteration $\hat{\mathcal{W}}_{\kappa+1}$.

As per [33], the functions for tuning the parameters are shown in Equations (1) and (2):

$$a_{\mathcal{K}} := \frac{a}{(\mathcal{A} + \mathcal{K})^\alpha} \tag{1}$$

$$c_{\mathcal{K}} := \frac{c}{\gamma^{\mathcal{K}}} \tag{2}$$

where a, $\mathcal{A}$, α, c, and γ are algorithmic hyperparameters of SPSA. Here, a is the initial scaling constant for the step size, $\mathcal{A}$ is the stability constant that shifts the denominator to reduce large updates in early iterations, α is the decay rate of the step size, typically chosen in $(0.5, 1]$ for convergence guarantees, c is the initial perturbation constant controlling the magnitude of random offsets, and γ is the decay factor controlling how quickly perturbations decrease across iterations.

These parameters are dimensionless and require tuning for the problem at hand. Following the SPSA literature [33,42], we set the initial values via preliminary experiments, and then performed a sensitivity analysis by varying one parameter at a time within a reasonable range while keeping the others fixed. For each configuration, we ran 10 independent trials and reported the mean results.

SPSA does not have an automatic stopping rule; thus, we specify the maximum number of iterations as the stopping rule. The iteration sequence specified here as the stopping criterion must be monotone and satisfy the condition

$$\lim_{k \to \infty} a_k = 0.$$

Let us illustrate how the SPSA algorithm is used as a feature selection technique. Assume **X** is a data matrix with dimensions $n \times p$, where n represents observations and p represents features. Assume **Y** is a response vector with dimensions $n \times 1$. The vector $\{\mathbf{X}, \mathbf{Y}\}$ constitutes a dataset. Let $\{X = X_1, X_2 \ldots X_p\}$ denote a feature set where the jth feature in X is represented by X_j. For a subset that is non-empty, represented as $X' \subset X$, we define $\mathcal{L}_C(X', Y)$ as the true value of the performance criterion of a wrapper classifier denoted by C on the dataset. We train the classifier C albeit with an unknown $\mathcal{L}_C$ and compute the error rate denoted by $y_C(X', Y)$.

In this study, the wrapper classifier C is implemented as a linear Support Vector Machine (SVM) with class-weighted loss to handle imbalance across datasets. This choice is motivated by the small-n, large-p nature of our datasets ($p \approx 36{,}000$–$45{,}000$ features vs. $n \leq 600$ samples), where linear models with regularization provide stable estimates and reduce the risk of overfitting. For robustness, we also verified results using logistic regression with an elastic net penalty, which promotes sparsity while maintaining stability under correlated features.

Since the true error $\mathcal{L}_C$ is unknown, we instead compute the empirical error rate denoted by $y_C(X', Y)$, which can be expressed as $y_C = \mathcal{L}_C + \varepsilon$, where ε represents the noise arising from finite-sample estimation, variability in cross-validation splits, and stochastic elements of training. Thus, y_C serves as a noisy but unbiased estimate of $\mathcal{L}_C$, and SPSA leverages this noisy feedback in approximating the gradient. The feature selection problem

can therefore be defined by the non-empty feature set X^*, and it can be determined by Equation (3).

$$X^* = \arg\min_{X' \subset X} y_C(X', Y) \tag{3}$$

3.1.2. Barzilai–Borwein (BB) Method

According to the non-monotone methods concept, the non-monotone feature remembers the data provided by previous iterations. One of the first non-monotone search methods, the Barzilai–Borwein (BB) method, is described as a gradient method with a two-point step size [43]. With the motivation from Newton's method, the BB method targets to approximate the Hessian matrix instead of doing the direct computation. Thus, it computes the series of objective values that are decreasing monotonically, and hence, the BB method performs better than the classical steep-descent methods in terms of both performance and cost of computation.

A lot of research has happened in steepest descent methods like the BB method and the Cauchy method [44], and the research concluded that the convergence analysis of the BB method found to linearly converge in a convex quadratic form [45,46]. The famous BB methods, well studied in different research areas, are Cauchy BB and cyclic BB. Cauchy BB is the combination of BB and the Cauchy methods, which performs better than the original BB and reduces the computation complexity by half, but it includes the steepest descent method, whereas cyclic BB has an extra process that determines the appropriate cycle length. Due to this shortcoming, we use the original BB method with a smoothing effect for our SPSA feature selection (SPSA-FS) algorithm.

3.1.3. Using the BB Method in SPSA-FS

In the SPSA feature selection algorithm, we improved the speed of the convergence by adopting a non-monotone BB step size strategy. Let $\hat{w}_k \in \mathbb{R}^p$ denote the estimated parameter vector that is the feature weights at iterations k. $\hat{g}_k = \hat{g}(\hat{w}_k)$ represents the estimated gradient of objective function with respect to $\hat{w}_k$. The gradient estimates here are noisy because of the stochastic nature of the optimization; thus, we apply the smoothing to stabilize the updates.

The BB step size at iteration k, denoted $\hat{a}_k$, can be computed as shown in Equation (4). This approximates the inverse Hessian using differences between the gradients and consecutive parameter vectors without computing the second derivatives.

$$\hat{a}_k = \frac{(\hat{w}_k - \hat{w}_{k-1})^\top (\hat{g}_k - \hat{g}_{k-1})}{(\hat{g}_k - \hat{g}_{k-1})^\top (\hat{g}_k - \hat{g}_{k-1})} \tag{4}$$

For the reduction in the step size fluctuations, we smooth the step size by taking the average of window τ iterations. It is shown in Equation (5):

$$\hat{b}_k = \frac{1}{\tau} \sum_{n=k-\tau+1}^{k} \hat{a}_n \tag{5}$$

where $\hat{b}_k$ in the above equation is the smoothed step size at iteration k.

Likewise, to stabilize the gradient estimates, we average the current gradient with the previous m gradients as shown in Equation (6), where $\bar{g}_k$ is the smoothed gradient that is used to update $\hat{w}_k$.

$$\bar{g}_k = \frac{1}{m+1} \sum_{n=k-m}^{k} \hat{g}_n \tag{6}$$

By using the smoothed step size and gradient estimates, the SPSA algorithm achieves more stable estimates and converges faster, especially in the cases of optimizing complex or noisy functions.

As explained above, the SPSA algorithm is an iterative stochastic optimization algorithm, regardless of the number of features, that approximates the gradients of the objective function with only a few functional evolutions per iteration, which gives SPSA scalability, noise tolerance, global search tendency, and computational efficiency. SPSA works well for high-dimensional features without exponential cost. It handles noisy evaluation metrics better than deterministic methods. Also, SPSA perturbs multiple features simultaneously, which helps to avoid getting easily trapped in local optima. Finally, SPSA required fewer evaluations than the methods that compute gradients or the methods that evaluate fitness scores feature by feature. The hyperparameters used for SPSA are described in Table 1.

Table 1. Hyperparameters for SPSA.

Hyperparameter	Values
Number of iterations (max_iters)	500
Initial step size (a)	0.01
Perturbation size (c)	0.05
Learning rate decay	0.602 (alpha), 0.101 (gamma)

3.2. Feature Selection Algorithms for Comparison

3.2.1. Neural Network Feature Selection Using Relative Change Scores (RelChaNet)

Network Pruning is a technique that identifies the less relevant features or neurons and removes them. This technique has been extensively studied since 1988 [47]. Among the recent advances in the pruning technique, the Neural Network Feature Selection Using Relative Change Scores (RelChaNet) method builds upon these foundational concepts by measuring the induced change in network parameters to guide the feature selection [48]. The authors introduced a lightweight feature selection algorithm that uses the pruning of neurons and input layer regrowth of a dense neural network. The neuron pruning happens when a gradient sum metric measures the relative change that occurs in a network once the feature enters, and in the meantime, the neurons grow again randomly. Figure 2 illustrates the relative change score calculation that is embedded in RelChaNet.
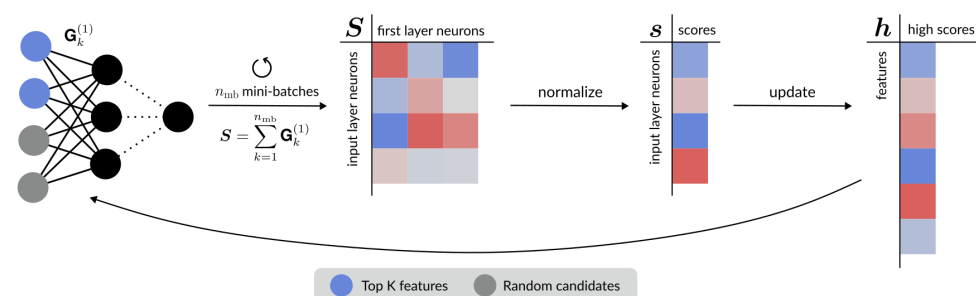
Figure 2. RelChaNet relative change scores calculation illustration [48].

Consider a neural network that consists of an input layer whose size is equal to the total number of features that need to be selected (K) and some candidate features. Multiple mini-batches are determined by the n_{mb} hyperparameter. The first layer gradients G^1 are combined in the matrix S. In the next step, this sum of gradients is normalized by the L^1 norm with regard to each input neuron. This is followed by applying Z-standards to the resulting vector, which produces the score vector s. The candidate scores are then used to update the high scores h.

Ultimately, all the features with K high scores remain in the network while the other features are redrawn randomly. Before training, the first layer weights are reinitialized, and the two hyperparameters used in RelChaNet adapt to the dataset characteristics fed to the network. RelChaNet overcomes the general drawbacks by allowing candidate features multiple mini-batches to demonstrate their relevance potential in the network and compares that relevance as an induced change rather than their absolute weights. The algorithm considers the network of the multi-layer perceptron with feed-forward architecture. This architecture is integrated into back-propagation training using the Adam optimizer.

Let us consider a dataset with features N, selected features K, and the number of hidden layer neurons n_{hidden}. The hyperparameters for the algorithm are the ratio of candidate features c_{ratio} considered at each iteration, and total mini-batches n_{mb}. Let us initialize the candidate features K_c with Equation (7).

$$K_c = round(c_{ratio}(N - K)) \tag{7}$$

The input layer size is calculated as the number of selected features K plus K_c. First, we choose the features randomly to populate the input layer; then, we start training the neural network that runs for n_{mb} mini-batches. The first-layer gradients are aggregated by addition. These gradient sums are normalized later, which results in a relative change score s_i, which is calculated by Equation (8).

$$s_i = \sum_{j=1}^{n_{hidden}} \|S_{ij}\| \, for \, i \in \{1, \dots, K + K_c\} \tag{8}$$

These scores are used for candidate features to update their high scores h. These features with high scores remain, and then new feature candidates are drawn randomly. This cycle will be repeated to gather as many features with a high score h, and also compare this score with incoming new features that are added in the next iterations. The hyperparameters used for RelChaNet are described in Table 2.

Table 2. Hyperparameters for RelChaNet.

Hyperparameter	Values
Number of layers	2
Hidden units per layer	50
Learning rate	0.01
Batch size	128
Number of epochs	200

3.2.2. ReliefF

ReliefF is a popular feature selection algorithm that is widely used in many industry data applications. It is a filter-based feature selection method that selects the best features by feature weight calculations [49]. Relief was proposed by Kira in their 1992 paper [50], and the proposed algorithm is limited to two-class classification problems. The Relief algorithm assigns different weights to the data features based on the correlation between classes and the features. The feature whose weight is greater than the set threshold will be selected as an important feature.

The main limitation of the Relief algorithm is that it only handles two-class classification problems. In 1994, Kononenko proposed the ReliefF algorithm, which is an extension of the Relief algorithm [51]. ReliefF can handle multiclass classification problems.

Let us assume that we have class labels of a certain training dataset $C = \{c_1, c_2, \dots, c_l\}$. A sample R_i is selected randomly from this training dataset; then, ReliefF searches for

K-Nearest Neighbors, which are also called hits of R_i from the same class, that is H_j ($j = 1, 2,$ ..., k), and hits of R_i from the different classes that is M_j(c) ($j = 1, 2, ..., k$). This procedure repeats m times. Therefore, the feature A weight is updated using Equation (9):

$$W(A) = W(A) - \sum_{j=1}^{k} \frac{diff(A,R_i,H_j)}{(m \times k)}$$
$$+ \sum_{c \notin class(R)} \sum_{j=1}^{k} \left[\frac{p(c)}{1-p(class(R))} \times \frac{diff(A,R_i,M_j(c))}{m \times k} \right] \tag{9}$$

where m is the total number of iterations and $diff(A, R_1, R_2)$ denotes the difference between samples R_1 and R_2 in feature A. This difference is calculated using Equation (10).

$$diff(A, R_1, R_2) = \begin{cases} \frac{|R_1[A] - R_2[A]|}{max(A) - min(A)} & \text{if } A \text{ is continuous} \\ 0, & \text{if } A \text{ is discrete and } R_1[A] = R_2[A] \\ 1, & \text{if } A \text{ is discrete and } R_1[A] \neq R_2[A] \end{cases} \tag{10}$$

The hyperparameters used for ReliefF are described in Table 3.

Table 3. Hyperparameters for ReliefF.

Hyperparameter	Values
Number of neighbors (k)	20
Sample size	full dataset

3.2.3. Genetic Algorithm Based Feature Selection Method (GA)

The Genetic Algorithm is an evolutionary algorithm that is inspired by the process of natural selection and genetics for finding optimal solutions in a vast solution space. According to the natural selection theory, the fittest individuals are selected, and then they are used to produce the offspring. The fittest parents' characteristics are passed on to the offspring using crossover and mutation for a better chance of survival. The GA contains two types of components. The first one defines the meta-parameters fitness function, selection strategy, crossover, mutation rates, and population size. The second component is an iterative evolutionary loop that applies the first component repeatedly so that it improves the population [52]. In this loop, the algorithm performs the following steps: (1) Fitness evaluation of each individual in the current population. (2) Select parents based on fitness values. (3) Offspring generation through crossover and mutation. (4) Producing the next generation. This evolutionary loop continues until a stopping criterion, such as the value of the maximum number of generations, is met.

Initial Population Generation

This is the first step in the GA implementation. The initial population consists of 50 chromosomes, each representing a randomly generated feature subset. For each chromosome, the genes are assigned randomly as 0 or 1, which indicates either exclusion or inclusion of corresponding features. In order to avoid redundancy, duplicates in the initial population are minimized. This results in diverse candidate feature subsets for the evolutionary algorithm to explore. The length of the chromosome is the total number of features in the subset.

Fitness Function

The fitness function evaluates each of the chromosome's feature subsets' quality based on the classification performance. In this step, we use the KNN (K-Nearest Neighbor)

accuracy, which is trained and tested on selected features. We set the KNN parameters as $k = 5$ and Euclidean distance as the metric. Before we proceed to classification, the features are normalized using z-score scaling for comparability across the dimensions. The fitness score of each chromosome is calculated as the average classification accuracy from 5-fold cross-validation on training data.

Selection

This is an important step in the process in which parent chromosomes are selected from the current population based on their fitness scores using tournament selection. In the selection method, a subset of individuals is selected randomly, and the fittest individual among them is selected as a parent. We selected tournament selection as a selection criterion, which ensures that the fittest individuals have more chances to get selected while also maintaining diversity. This selected parent then proceeds to the next step.

Crossover and Mutation

In this step, offspring chromosomes that form the next generation are generated by the crossover and mutation process. Here, we implement two-point crossover, where two crossover points are randomly selected along the parent chromosomes. The segments between points are swapped to produce offspring. Here, we set the crossover rate as 0.5, which means fifty percent of parents that are selected go through the crossover process, and the rest of the parents are unchanged. During the mutation process, random alterations are introduced to offspring for genetic diversity and to explore the search space. A mutation flips the individual genes with a mutation rate of 0.01 per gene. For example, if there are 40,000 features, then due to this mutation rate, approximately 400 mutations per chromosome per generation happen, which results in balancing the exploration.

Creating Next Generation

The next generation is created by the replacement of the whole current population with newly created offspring. This replacement will make sure that only new chromosomes will survive for the next iteration. Among all the individuals in the final generation, the most fitted chromosomes (the feature subset that yields the best classification accuracy) are selected as the optimal feature set.

Stopping Criterion

There should be a general stopping criterion for terminating the process of GA. Here, we used a fixed number of iterations as the stopping criterion, which we set to 20. Once the limit is reached, the GA execution is terminated, and the best-performing chromosome from the last generation returns as the selected feature subset.

The hyperparameters used for GA are described in Table 4.

Table 4. Hyperparameters for the Genetic Algorithm.

Hyperparameter	Values
Population size	50
Number of generations	50
Crossover rate	0.5
Mutation rate	0.01
Selection method	Tournament Selection
Total Iterations	20

3.2.4. Mutual Information-Based Feature Selection (MI)

The filter-based feature selection methods rank features based on their association with the target class. In simple filter approaches, the features are individually scored, and the features with high scores are selected. But in greedy methods, dependencies between features are considered by selecting features iteratively, which provides the highest incremental contribution, given that the features are already selected. At each step, the feature, together with the previously selected set, maximizes the relevance and is added. This process continues until the desired number of features is selected. In this research, the MI feature selection uses a simple ranking approach where features are ranked based on MI scores with class label, without considering dependencies among features.

MI of two random variables is a quantitative measurement of dependency between the variables [53]. This will be defined by the probability density function (PDF) of variables, say X, Y and joint (X, Y) as f_X, f_Y and $f_{X,Y}$, respectively, [54].

$$M\{X; Y\} = \iint f_{X,Y}(x, y) \log \left(\frac{f_{(X,Y)}(x, y)}{f_X(x) f_Y(y)} \right) d_x d_y \tag{11}$$

If the variables X and Y are completely independent, then the joint PDF is equal to the product of the PDF of X and the PDF of Y, which will be equated as $f_{X,Y} = f_X * f_y$, and MI becomes zero.

$$MI(X; Y) = 0 \tag{12}$$

Entropy is a measure of the uncertainty or randomness in a random variable. For a variable X, entropy is defined as:

$$h(X) = - \int f_X(x) \log f_X(x) dx \tag{13}$$

MI is expressed in terms of entropy as:

$$MI(X; Y) = h(Y) - h(Y|X) \tag{14}$$

where $h(Y/X)$ is the uncertainty of Y when X is known. If Y and X are independent, then $h(Y/X) = h(Y)$ and $MI(X; Y) = 0$.

In feature selection problems, usually the features X are continuous and the class label Y is discrete including ours. So, here we estimate the MI between the continuous features and discrete labels, which requires computing the conditional PDF of the feature given each class. This can be carried out using techniques like kernel density estimation or histogram binning [55]. The MI between X and Y with possible $\mathcal{Y}$ is calculated by Equation (15):

$$MI(X; Y) = \sum_{y \in \mathcal{Y}} P(Y = y) \int f_{X|Y}(x|y) \log \frac{f_{X|Y}(x|y)}{f_X(x)} dx \tag{15}$$

where $P(Y = y)$ is the prior probability of class y, $f_{X|Y}(x|y)$ is the conditional PDF of X given $Y = y$, and $f_X(x)$ is the marginal PDF of X. By using this estimation, features are ranked based on MI scores with a class label.The hyperparameters used for MI are described in Table 5.

Table 5. Hyperparameters for Mutual Information.

Hyperparameter	Values
Number of bins (discretization)	10
Estimator type	K-Nearest Neighbor
k-neighbors	10

3.2.5. Simulated Annealing (SA)

Simulated Annealing (SA) is a stochastic technique that is inspired by statistical mechanics. SA is used for finding globally optimal solutions to large optimization problems. The SA algorithm works with the assumption that some parts of the current solution belong to a potentially better one, and thus, these parts should be retained by exploring the current solution's neighbors. With the assumption of minimizing the objective function, SA jumps from hill to hill, and thus, escapes or avoids sub-optimal solutions. When a system, say S, contains a set of possible states in thermal equilibrium at a temperature T, the probability that it is in a certain state s is called $P_T(s)$. $P_T(s)$ depends on T and $E(s)$ of state s. That probability follows the Boltzmann distribution.

MI is expressed in terms of entropy as:

$$P_T(s) = \frac{exp\left(-\frac{E(s)}{kT}\right)}{Z} \tag{16}$$

where k is Boltzmann constant and Z acts as a normalization factor and it is defined as:

$$Z = \sum_{s \in S} exp\left(-\frac{E(s)}{kT}\right) \tag{17}$$

Consider s a current state as described above, and s' as neighboring state. The probability of the transition from s to s' can be formulated as:

$$P_T(s \to s') = \frac{P_T(s')}{P_T(s)} = exp\left(-\frac{\triangle E}{kT}\right) \tag{18}$$

where $\triangle E = E(S') - E(S)$.

If $P_T(s') \geq P_T(s)$, then the move is accepted and, if $P_T(s') < P_T(s)$, then the move is accepted with probability of $P_T(s, s') < 1$. The probability depends on the current temperature T, and it decreases when T does. There is a probability of T being lower at the end, in which the state is called the freezing point, and at this state, the transition is unlikely and the system is considered frozen. At this state, to increase the chance of maximizing the probability of finding the minimal energy state, thermal equilibrium should be reached. In order to reach equilibrium, annealing is scheduled to escape becoming stuck at a local minimum. Hence, the SA algorithm is introduced, and in SA, T is initially set at a high value to approximate thermal equilibrium. Then, small decrements of T are performed, and the process is iterated until the system is considered frozen. Reaching a near-optimal solution depends on how well the cooling schedule is designed, but it results in an inherently slow process because of the thermal equilibrium requirements at every point of temperature T.

To perform SA, we need four components to be performed. The components are configuration, move set, objective function, and cooling schedule. In the configuration step, the model represents all the possible solutions that the system can take, and then, it is used to find a near-optimal solution. Move sets are the computations that we need to perform to move from one state to another as part of the annealing process. The objective function is used to measure how good and optimal a given current state is. The cooling schedule

anneals the problem from a random solution to a good solution. This component schedules when to reduce the current temperature and when to stop the annealing process.

At the beginning of the SA process, an initial solution is selected randomly and is assumed to be an optimal solution. If T does not satisfy the termination condition, then the neighboring solution is selected and the cost is calculated for that solution. If the cost of the newly selected neighbor solution is less than or equal to the current optimal solution, then the current optimal solution is replaced by the newly selected neighbor solution [56].

The hyperparameters used for SA are described in Table 6.

Table 6. Hyperparameters for Simulated Annealing.

Hyperparameter	Values
Initial temperature	300
Cooling schedule	Exponential decay
Number of iterations	1000
Neighborhood size	5

3.2.6. Minimum Redundancy Maximum Relevance (MRMR)

The Minimum Redundancy Maximum Relevance (MRMR) method was first introduced by Ding and Peng in [57] to address redundancy problems with high dimensional and high-throughput datasets related to cancer. The MRMR method helps in identifying the features that are most relevant to the class labels and less redundant with respect to each other, and thus, results in improving the classification performance.

The MRMR algorithm works in a filter-based framework using Mutual Information (MI) to evaluate two criteria. The first criterion is relevance, which is quantified as Mutual Information $I(f;c)$ between a candidate feature f and the target class label c. The features are selected based on high Mutual Information. The second criterion is redundancy, which is defined as the average MI $I(f_i; f_j)$ between the candidate feature and each feature that has been selected already. The features are selected based on the minimum average MI to avoid the highly correlated (redundant) features [55].

At each step, a new feature is selected from the unselected features by maximizing the following condition:

$$Score(f) = I(f;c) - \frac{1}{|S|} \sum_{s \in S} I(f;s) \tag{19}$$

where S is the current feature set that is selected within the subset of size $|S|$. The MRMR optimization is formulated as:

$$MRMR(S) = \frac{1}{|S|} \sum_{f_i \in S} I(f_i; c) - \frac{1}{|S|^2} \sum_{f_i, f_j \in S} I(f_i; f_j) \tag{20}$$

The greedy incremental process continues until a predefined number of features are selected. The greedy incremental algorithm follows the steps below:

- Initialize S as empty;
- At each step, evaluate the remaining candidate features using the above score condition;
- Add the feature that maximizes the score;
- Repeat the process until the desired number of features is selected.

The hyperparameters used for MRMR are described in Table 7.

Table 7. Hyperparameters for MRMR.

Hyperparameter	Values
Number of features to select	5%, 10%, 15% of total features
Feature scoring function	Mutual Information (MI)

3.3. Classification Models

After generating the feature subsets from seven feature selection methods defined in the above section, we pass those feature subsets to different classification models in order to see how the features generated from different feature selection algorithms affect the classification model performance. The classifier's objective is to learn how to classify the objects by analyzing the dataset, where we know which classes the instance belongs to. We used six different classification models in this research, and we briefly describe each of the models in the subsections below.

3.3.1. Decision Tree (DT)

The instances are representations of attribute value vectors, and the input data that is fed to the classifiers has such types of vectors where each vector belongs to a class. The output typically consists of mapping from attribute values to classes, and with learning, the model is able to classify known and unseen instances [58].

The Decision Tree method is an example model of representation of mappings that contains attribute nodes linked to multiple sub-trees or leaves or decision nodes that are labeled with classes. The decision node computes the outcome or decision based on an attribute value, and each decision is associated with one sub-tree. In DT, an instance at the root node is classified, and the outcome of that node will be a sub-tree. This process will continue until the outcome of that instance is determined. The depth of the Decision Tree is based on how many sub-trees it was divided into, and this determines the total conditions used in the decision rules, which is not a fixed number [59]. The hyperparameters used for DT are described in Table 8.

Table 8. Hyperparameters for Decision Tree.

Hyperparameter	Values
max_depth	20
min_samples_split	10
min_samples_leaf	5
criterion	gini
max_features	sqrt

3.3.2. K-Nearest Neighbors (KNN)

KNN is one of the popular machine learning classification methods, which works on the principle of classifying unlabeled data based on its Nearest Neighbors. The concept of the KNN method was first proposed by Fix and Hodge in 1951 [60], and later developed by Cover and Heart in 1967 [61]. KNN is also used for prediction problems in which the label of the sample is predicted as the one with the majority label among its Nearest Neighbors.

KNN classifies the objects according to the distance between two samples. In general, the Euclidean distance formula is used to calculate the distance between two training or testing objects [62]. The formula is given in Equation (21).

$$d_{xy} = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \tag{21}$$

The hyperparameters used for KNN are described in Table 9.

Table 9. Hyperparameters for K-Nearest Neighbors.

Hyperparameter	Values
n_neighbors	10
weights	distance
metric	minkowski (p = 2 default)
leaf_size	20

3.3.3. Extreme Gradient Boosting (XGB)

The XGBoost algorithm, an ensemble learning method introduced by Chen and Guestrin [63] in 2016, improved Gradient Boosting by optimizing computational efficiency and scalability. XGB has been implemented in several programming languages and software libraries since its introduction and makes it accessible for performing both regression and classification tasks.

This method is a hybrid model of multiple base learners. It explores different base learners and picks a learning function that reduces the loss. The idea of 'ensembling' the additive models is to train the predictors sequentially, and correct the predecessor by fitting the new predictor to the residual errors made by the previous predictor. In each step, the model optimizes the parameters. The inference and training of this learning method can be expressed by Equations (22) and (23):

$$\theta^*, f^* = \arg\min_{\theta, f} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}\left[f^{(i)}(x^{(i)}; \theta), y^{(i)}\right] \tag{22}$$

$$\hat{p}(y|x; \theta^*) = f^*(\theta, x) \tag{23}$$

where θ and f represent both the model set and its parameter set. $\mathcal{L}$ is the model train loss function. $\hat{p}(y \mid x; \theta^*)$ is the predicted conditional probability of the output y given input x and optimized parameters θ^*. The hyperparameters used for XGB are described in Table 10.

Table 10. Hyperparameters for Extreme Gradient Boosting (XGB).

Hyperparameter	Value
Booster	gbtree
Learning rate	0.1
Maximum depth	6
Number of estimators	100
Subsample	0.8
Column sample by tree	0.8
Gamma	0
Minimum child weight	1
Regularization parameter (L2)	1
Regularization parameter (L1)	0

3.3.4. Logistic Regression (LR)

LR is used frequently for binary and linear classification tasks [64]. LR is used for estimating the probabilities of classes because it models the associations with the logistic data distribution. LR performs well with linearly separable classes, and this method is best used to identify the class decision boundaries [65]. This method focuses on the relationship between independent variables ($x_0, x_1, \ldots, x_n$) and a dependent variable Y. The logistic function, which is a sigmoid function, is used for the logistic model calculation. In this calculation, each value between negative infinity and positive infinity is generated as the range of input and output values between 0 and 1 [66].

Logistic regression is able to interpret the vector variables in the data and to conduct the coefficient or weight evaluation for each of those input variables, and in turn is able to predict the class that the vector belongs to. LR is the method used for the datasets in which the independent variables are known the results. The hyperparameters used for LR are described in Table 11.

Table 11. Hyperparameters for Logistic Regression.

Hyperparameter	Values
penalty	lasso
C	5
solver	lbfgs
max_iter	200
class_weight	balanced

3.3.5. Support Vector Machine (SVM)

SVM determines an optimal hyperplane that maximizes the margin between the hyperplane and the closest data point to that hyperplane. The patterns observed in the optimal hyperplane are called support vectors. Identifying a hyperplane that is optimal using SVM can be seen in Figure 3.
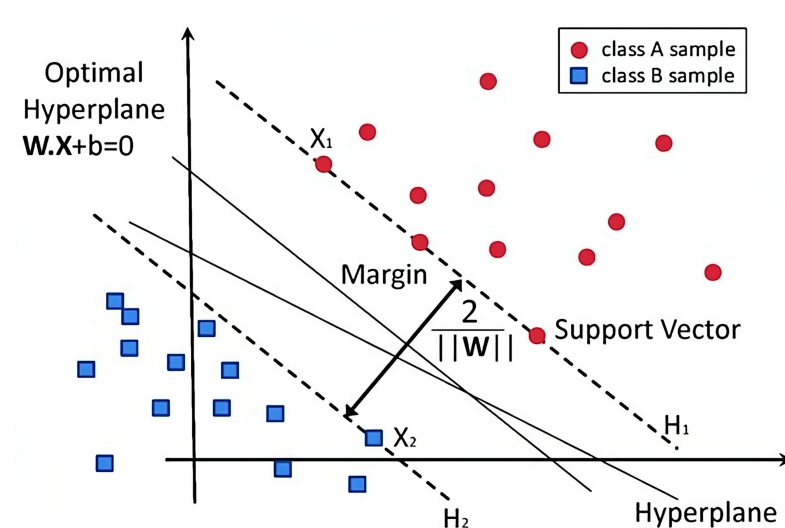
Figure 3. Support Vector Machine (SVM) data classification [67].

The SVM hyperplane is the set of points x satisfying Equation (24):

$$g(x) = w^T x + b = 0 \tag{24}$$

where w is the weight vector orthogonal to the hyperplane and b is an offset from the origin. In the case of linear SVMs, $g(x) \geq 0$ is considered the closest point of a class, and $g(x) < 0$ is considered as the closest point of another class [68].

The distance between two support vectors is defined by Equation (25):

$$d = \frac{2}{\|w\|^2} \tag{25}$$

where w is the optimal weight vector orthogonal to the seperating hyperplane, which is obtained through SVM optimization.

d should be increased for better separation, and proportionally *w* should be reduced using the Lagrange function in Equation (26):

$$L_p(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{l=1}^{n} \alpha_l \{ y_l (W^T x_l + b) - 1 \} \tag{26}$$

where $y_l(W^T x_l + b) \geq 1$ $i = 1, 2, \ldots, n$ and $y_l = \{+1, -1\}$ represent the class labels and α_l is the Lagrange multiplier. L_p should be reduced for optimal w and b computation. The hyperparameters used for SVM are described in Table 12.

Table 12. Hyperparameters for Support Vector Machine.

Hyperparameter	Values
C	5
kernel	rbf
gamma	scale
class_weight	balanced
max_iter	−1

3.3.6. Light Gradient Boosting Machine (LGBM)

LGBM is a variant method based on the Gradient Boosting Decision Tree (GBDT) and has better optimization than the XGB method. GBDT is an ensemble algorithm that combines multiple Decision Trees as base learners [69]. Each newly added tree pays increased attention to the samples that are misclassified by the previous trees. Through repetitive training of new Decision Trees and increasing their weights, GBDT gradually reduces model error and improves classification accuracy [70].

The LGBM uses the GBDT concept in its method, and the core concept involves sorting and bucketing the attributes to ensure all split points are checked. In training, LGBM selects leaf nodes for splitting and growing, and thus reduces the loss function. LGBM also introduces Gradient-Based One-Side Sampling (GOSS) to improve the effectiveness of model training. GOSS basically concentrates on the samples that have larger gradients, ignores the samples with low gradients, and amplifies the small gradient data weight. This process allows for effective utilization of large gradient samples and also retains some information from small gradient data samples that are disregarded [71].

The hyperparameters used for LGBM are described in Table 13.

Table 13. Hyperparameters for LightGBM classifier.

Hyperparameter	Values
num_leaves	128
max_depth	−1
learning_rate	0.01
n_estimators	100
min_data_in_leaf	50
feature_fraction	1.0
bagging_fraction	1.0
bagging_freq	1

3.4. Description of Datasets

The datasets used in this study were obtained from the Cancer Genome Atlas (TCGA) repository. TCGA is a cancer genomics program that characterizes 20,000 primary cancer and matched normal samples of 33 cancer types in total. This genomics program started in 2006 as a joint effort of both the National Cancer Institute and the National Human Genome Research Institute in the United States of America, and brought researchers from

several institutions together. Due to the efforts of this program, TCGA was able to create 2.5 petabytes of data of transcriptomics, epigenomics, genomics, and proteomics data [72].

From this website, we collected a total of 10 types of cancer genomic datasets to use in this paper: the Colon and Rectal Adenocarcinomas (COAD), Head and Neck Squamous Cell Carcinoma (HNSC), Kidney Chromophobe (KICH), Kidney Renal Papillary (KIRP), Liver Hepatocellular Carcinoma (LIHC), Lung Squamous Cell (LUSC), Prostate Adenocarcinoma (PRAD), Stomach Adenocarcinoma (STAD), Thyroid Cancer (THCA), and Uterine Corpus Endometrioid Carcinoma (UCEC). All datasets are high-dimensional data that have features between 35,924 to 44,878. We used the datasets without applying additional preprocessing or normalization steps. This decision was made to ensure that all feature selection methods were evaluated on identical input data, thereby isolating the effect of feature selection from any influence of preprocessing techniques. While data normalization and transformation are often applied in research studies, our focus was on the comparative evaluation of feature selection algorithms under a consistent setting.

The summary of each dataset is shown in Table 14.

Table 14. Summary of Cancer datasets.

Dataset	Samples	Features	Class 1	Class 2
COAD	522	37,677	509	13
HNSC	564	35,958	535	29
KICH	91	43,806	60	31
KIRP	322	44,874	236	86
LIHC	421	35,924	322	99
LUSC	553	44,894	494	59
PRAD	553	44,824	472	81
STAD	448	44,878	358	80
THCA	564	36,120	380	184
UCEC	588	36,086	345	243

3.5. Experiment Setup

We passed all ten cancer datasets and fed them as input to all seven feature selection methods that we considered in this research—SPSA (our proposed model), RelChaNet, ReliefF, GA, MI, SA, and MRMR. We selected the top 5%, 10%, and 15% of features from each of the ten datasets. We used all the 30 dimensionally reduced datasets (top 5%, 10%, and 15% feature subsets in each of the 10 cancer datasets) and passed them as input to six classification models and calculated the performance metrics. We divided each dataset for training and testing with a split of 80% and 20%, respectively.

3.6. Evaluation Metrics

To compare how the subsets of data produced by different feature selection algorithms as well as how they perform with the classification models, we considered performance metrics such as Accuracy, Precision, Balanced Accuracy, Recall, and F1 Scores [73].

Accuracy is the ratio of all classifications that are correct, whether they are negative or positive. It is calculated using Equation (27).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{27}$$

Recall is also known as the true positive rate, which calculates all positives that are classified as positives. It is calculated using Equation (28).

$$Recall = \frac{TP}{TP + FN} \tag{28}$$

Precision is the proportion of all positive classifications that are actually positive. It is calculated using Equation (29).

$$Precision = \frac{TP}{TP + FP} \tag{29}$$

F1 Score is a harmonic mean of both Precision and Recall, which balances both. This metric is preferred over Accuracy for class-imbalanced datasets. It is calculated using Equation (30).

$$F1 = \frac{2TP}{2TP + FP + FN} \tag{30}$$

Sensitivity is the same as Recall explained above, and Specificity is used to measure the proportion of True Negatives over the Total Negatives. It is calculated using Equation (31).

$$Specificity = \frac{TN}{TN + FP} \tag{31}$$

Balanced Accuracy is the arithmetic mean of sensitivity and specificity. It is used in cases of imbalanced data. It is calculated using Equation (32).

$$Balanced\,Accuracy = \frac{Sensitivity + Specificity}{2} \tag{32}$$

In all the above evaluation metric equations, TP is True Positive, TN is True Negative, FP is False Positive, and FN is False Negative.

3.7. Computational Resource Consumption Measurement

For the experiments, we used the Python (version 3.12.11) programming language to re-implement all the feature selection methods and classification models. For the experiments, we used the Center for Computationally Assisted Science and Technology (CCAST), an advanced computing infrastructure at North Dakota State University. We ran all the experiments on a JupyterLab setup of a 16-Core CPU with 128 GB memory with a GPU allocation. Table 15 provides the average execution runtime (in seconds) of the feature selection algorithms across all ten datasets, as well as the average execution runtime of the classifiers combined with the feature selection algorithms across the same datasets.

Table 15. Average total runtime (in seconds) for feature selection algorithms and classifiers, averaged across all ten datasets.

Task	Feature Selection Algorithms	Total Runtime (in Seconds)
Feature Selection Algorithms	SPSA	2108
	RelChaNet	2529
	ReliefF	2646
	GA	2123
	MI	1963
	SA	4986
	MRMR	7395
Feature Selection Algorithm + Classifiers	SPSA	3609
	RelChaNet	4050
	ReliefF	4171
	GA	3643
	MI	3460
	SA	8496
	MRMR	11,974

4. Results

We applied seven feature selection methods on the cancer datasets and extracted the top attributes that will help to improve the performance of the classification models. According to all the tables in the Appendix A, we can see that our model SPSA achieved mostly higher Balanced Accuracy compared to ReliefF, RelChaNet, GA, and MI in case of all the top 5%, 10%, and 15% of feature subsets.

For the DT classification, SPSA achieved a 100% Accuracy for COAD's and KICH's top 5, 10, 15 percent feature sets as shown in Tables A1, A2, and A3, respectively. SPSA often achieves near-perfect or perfect results across the datasets and different feature selection percentages, suggesting that SPSA is effective and robust in selecting features across different datasets. As for RelChaNet, its performance is generally good but varies more significantly between datasets. For example, Accuracy and F1 Score drop notably for certain datasets (e.g., THCA and UCEC). Regarding ReliefF, it shows competitive results, generally close to SPSA's performance, although there is some variation, such as slightly lower F1 Scores and Recall for certain datasets. GA scored average to very well in performance metrics with the small feature set (5%) with all the datasets, but it did not do better with the 10% and 15% feature sets. With the MI, the feature selection performance is good with the 10% feature set among most of the datasets, but with smaller and larger feature sets, it did not show good performance compared to other feature selection methods, and in some cases, it performed worse on most of the datasets. SA shows gradual improvement with higher feature subsets (10% and 15%). This feature selection method is slightly less stable in performance compared to SPSA or MI feature selection methods. MRMR is more consistent and has better values for Accuracy, Precision, F1 Score, and Balanced Accuracy, but lower and less consistent Recall scores for the 10% feature subset, and average values of Recall for the 15% feature subset. In particular, SPSA consistently ranks among the top methods for most datasets and subsets, with its strongest performance on COAD, KICH, LIHC, STAD, and THCA, while in a few cases (e.g., PRAD and UCEC), MRMR or SA slightly surpass its results. This shows that SPSA provides stable and reliable feature selection across subsets, often matching or outperforming traditional methods.

For the KNN classification, SPSA achieved 100% Accuracy for COAD's and KICH's top 5 and 10 percent feature sets as shown in Tables A4, A5, and A6, respectively. ReliefF also achieved 100% Accuracy for the same feature sets on COAD. SPSA shows consistently good values across different datasets, maintaining high classification performance even with a reduced number of features, whereas ReliefF shows slightly less consistent results, but still offers competitive performance and high Accuracy for many datasets. RelChaNet, on the other hand, tends to underperform, particularly with fewer features, and appears to be less robust across datasets compared to the other two methods. The GA feature selection method did not work well with KNN and scored low on all the datasets except for COAD, where it performed best. MI worked well with most of the small feature sets on all the datasets, but it performed occasionally average and, most of the time, worse than the 15% feature set. The results for SA indicate that the feature subsets that are generated have lower Accuracy and Precision compared to other methods in most cases, except for very few datasets where the Precision is high. The Balanced Accuracy is lowest across all feature subset sizes and the datasets. The MRMR feature subset performance is more variable. It shows high Accuracy and Recall for the KIRP and PRAD datasets, and for other datasets, it performed poorly for Balanced Accuracy and Recall. The Balanced Accuracy is frequently lower than SPSA and ReliefF. Overall, SPSA shows strong and consistent performance across datasets, especially at 5% features, where it often achieves near-perfect Accuracy and Recall compared to other methods. At 10% and 15%, it remains competitive, though

in some datasets, MRMR or GA slightly surpass it, indicating SPSA's advantage is most pronounced with smaller feature subsets.

For the LGBM classification, SPSA, RelChaNet, and ReliefF feature selection methods achieved 100% Accuracy for all COAD's feature sets as shown in Tables A7, A8, and A9, respectively. SPSA tends to be the most stable and highest-performing method across the different datasets and feature levels. RelChaNet occasionally showed variability and lower performance, suggesting potential dataset-specific challenges. ReliefF was comparable to SPSA in many cases, but occasionally showed variability, particularly for smaller feature sets. GA did not perform well on all datasets except for the COAD and KICH datasets, where it performed on par with SPSA on the 10% feature set. MI also worked well with COAD, but it performed poorly with the other datasets compared to the other feature selection methods. The SA feature subsets performed poorly for some datasets, like COAD and LUSC, at 15%, but were excellent in a few cases. The Balanced Accuracy is lowest among all feature selection methods, especially for the 15% feature subsets across the ten datasets. The MRMR feature subsets generated mixed results, where the Accuracy and Balanced Accuracy were lowest, especially for the 15% subset. Some datasets show decent Recall but poor F1 and Balanced Accuracy. Overall, SPSA demonstrates consistently strong performance across 5%, 10%, and 15% feature subsets, often matching or surpassing other methods in Accuracy and F1 while maintaining higher Balanced Accuracy. Unlike methods such as SA or MRMR that show variability, SPSA remains stable and reliable across all datasets.

For the LR classification, SPSA achieved 100% Accuracy for all KICH's top 5 and 10 percent feature sets, as shown in Tables A10, A11, and A12, respectively. SPSA emerges as the most robust method, maintaining high performance with fewer features. RelChaNet offers a good balance, with strengths at mid-level feature percentages, but some sensitivity to feature set size. ReliefF shows potential in specific datasets but lacks the broad consistency demonstrated by the other two methods. GA is effective on the LIHC dataset and showed good results with 10% and 15% feature sets for the majority of the datasets. MI also performed and achieved good scores on the KICH dataset; however, the performance was worse with most of the 10% and 15% feature sets. The SA feature subsets produced high Accuracy, Precision, Recall, and F1 Score results across most datasets. It frequently achieved perfect scores for the UCEC, KIRP, and HNSC datasets and has a strong Balanced Accuracy across all datasets. The MRMR feature subsets have very strong performance with perfect or near-perfect Precision, Recall, and F1 Scores. This feature selection method generated excellent results for the THCA, UCEC, HNSC, and PRAD datasets for all feature subsets. Overall, SPSA shows stable and competitive performance across all subsets. At 5%, it secures strong results—often surpassing RelChaNet, ReliefF, and MI—while remaining close to GA. At 10%, it maintains reliable Accuracy and Recall across datasets, outperforming weaker methods though occasionally behind SA and MRMR. By 15%, SPSA continues to deliver high scores, particularly in THCA and UCEC, confirming its robustness and consistent competitiveness with other leading methods.

For SVM classification, SPSA achieved 100% Accuracy for only COAD's top 5 percent feature set but outperformed ReliefF and RelChaNet for all the other datasets' feature sets as shown in Tables A13, A14, and A15, respectively. SPSA often leads in Precision and Recall metrics across most datasets, showcasing its ability to identify high-importance features that strongly influence classification. RelChaNet provides more stable but generally moderate performance, sometimes closing in on SPSA's results but also demonstrating variable results with feature changes. ReliefF's performance suggests it is less effective in filtering out critical features, leading to consistently lower performance results. GA did not perform well overall among most of the datasets, but it scored well on the PRAD dataset,

and with the 15% feature set of the STAD dataset. MI performed well with the 15% feature set on most datasets, but produced low Accuracy among the 5% and 10% feature sets for all datasets. The SA feature subsets are among the best performers, especially for the 15% subset, with an Accuracy of above 0.99 among the KIRP, PRAD, STAD, and UCEC datasets. The Precision, F1 Score, and Recall are high in many datasets, but the Balanced Accuracy is slightly behind SPSA and MRMR feature selection methods. The MRMR feature subsets have variable Accuracy and F1 Scores and are less consistent across the datasets. The Recall drops significantly, and the Balanced Accuracy achieved mixed results—with some datasets, it is good, but it is poor in others. Overall, SPSA shows strong and consistent performance across 5%, 10%, and 15% feature subsets. At 5%, it achieves near-perfect results on several datasets, clearly outperforming ReliefF, RelChaNet, and GA. With 10% features, SPSA maintains robust Accuracy and balanced results, especially in LIHC, THCA, and UCEC. At 15%, it remains competitive—particularly in THCA and UCEC—though SA and MRMR occasionally surpass it.

For the XGB classification, SPSA achieved 100% Accuracy for COAD's and KICH's top 5, 10, 15 percent feature sets, and THCA's top 5 percent feature set as shown in Tables A16, A17 and A18, respectively. SPSA generally achieves very high Accuracy and consistent metrics across the different datasets. RelChaNet displays lower Accuracy and performance metrics compared to SPSA and ReliefF in most cases. ReliefF generally performs on par with SPSA, often showing similarly high Accuracy and metrics across most datasets and feature percentages. Most of the time, GA did not perform well across most of the datasets, but performed well in Accuracy and Precision for the UCEC and COAD datasets. MI performed subpar across all datasets except KIRP, where it achieved a score as good as SPSA's. The SA feature subsets performed occasionally decently and achieved perfect values for the 15% feature subset for the LUSC dataset, but they have high inconsistency, often with poor Balanced Accuracy and Recall, and struggle with most of the datasets, which we can interpret due to the lack of robustness across the datasets. The MRMR feature subsets performed well for the PRAD dataset at 5%, and for STAD at 10%, but they have poor performance for Balanced Accuracy and Precision, especially for the KICH and KIRP datasets. Overall, SPSA consistently delivered strong and stable results across 5%, 10%, and 15% subsets, often matching or outperforming ReliefF, GA, and MI, with perfect scores in datasets like COAD, KICH, and THCA. While MRMR occasionally surpassed SPSA at higher subsets (e.g., PRAD, STAD), SPSA generally proved more reliable and robust than SA and RelChaNet, maintaining high Accuracy and balanced performance across datasets.

A colored heat map table representation of the Balanced Accuracy scores for each classification model used in this research across all ten cancer datasets is shown in Figure 4 for the 5% feature set. Please note that we opted to only display the best values, which are the 5% feature set results.

Classifier	Feature Selection	COAD	HNSC	KICH	KIRP	LIHC	LUSC	PRAD	STAD	THCA	UCEC
DT	SPSA	1.0000	0.9289	0.9722	0.8039	0.8575	0.7541	0.8503	0.8981	0.9197	0.8086
	RelChaNet	1.0000	0.9413	0.9444	0.7920	0.8678	0.8130	0.8609	0.7239	0.4375	0.7424
	ReliefF	1.0000	0.7920	0.7569	0.8535	0.8175	0.8280	0.8375	0.8852	0.8621	0.7943
	GA	0.9385	0.8575	0.9275	0.8075	0.7635	0.8459	0.8839	0.9305	0.8943	0.8462
	MI	0.9722	0.8130	1.0000	0.7974	0.8157	0.8459	0.8849	0.8823	0.8746	0.7460
	SA	0.8605	0.8394	0.8654	1.0000	0.8278	0.8185	0.8495	0.8629	0.8294	0.8293
	MRMR	1.0000	0.8647	0.8749	0.8654	0.9053	0.8195	0.9862	0.9721	0.9053	0.9586
KNN	SPSA	0.9275	0.8562	0.7653	0.7594	0.8056	0.8167	0.8276	0.8746	0.8953	0.8497
	RelChaNet	0.8647	0.8346	0.6589	0.6946	0.7590	0.7640	0.7649	0.8195	0.5043	0.6195
	ReliefF	0.9179	0.8296	0.7346	0.6358	0.8236	0.7934	0.7956	0.8257	0.8459	0.7058
	GA	0.8954	0.8523	0.6587	0.6595	0.7295	0.8158	0.7677	0.7727	0.8678	0.7698
	MI	0.8754	0.8426	0.8076	0.8023	0.7369	0.7798	0.8064	0.8403	0.8496	0.7596
	SA	0.8596	0.8821	0.8936	0.6743	0.7395	0.8295	0.8384	0.8547	0.7934	0.7397
	MRMR	0.8691	0.7195	0.8563	0.8745	0.7295	0.8736	0.7854	0.8469	0.8821	0.8596
LightGBM	SPSA	0.9265	0.9056	0.9670	0.8898	0.9284	0.8743	0.9076	0.8896	0.9048	0.8959
	RelChaNet	0.8769	0.8658	0.8047	0.8296	0.8429	0.8157	0.8645	0.8494	0.5296	0.8195
	ReliefF	0.8976	0.8368	0.8486	0.8285	0.8158	0.8168	0.8196	0.8494	0.8694	0.8195
	GA	0.8597	0.8459	0.8076	0.7986	0.8487	0.8185	0.8594	0.8396	0.8589	0.8789
	MI	0.8796	0.8495	0.8175	0.8538	0.8478	0.8276	0.8438	0.8754	0.8965	0.8478
	SA	0.8695	0.7296	0.8691	0.7296	0.6289	0.7285	0.8196	0.7296	0.7296	0.8591
	MRMR	0.7257	0.7504	0.7504	0.8318	0.8219	0.7968	0.7504	0.8659	0.8695	0.7089
LR	SPSA	0.8940	0.8056	0.9058	0.8495	0.8596	0.7598	0.7698	0.8578	0.8958	0.8768
	RelChaNet	0.8494	0.7089	0.8592	0.7285	0.7290	0.6459	0.7209	0.8395	0.4896	0.8589
	ReliefF	0.7495	0.5849	0.8305	0.8694	0.7056	0.4986	0.6792	0.8196	0.8749	0.8318
	GA	0.8539	0.7057	0.8596	0.7195	0.7098	0.6289	0.7296	0.8056	0.8695	0.8478
	MI	0.8945	0.6329	0.8974	0.7098	0.7109	0.7286	0.6589	0.8194	0.8967	0.8496
	SA	0.8295	0.7945	0.8168	0.8195	0.8175	0.8495	0.8296	0.8594	0.8494	0.8058
	MRMR	0.7945	0.8058	0.8657	0.8438	0.8068	0.9670	0.8478	0.8675	0.8896	0.8158
SVM	SPSA	0.9298	0.8594	0.8745	0.6849	0.8639	0.8275	0.8467	0.8943	0.8593	0.8823
	RelChaNet	0.8958	0.8417	0.6095	0.6295	0.7349	0.7594	0.7548	0.7689	0.5096	0.8536
	ReliefF	0.8386	0.8398	0.6285	0.6257	0.6349	0.7697	0.7386	0.6278	0.8469	0.7195
	GA	0.8756	0.8682	0.7598	0.6295	0.8193	0.7395	0.7632	0.7496	0.5837	0.8727
	MI	0.8396	0.8692	0.6894	0.6197	0.8295	0.7941	0.7943	0.6295	0.7975	0.8273
	SA	0.7267	0.6978	0.7950	0.7956	0.6495	0.7594	0.8257	0.8652	0.7246	0.7950
	MRMR	0.8560	0.8594	0.8974	0.7329	0.8248	0.8265	0.8056	0.8256	0.7549	0.8084
XGB	SPSA	1.0000	0.9586	1.0000	0.8365	0.8496	0.8946	0.8826	0.9286	0.9893	0.8891
	RelChaNet	0.9256	0.8932	0.8495	0.7947	0.8185	0.8175	0.8284	0.8391	0.5973	0.8592
	ReliefF	0.9862	0.9273	1.0000	0.8672	0.8564	0.8305	0.8594	0.8475	0.8647	0.8197
	GA	1.0000	0.8596	0.8175	0.7648	0.8395	0.8175	0.8485	0.8238	0.8465	0.8625
	MI	0.9589	0.9285	0.9486	0.8725	0.7963	0.8283	0.8593	0.8645	0.8392	0.8939
	SA	0.8678	0.8923	0.8843	0.8782	0.8086	0.9063	0.7943	0.8519	0.7829	0.8823
	MRMR	0.9166	0.8450	0.8462	0.8256	0.8349	0.9600	0.9166	0.8280	0.8130	0.9411

Figure 4. Heat map table representing Balanced Accuracy scores for all the classifiers on 5% Feature Subset.

4.1. Statistical Analysis

Next, we study the effects of the features selected from the seven feature selection methods on Accuracy among the ten cancer datasets. We used the Friedman test, which is a non-parametric statistical test, to determine whether or not there is a statistically significant difference between the paired treatments, where treatments are arranged in a randomized repeated-measure design.

We are doing this statistical test only for the top 5% feature selection data. The Friedman test for this research uses the following null and alternative hypotheses:

The null hypothesis (H_0): The seven feature selection methods used in this research have an equal effect on Accuracy among the ten cancer datasets.

The alternative hypothesis (H_a): At least one feature selection method used in this research has a different effect from the others based on Accuracy among the ten cancer datasets.

4.1.1. DT

First, we calculated the summary statistics of the DT balanced classifier Accuracy scores of all ten datasets after applying the feature selection algorithms. We visualized the above summary with a violin plot in Figure 5.

Later, we applied the Friedman test and we obtained the test statistic of 32.61, and the *p*-value of 0.00001. Since the *p*-value is less than 0.05, which is statistically significant, the null hypothesis should be rejected. Therefore, we have sufficient evidence to conclude that the type of feature selection method used leads to statistically significant differences in Accuracy scores between the ten cancer datasets.

Figure 5. Representation of DT's Accuracy scores as a violin plot.

Next, we performed the Nemenyi post hoc test to identify which feature selection methods have different effects on Accuracy. The Nemenyi post hoc test returns the following *p*-values for each pairwise comparison of means as shown in Table 16.

Table 16. *p*-values for each pairwise comparison of means for DT classifier.

	SPSA	RelChaNet	ReliefF	GA	MI	SA	MRMR
SPSA	1.000000	0.255021	0.830311	0.003006	0.001625	0.000173	0.679535
RelChaNet	0.255021	1.000000	0.967552	0.744477	0.645473	0.309844	0.994057
ReliefF	0.830311	0.967552	1.000000	0.184945	0.129608	0.031322	0.999975
GA	0.003006	0.744477	0.184945	1.000000	0.999999	0.994057	0.309844
MI	0.001625	0.645473	0.129608	0.999999	1.000000	0.998617	0.230008
SA	0.000173	0.309844	0.031322	0.994057	0.998617	1.000000	0.066581
MRMR	0.679535	0.994057	0.999975	0.309844	0.230008	0.066581	1.000000

At $\alpha = 0.05$, the pairs below have statistically significant differences in the Accuracy scores among the ten cancer datasets.

- SPSA vs. GA;
- SPSA vs. MI;
- SPSA vs. SA;
- ReliefF vs. SA.

4.1.2. KNN

The summary statistics of the KNN classifier balanced classifier scores of all ten datasets after the applied feature selection algorithms are visualized as a violin plot in Figure 6.

For the Friedman test, we obtained the test statistic as 34.37 and the *p*-value as 0.000006. Since the *p*-value is less than 0.05, which is statistically significant, the null hypothesis should be rejected. Therefore, we have sufficient evidence to conclude that the type of feature selection method used leads to statistically significant differences in Accuracy scores among the ten cancer datasets.
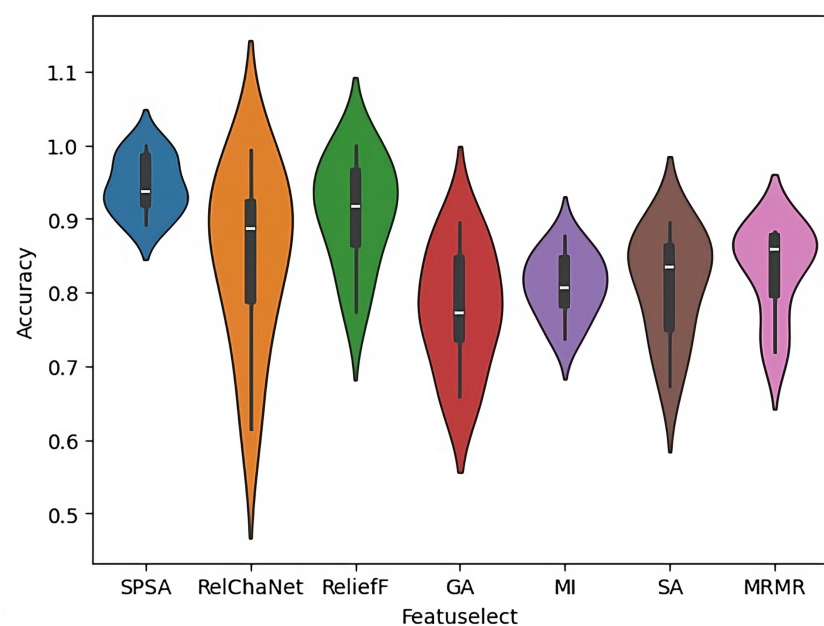
Figure 6. Representation of KNN's Accuracy scores as a violin plot.

Next, the Nemenyi post hoc test returns the following *p*-values for each pairwise comparison of means as shown in Table 17.

Table 17. *p*-values for each pairwise comparison of means for KNN classifier.

	SPSA	RelChaNet	ReliefF	GA	MI	SA	MRMR
SPSA	1.000000	0.042787	0.877672	0.000051	0.000439	0.005411	0.009475
RelChaNet	0.042787	1.000000	0.575532	0.610712	0.877672	0.996166	0.999241
ReliefF	0.877672	0.575532	1.000000	0.009475	0.042787	0.206646	0.281653
GA	0.000051	0.610712	0.009475	1.000000	0.999241	0.932179	0.877672
MI	0.000439	0.877672	0.042787	0.999241	1.000000	0.996166	0.987221
SA	0.005411	0.996166	0.206646	0.932179	0.996166	1.000000	0.999999
MRMR	0.009475	0.999241	0.281653	0.877672	0.987221	0.999999	1.000000

At $\alpha = 0.05$, the pairs below have statistically significant differences in the Accuracy scores among the ten cancer datasets.

- SPSA vs. RelChaNet;

- SPSA vs. GA;
- SPSA vs. MI;
- SPSA vs. SA;
- SPSA vs. MRMR;
- ReliefF vs. MI.

4.1.3. LGBM

The summary statistics of the LGBM classifier balanced classifier scores of all ten datasets after feature selection algorithms applied are visualized as a violin plot in Figure 7.

For the Friedman test, we obtained the test statistic as 47.06 and the *p*-value as 0.00000001. Since the *p*-value is less than 0.05, which is statistically significant, the null hypothesis should be rejected. Therefore, we have sufficient evidence to conclude that the type of feature selection method used leads to statistically significant differences in Accuracy scores among the ten cancer datasets.
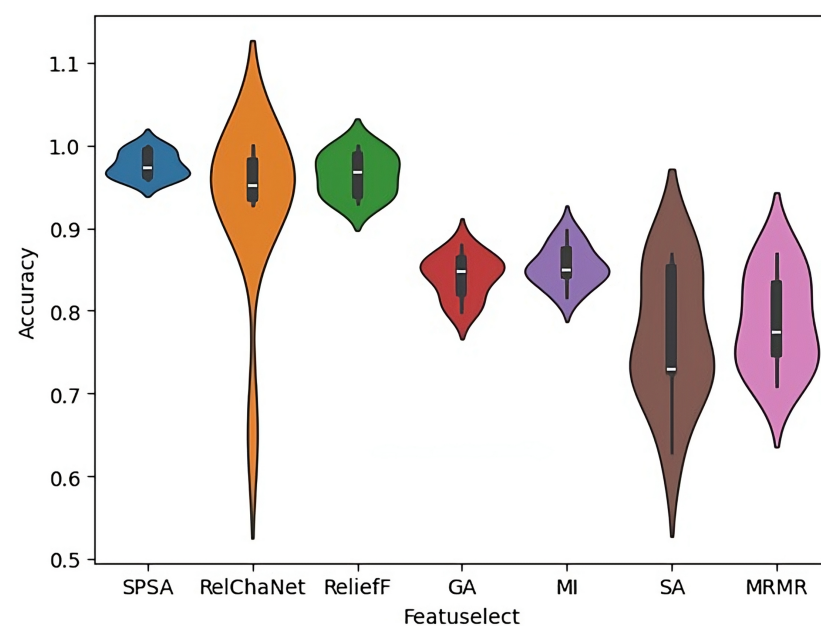
Figure 7. Representation of LGBM's Accuracy scores as a violin plot.

Next, the Nemenyi post hoc test returns the following *p*-values for each pairwise comparison of means as shown in Table 18.

Table 18. *p*-values for each pairwise comparison of means for LGBM classifier.

	SPSA	RelChaNet	ReliefF	GA	MI	SA	MRMR
SPSA	1.000000	0.679535	0.957723	0.001625	0.022624	0.000014	0.000024
RelChaNet	0.679535	1.000000	0.996166	0.230008	0.679535	0.013551	0.019133
ReliefF	0.957723	0.996166	1.000000	0.049754	0.281653	0.001316	0.002001
GA	0.001625	0.230008	0.049754	1.000000	0.991137	0.945976	0.967552
MI	0.022624	0.679535	0.281653	0.991137	1.000000	0.575532	0.645473
SA	0.000014	0.013551	0.001316	0.945976	0.575532	1.000000	1.000000
MRMR	0.000024	0.019133	0.002001	0.967552	0.645473	1.000000	1.000000

At $\alpha = 0.05$, the pairs below have statistically significant differences in Accuracy scores among the ten cancer datasets.

- SPSA vs. GA;

- SPSA vs. MI;
- SPSA vs. SA;
- SPSA vs. MRMR;
- RelChaNet vs. SA;
- RelChaNet vs. MRMR;
- ReliefF vs. GA;
- ReliefF vs. SA;
- ReliefF vs. MRMR.

4.1.4. LR

The summary statistics of the LR classifier balanced classifier scores of all ten datasets after the applied feature selection algorithms are visualized with a violin plot in Figure 8.
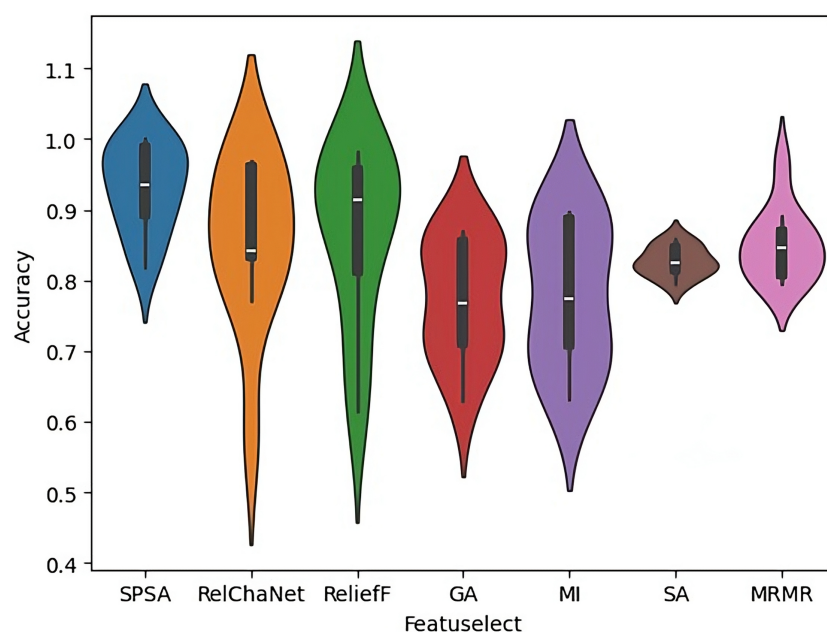
Figure 8. Representation of LR's Accuracy scores as a violin plot.

For the Friedman test, we obtained the test statistic as 30.59 and the *p*-value as 0.00003. Since the *p*-value is less than 0.05, which is statistically significant, the null hypothesis should be rejected. Therefore, we have sufficient evidence to conclude that the type of feature selection method used leads to statistically significant differences in Accuracy scores among the ten cancer datasets.

Next, the Nemenyi post hoc test returns the following *p*-values for each pairwise comparison of means as shown in Table 19.

Table 19. *p*-values for each pairwise comparison of means for the LR classifier.

	SPSA	RelChaNet	ReliefF	GA	MI	SA	MRMR
SPSA	1.000000	0.436086	0.230008	0.000051	0.000856	0.002001	0.049754
RelChaNet	0.436086	1.000000	0.999823	0.087853	0.339510	0.470241	0.957723
ReliefF	0.230008	0.999823	1.000000	0.206646	0.575532	0.712624	0.996166
GA	0.000051	0.087853	0.206646	1.000000	0.996166	0.982115	0.575532
MI	0.000856	0.339510	0.575532	0.996166	1.000000	0.999993	0.916230
SA	0.002001	0.470241	0.712624	0.982115	0.999993	1.000000	0.967552
MRMR	0.049754	0.957723	0.996166	0.575532	0.916230	0.967552	1.000000

At $\alpha = 0.05$, the pairs below have statistically significant differences in Accuracy scores among the ten cancer datasets.

- SPSA vs. GA;
- SPSA vs. MI;
- SPSA vs. SA;
- SPSA vs. MRMR.

4.1.5. SVM

The summary statistics of the SVM classifier balanced classifier scores of all ten datasets after feature selection algorithms applied are visualized with a violin plot in Figure 9.
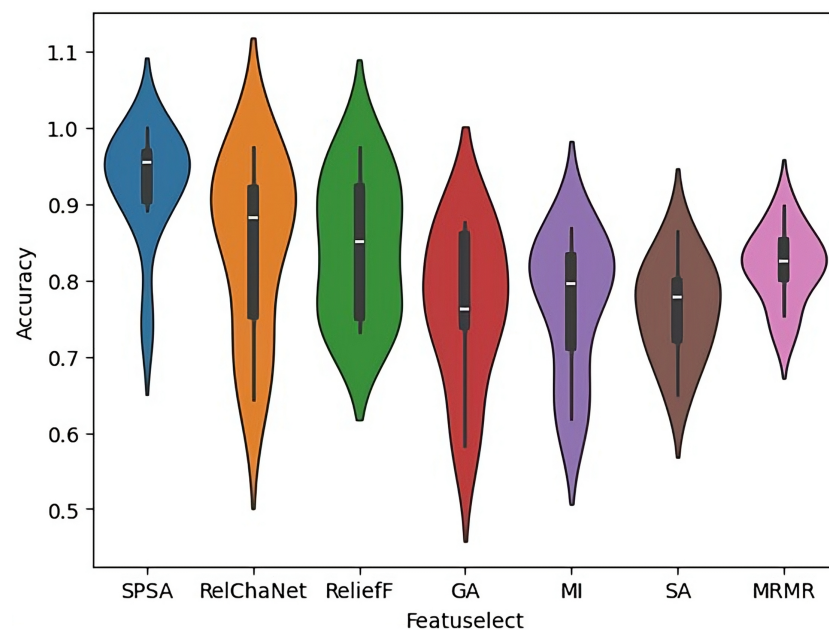
Figure 9. Representation of SVM's Accuracy scores as a violin plot.

For the Friedman test, we obtained the test statistic as 28.51 and the *p*-value as 0.000075. Since the *p*-value is less than 0.05, which is statistically significant, the null hypothesis should be rejected. Therefore, we have sufficient evidence to conclude that the type of feature selection method used leads to statistically significant differences in Accuracy scores among the ten cancer datasets.

Next, the Nemenyi post hoc test returns the following *p*-values for each pairwise comparison of means as shown in Table 20.

Table 20. *p*-values for each pairwise comparison of means for SVM classifier.

	SPSA	**RelChaNet**	**ReliefF**	**GA**	**MI**	**SA**	**MRMR**
SPSA	1.000000	0.470241	0.087853	0.000439	0.000439	0.001625	0.022624
RelChaNet	0.470241	1.000000	0.982115	0.230008	0.230008	0.402786	0.855064
ReliefF	0.087853	0.982115	1.000000	0.744477	0.744477	0.898068	0.999241
GA	0.000439	0.230008	0.744477	1.000000	1.000000	0.999928	0.945976
MI	0.000439	0.230008	0.744477	1.000000	1.000000	0.999928	0.945976
SA	0.001625	0.402786	0.898068	0.999928	0.999928	1.000000	0.991137
MRMR	0.022624	0.855064	0.999241	0.945976	0.945976	0.991137	1.000000

At $\alpha = 0.05$, the pairs below have statistically significant differences in Accuracy scores among the ten cancer datasets.

- SPSA vs. GA;
- SPSA vs. MI;
- SPSA vs. SA;
- SPSA vs. MRMR.

4.1.6. XGB

The summary statistics of the XGB classifier balanced classifier scores of all ten datasets after feature selection algorithms applied are visualized as a violin plot in Figure 10.
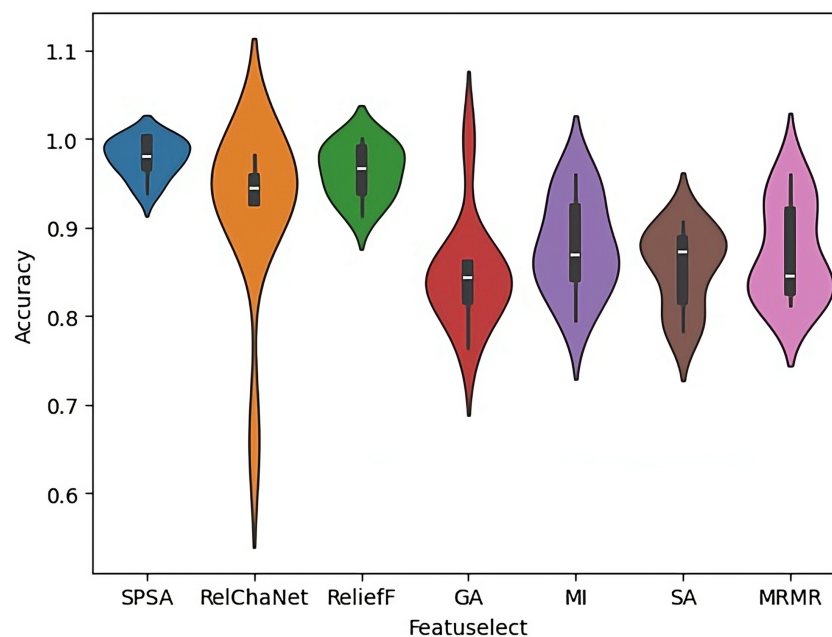
Figure 10. Representation of XGB's Accuracy scores as a violin plot.

For the Friedman test, we obtained the test statistic as 36.31 and the *p*-value as 0.000002. Since the *p*-value is less than 0.05, which is statistically significant, the null hypothesis should be rejected. Therefore, we have sufficient evidence to conclude that the type of feature selection method used leads to statistically significant differences in Accuracy scores among the ten cancer datasets.
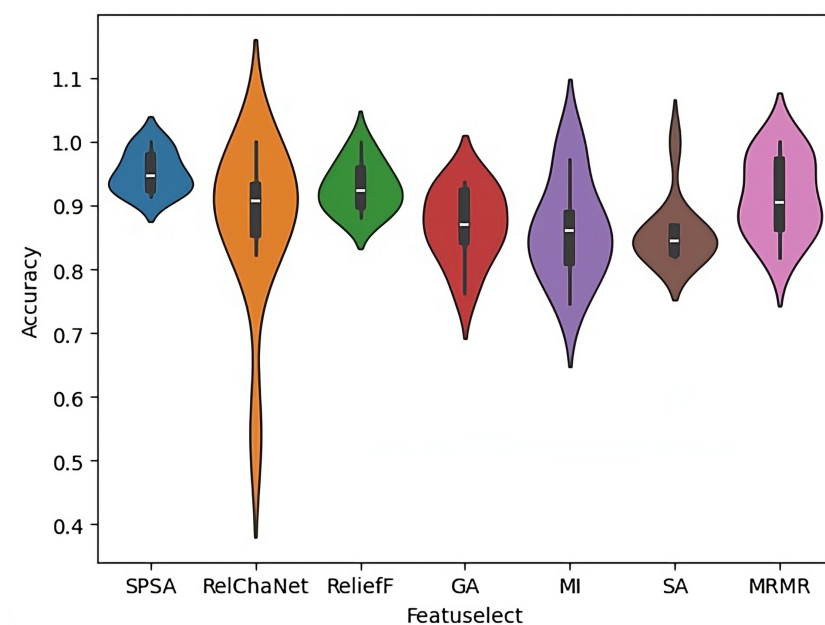
Next, the Nemenyi post hoc test returns the following *p*-values for each pairwise comparison of means as shown in Table 21.

Table 21. *p*-values for each pairwise comparison of means for the XGB classifier.

	SPSA	RelChaNet	ReliefF	GA	MI	SA	MRMR
SPSA	1.000000	0.339510	0.945976	0.000220	0.006543	0.000220	0.004462
RelChaNet	0.339510	1.000000	0.932179	0.255021	0.774849	0.255021	0.712624
ReliefF	0.945976	0.932179	1.000000	0.013551	0.146463	0.013551	0.114268
GA	0.000220	0.255021	0.013551	1.000000	0.982115	1.000000	0.991137
MI	0.006543	0.774849	0.146463	0.982115	1.000000	0.982115	1.000000
SA	0.000220	0.255021	0.013551	1.000000	0.982115	1.000000	0.991137
MRMR	0.004462	0.712624	0.114268	0.991137	1.000000	0.991137	1.000000

At $\alpha = 0.05$, the pairs below have statistically significant differences in Accuracy scores among the ten cancer datasets.

- SPSA vs. GA;
- SPSA vs. MI;

- SPSA vs. SA;
- SPSA vs. MRMR;
- ReliefF vs. GA;
- ReliefF vs. SA.

5. Conclusions

This research successfully demonstrated the effectiveness of the Simultaneous Perturbation Stochastic Approximation (SPSA) method for feature selection in large-scale cancer classification tasks, making an advancement in the application of the SPSA technique to high-dimensional genomic datasets. Our comprehensive experimental evaluation across datasets containing over 35,000 features establishes SPSA as a viable and superior alternative to existing feature selection methodologies for cancer detection applications. The experimental results provide compelling evidence for the efficacy of the SPSA-based approach. Through systematic evaluation using six diverse classification algorithms (Decision Trees, K-Nearest Neighbors, LightGBM, Logistic Regression, XGBoost, and Support Vector Machines), we demonstrated that SPSA-generated feature subsets consistently achieve superior classification performance compared to four state-of-the-art feature selection methods. Our approach yielded mostly higher and often perfect classification Accuracy across nearly all ten reduced-dimensional datasets, while maintaining competitive computational efficiency with average and frequently lower computation times.

The robustness of our findings is underscored by the comprehensive evaluation framework employing multiple performance metrics, including Accuracy, Balanced Accuracy, Precision, Recall, and F1 Score. The consistent advantage of SPSA-based feature selection across these diverse metrics and multiple classifier architectures validates the method's reliability and generalizability for high-dimensional cancer classification tasks.

Our investigation revealed that while SPSA maintains consistently high performance across most classifier combinations, there are isolated instances of reduced performance when integrated with certain classifiers. However, these cases represent minimal exceptions rather than systematic limitations, and the overall performance profile strongly favors the SPSA approach.

The successful application of SPSA to datasets exceeding 35,000 features establishes a new benchmark for feature selection in high-dimensional biomedical data analysis. We believe that researchers working with high-dimensional genomic, proteomic, or other biomedical datasets can leverage the SPSA-based feature selection method to significantly improve the Accuracy and reliability of their machine learning models.

This work opens several avenues for future research, including the exploration of hybrid approaches combining SPSA with other optimization techniques, the investigation of adaptive parameter tuning for different dataset characteristics, and the extension to multiclass cancer classification problems.

Author Contributions: Conceptualization, S.P.; Methodology, S.P.; Software, S.P.; Validation, S.P.; Formal analysis, S.P.; Investigation, S.P.; Resources, S.L.; Writing—original draft, S.P.; Writing—review and editing, S.P. and S.L.; Visualization, S.P.; Supervision, S.L. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The original data presented in the study are openly available at https://www.nature.com/articles/ng.2764#rightslink (accessed on 18 September 2025).

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Table A1. Decision Trees with 5% feature selection.

Dataset	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy
COAD	SPSA	1.0000	1.0000	1.0000	1.0000	1.0000
	RelChaNet	1.0000	1.0000	1.0000	1.0000	1.0000
	ReliefF	1.0000	1.0000	1.0000	1.0000	1.0000
	GA	1.0000	1.0000	1.0000	1.0000	0.9385
	MI	1.0000	1.0000	1.0000	1.0000	0.9722
	SA	0.9703	0.9664	0.9398	0.9524	0.8605
	MRMR	1.0000	1.0000	1.0000	1.0000	1.0000
HNSC	SPSA	0.9764	0.9879	0.7778	0.8510	0.9289
	RelChaNet	0.9588	0.7853	0.8734	0.8224	0.9413
	ReliefF	0.9705	0.8302	0.9320	0.8731	0.7920
	GA	0.9705	0.8302	0.9320	0.8731	0.8575
	MI	0.9647	0.8119	0.8765	0.8406	0.8130
	SA	0.9606	0.9502	0.9397	0.9448	0.8394
	MRMR	0.9823	0.9821	0.9775	0.9797	0.8647
KICH	SPSA	1.0000	1.0000	1.0000	1.0000	0.9722
	RelChaNet	0.9285	0.9167	0.9444	0.9251	0.9444
	ReliefF	0.9285	0.9167	0.9444	0.9251	0.7569
	GA	1.0000	1.0000	1.0000	1.0000	0.9275
	MI	1.0000	1.0000	1.0000	1.0000	1.0000
	SA	0.9823	0.9821	0.9775	0.9797	0.8654
	MRMR	0.9823	0.9778	0.9822	0.9799	0.8749
KIRP	SPSA	0.9278	0.8991	0.9263	0.9112	0.8039
	RelChaNet	0.8247	0.7803	0.7584	0.7679	0.7920
	ReliefF	0.9175	0.8829	0.9315	0.9016	0.8535
	GA	0.9278	0.8991	0.9263	0.9112	0.8075
	MI	0.8247	0.7803	0.7584	0.7679	0.7974
	SA	1.0000	1.0000	1.0000	1.0000	1.0000
	MRMR	0.9823	0.9821	0.9775	0.9797	0.8654
LIHC	SPSA	0.9212	0.8909	0.8909	0.8909	0.8575
	RelChaNet	0.9055	0.8754	0.8576	0.8659	0.8678
	ReliefF	0.8976	0.8497	0.8869	0.8656	0.8175
	GA	0.9055	0.8754	0.8576	0.8659	0.7635
	MI	0.9055	0.8754	0.8576	0.8659	0.8157
	SA	0.9481	0.9240	0.9120	0.9178	0.8278
	MRMR	0.9285	0.9500	0.9000	0.9181	0.9053
LUSC	SPSA	0.9277	0.8308	0.7643	0.7925	0.7541
	RelChaNet	0.9096	0.7702	0.7297	0.7476	0.8130
	ReliefF	0.9156	0.7819	0.7819	0.7819	0.8280
	GA	0.9277	0.8308	0.7643	0.7925	0.8459
	MI	0.9277	0.8308	0.7643	0.7925	0.8459
	SA	0.9448	0.9279	0.9179	0.9227	0.8185
	MRMR	0.9823	0.9313	0.8858	0.9071	0.8195
PRAD	SPSA	0.9457	0.9437	0.8298	0.8748	0.8503
	RelChaNet	0.9277	0.8726	0.8192	0.8428	0.8609
	ReliefF	0.8855	0.7681	0.7773	0.7726	0.8375
	GA	0.9277	0.8726	0.8192	0.8428	0.8839
	MI	0.9457	0.9437	0.8298	0.8748	0.8849
	SA	0.9285	0.9500	0.9000	0.9181	0.8495
	MRMR	1.0000	1.0000	1.0000	1.0000	0.9862

Table A1. *Cont.*

Dataset	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy
STAD	SPSA	0.9481	0.9146	0.9259	0.9201	0.8981
	RelChaNet	0.8962	0.8289	0.8796	0.8500	0.7239
	ReliefF	0.9407	0.9074	0.9074	0.9074	0.8852
	GA	0.9481	0.9146	0.9259	0.9201	0.9305
	MI	0.9481	0.9146	0.9259	0.9201	0.8823
	SA	0.9823	0.9821	0.9775	0.9797	0.8629
	MRMR	1.0000	1.0000	1.0000	1.0000	0.9721
THCA	SPSA	0.9647	0.9597	0.9597	0.9597	0.9197
	RelChaNet	0.5407	0.4891	0.4896	0.4889	0.4375
	ReliefF	0.9588	0.9650	0.9411	0.9518	0.8621
	GA	0.9647	0.9597	0.9597	0.9597	0.8943
	MI	0.9647	0.9597	0.9597	0.9597	0.8746
	SA	0.9457	0.9287	0.7744	0.8300	0.8294
	MRMR	0.9285	0.9500	0.9000	0.9181	0.9053
UCEC	SPSA	0.9152	0.9120	0.9136	0.9127	0.8086
	RelChaNet	0.8474	0.8423	0.8437	0.8429	0.7424
	ReliefF	0.8813	0.8815	0.8725	0.8762	0.7943
	GA	0.9152	0.9120	0.9136	0.9127	0.8462
	MI	0.8813	0.8815	0.8725	0.8762	0.7460
	SA	0.9690	0.9661	0.9545	0.9601	0.8293
	MRMR	0.9941	0.9969	0.9444	0.9690	0.9586

Table A2. Decision Trees with 10% feature selection.

Dataset	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy
COAD	SPSA	1.0000	1.0000	1.0000	1.0000	1.0000
	RelChaNet	0.9872	0.9935	0.7500	0.8301	1.0000
	ReliefF	1.0000	1.0000	1.0000	1.0000	0.9722
	GA	0.9872	0.9935	0.7500	0.8301	0.8949
	MI	0.9872	0.9935	0.7500	0.8301	0.8949
	SA	0.9448	0.9199	0.9294	0.9245	0.8517
	MRMR	0.9638	0.9805	0.8333	0.8901	0.8493
HNSC	SPSA	0.9764	0.9879	0.7778	0.8510	0.8764
	RelChaNet	0.9647	0.8119	0.8765	0.8406	0.7746
	ReliefF	0.9647	0.8045	0.9289	0.8542	0.8357
	GA	0.9705	0.8302	0.9320	0.8731	0.8678
	MI	0.9705	0.8302	0.9320	0.8731	0.8057
	SA	0.9285	0.9500	0.9000	0.9181	0.8926
	MRMR	0.9823	0.9313	0.8858	0.9071	0.8195
KICH	SPSA	1.0000	1.0000	1.0000	1.0000	0.9444
	RelChaNet	0.9642	0.9545	0.9722	0.9619	0.5000
	ReliefF	0.8928	0.8846	0.9167	0.8893	0.8295
	GA	0.9705	0.8302	0.9320	0.8731	0.8742
	MI	0.9647	0.8045	0.9289	0.8542	0.9257
	SA	0.9491	0.9505	0.9445	0.9472	0.9017
	MRMR	0.9882	0.9939	0.8889	0.9344	0.9273
KIRP	SPSA	0.9381	0.9141	0.9334	0.9230	0.8231
	RelChaNet	0.8556	0.8161	0.8161	0.8161	0.8130
	ReliefF	0.9072	0.8745	0.9001	0.8858	0.7954
	GA	0.9647	0.8045	0.9289	0.8542	0.7658
	MI	0.8556	0.8161	0.8161	0.8161	0.8549
	SA	1.0000	1.0000	1.0000	1.0000	0.9893
	MRMR	0.9823	0.9313	0.8858	0.9071	0.8836

Table A2. *Cont.*

Dataset	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy
LIHC	SPSA	0.9448	0.9380	0.9064	0.9208	0.9075
	RelChaNet	0.9291	0.8985	0.9076	0.9029	0.9063
	ReliefF	0.8976	0.8522	0.8754	0.8628	0.8175
	GA	0.8865	0.8499	0.8738	0.8604	0.8450
	MI	0.9291	0.8985	0.9076	0.9029	0.9592
	SA	0.9882	0.9939	0.8889	0.9344	0.9273
	MRMR	1.0000	1.0000	1.0000	1.0000	0.9483
LUSC	SPSA	0.9337	0.8435	0.7920	0.8149	0.6261
	RelChaNet	0.9277	0.8131	0.8131	0.8131	0.7920
	ReliefF	0.8433	0.6089	0.6194	0.6137	0.7585
	GA	0.9448	0.9380	0.9064	0.9208	0.8459
	MI	0.9277	0.8131	0.8131	0.8131	0.7850
	SA	1.0000	1.0000	1.0000	1.0000	0.9721
	MRMR	1.0000	1.0000	1.0000	1.0000	1.0000
PRAD	SPSA	0.9397	0.8782	0.8782	0.8782	0.8644
	RelChaNet	0.9216	0.8387	0.8504	0.8444	0.8711
	ReliefF	0.9096	0.8199	0.8087	0.8141	0.7936
	GA	0.8975	0.7359	0.7474	0.7415	0.8238
	MI	0.9216	0.8387	0.8504	0.8444	0.8936
	SA	1.0000	1.0000	1.0000	1.0000	1.0000
	MRMR	1.0000	1.0000	1.0000	1.0000	1.0000
STAD	SPSA	0.9555	0.9216	0.9444	0.9324	0.9166
	RelChaNet	0.9259	0.8777	0.8981	0.8873	0.8540
	ReliefF	0.9333	0.8857	0.9167	0.8999	0.8239
	GA	0.9216	0.8387	0.8504	0.8444	0.8346
	MI	0.9259	0.8777	0.8981	0.8873	0.8923
	SA	1.0000	1.0000	1.0000	1.0000	0.9382
	MRMR	1.0000	1.0000	1.0000	1.0000	1.0000
THCA	SPSA	0.9647	0.9752	0.9455	0.9584	0.9600
	RelChaNet	0.9470	0.9414	0.9372	0.9392	0.9259
	ReliefF	0.9529	0.9401	0.9557	0.9472	0.8852
	GA	0.9555	0.9216	0.9444	0.9324	0.7832
	MI	0.9529	0.9503	0.9415	0.9457	0.8861
	SA	0.9484	0.9392	0.9282	0.9335	0.8296
	MRMR	0.9823	0.9313	0.8858	0.9071	0.8195
UCEC	SPSA	0.8926	0.8887	0.8903	0.8895	0.8519
	RelChaNet	0.8644	0.8612	0.8581	0.8595	0.7210
	ReliefF	0.8587	0.8577	0.8492	0.8527	0.6954
	GA	0.8587	0.8577	0.8492	0.8527	0.7349
	MI	0.8587	0.8577	0.8492	0.8527	0.7239
	SA	0.9629	0.9611	0.9213	0.9396	0.8475
	MRMR	0.9491	0.9505	0.9445	0.9472	0.9017

Table A3. Decision Trees with 15% feature selection.

Dataset	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy
COAD	SPSA	1.0000	1.0000	1.0000	1.0000	0.9967
	RelChaNet	0.9936	0.9968	0.8750	0.9269	0.9289
	ReliefF	0.9936	0.9968	0.8750	0.9269	0.5000
	GA	0.9936	0.9968	0.875	0.9269	0.9500
	MI	0.9647	0.8119	0.8765	0.8406	0.9367
	SA	0.9882	0.9939	0.8889	0.9344	0.8593
	MRMR	0.9484	0.9392	0.9282	0.9335	0.8296

Table A3. *Cont.*

Dataset	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy
HNSC	SPSA	0.9764	0.8827	0.8827	0.8827	0.9289
	RelChaNet	0.9647	0.8045	0.9289	0.8542	0.8826
	ReliefF	0.9705	0.8302	0.9320	0.8731	0.8245
	GA	0.9647	0.8045	0.9289	0.8542	0.9063
	MI	0.9647	0.8119	0.8765	0.8406	0.9256
	SA	0.9448	0.9199	0.9294	0.9245	0.8517
	MRMR	1.0000	1.0000	1.0000	1.0000	0.9364
KICH	SPSA	1.0000	1.0000	1.0000	1.0000	1.0000
	RelChaNet	0.9285	0.9500	0.9000	0.9181	0.4972
	ReliefF	1.0000	1.0000	1.0000	1.0000	1.0000
	GA	0.9285	0.9500	0.9000	0.9181	0.8043
	MI	0.9285	0.9500	0.9000	0.9181	0.8623
	SA	0.9448	0.9199	0.9294	0.9245	0.8385
	MRMR	0.9527	0.9441	0.9230	0.9330	0.8496
KIRP	SPSA	0.9175	0.8949	0.8949	0.8949	0.8423
	RelChaNet	0.8556	0.8202	0.8039	0.8114	0.8130
	ReliefF	0.8865	0.8499	0.8738	0.8604	0.7423
	GA	0.8556	0.8161	0.8161	0.8161	0.8249
	MI	0.8865	0.8499	0.8738	0.8604	0.8256
	SA	0.9555	0.9306	0.9306	0.9306	0.8364
	MRMR	1.0000	1.0000	1.0000	1.0000	0.9591
LIHC	SPSA	0.9448	0.9279	0.9179	0.9227	0.9127
	RelChaNet	0.9212	0.8852	0.9024	0.8933	0.8575
	ReliefF	0.8976	0.8522	0.8754	0.8628	0.8349
	GA	0.9212	0.8852	0.9024	0.8933	0.8395
	MI	0.9448	0.9380	0.9064	0.9208	0.8409
	SA	0.9823	0.9313	0.8858	0.9071	0.8195
	MRMR	1.0000	1.0000	1.0000	1.0000	1.0000
LUSC	SPSA	0.9156	0.7795	0.8063	0.7920	0.8340
	RelChaNet	0.9096	0.7651	0.8273	0.7913	0.8374
	ReliefF	0.8975	0.7359	0.7474	0.7415	0.7836
	GA	0.9156	0.7819	0.7819	0.7819	0.8035
	MI	0.9277	0.8308	0.7643	0.7925	0.7829
	SA	0.9941	0.9969	0.9444	0.9690	0.9586
	MRMR	0.9698	0.9830	0.8958	0.9332	0.8196
PRAD	SPSA	0.9397	0.8782	0.8782	0.8782	0.8538
	RelChaNet	0.9337	0.8703	0.8574	0.8637	0.8782
	ReliefF	0.9216	0.8387	0.8504	0.8444	0.8579
	GA	0.9216	0.8387	0.8504	0.8444	0.8592
	MI	0.9216	0.8387	0.8504	0.8444	0.8936
	SA	0.9823	0.9313	0.8858	0.9071	0.8836
	MRMR	0.9629	0.9611	0.9213	0.9396	0.8475
STAD	SPSA	0.9555	0.9216	0.9444	0.9324	0.9351
	RelChaNet	0.9407	0.8996	0.9213	0.9099	0.9259
	ReliefF	0.9481	0.9240	0.9120	0.9178	0.8720
	GA	0.9481	0.9240	0.9120	0.9178	0.8823
	MI	0.9481	0.9240	0.9120	0.9178	0.8843
	SA	0.9448	0.9199	0.9294	0.9245	0.8385
	MRMR	0.9882	0.9939	0.8889	0.9344	0.9656
THCA	SPSA	0.9529	0.9503	0.9415	0.9457	0.9411
	RelChaNet	0.9352	0.9245	0.9285	0.9264	0.7943
	ReliefF	0.9352	0.9245	0.9285	0.9264	0.8345
	GA	0.9352	0.9245	0.9285	0.9264	0.8289
	MI	0.9529	0.9503	0.9415	0.9457	0.8340
	SA	0.9851	0.9769	0.9769	0.9769	0.8296
	MRMR	1.0000	1.0000	1.0000	1.0000	1.0000

Table A3. *Cont.*

Dataset	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy
UCEC	SPSA	0.8870	0.8887	0.8773	0.8818	0.8416
	RelChaNet	0.8587	0.8542	0.8635	0.8565	0.7214
	ReliefF	0.8870	0.8821	0.8875	0.8843	0.7529
	GA	0.8870	0.8821	0.8875	0.8843	0.7560
	MI	0.8870	0.8821	0.8875	0.8843	0.7540
	SA	0.9638	0.9805	0.8333	0.8901	0.8493
	MRMR	0.9823	0.9313	0.8858	0.9071	0.8195

Table A4. K-Nearest Neighbors with 5% feature selection.

Dataset	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy
COAD	SPSA	1.0000	1.0000	1.0000	1.0000	0.9275
	RelChaNet	0.9936	0.9968	0.8750	0.9269	0.8647
	ReliefF	1.0000	1.0000	1.0000	1.0000	0.9179
	GA	1.0000	1.0000	1.0000	1.0000	0.8954
	MI	1.0000	1.0000	1.0000	1.0000	0.8754
	SA	0.9745	0.4873	0.5000	0.4935	0.8596
	MRMR	0.9588	0.9792	0.6111	0.6712	0.8691
HNSC	SPSA	0.9882	0.9939	0.8889	0.9344	0.8562
	RelChaNet	0.9705	0.9849	0.7222	0.8000	0.8346
	ReliefF	0.9882	0.9939	0.8889	0.9344	0.8296
	GA	0.9705	0.9849	0.7222	0.8000	0.8523
	MI	0.9705	0.9849	0.7222	0.8000	0.8426
	SA	0.9647	0.9692	0.9502	0.9589	0.8821
	MRMR	0.8474	0.8477	0.8355	0.8401	0.7195
KICH	SPSA	0.8928	0.8918	0.8722	0.8805	0.7653
	RelChaNet	0.7857	0.7667	0.7667	0.7667	0.6589
	ReliefF	0.8571	0.8625	0.8222	0.8363	0.7346
	GA	0.7857	0.7667	0.7667	0.7667	0.6587
	MI	0.9285	0.9500	0.9000	0.9181	0.8076
	SA	0.9717	0.9718	0.9698	0.9708	0.8936
	MRMR	0.9548	0.9577	0.9493	0.9530	0.8563
KIRP	SPSA	0.8969	0.8852	0.8443	0.8616	0.7594
	RelChaNet	0.8144	0.7788	0.7148	0.7355	0.6946
	ReliefF	0.7731	0.7095	0.6988	0.7036	0.6358
	GA	0.8144	0.7788	0.7148	0.7355	0.6595
	MI	0.9175	0.9033	0.8827	0.8922	0.8023
	SA	0.8074	0.9030	0.5185	0.4820	0.6743
	MRMR	0.9642	0.9737	0.9500	0.9602	0.8745
LIHC	SPSA	0.9291	0.9398	0.8615	0.8927	0.8056
	RelChaNet	0.8740	0.9012	0.7448	0.7876	0.7590
	ReliefF	0.9055	0.9236	0.8115	0.8506	0.8236
	GA	0.8503	0.8826	0.6948	0.7340	0.7295
	MI	0.8740	0.9012	0.7448	0.7876	0.7369
	SA	0.8975	0.9485	0.5278	0.5255	0.7395
	MRMR	0.8554	0.4277	0.5000	0.4610	0.7295
LUSC	SPSA	0.9216	0.8908	0.6633	0.7188	0.8167
	RelChaNet	0.8975	0.7534	0.5766	0.6029	0.7640
	ReliefF	0.9156	0.8370	0.6599	0.7079	0.7934
	GA	0.9216	0.8908	0.6633	0.7188	0.8158
	MI	0.9096	0.8565	0.6077	0.6496	0.7798
	SA	0.9661	0.9650	0.9650	0.9650	0.8295
	MRMR	0.9647	0.9820	0.6667	0.7409	0.8736

Table A4. *Cont.*

Dataset	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy
PRAD	SPSA	0.9397	0.9010	0.8436	0.8690	0.8276
	RelChaNet	0.8975	0.8236	0.7151	0.7530	0.7649
	ReliefF	0.9156	0.8891	0.7430	0.7920	0.7956
	GA	0.9156	0.8891	0.7430	0.7920	0.7677
	MI	0.9216	0.9581	0.7292	0.7924	0.8064
	SA	0.8975	0.9485	0.5278	0.5255	0.8384
	MRMR	0.8975	0.9465	0.6458	0.6976	0.7854
STAD	SPSA	0.9629	0.9371	0.9491	0.9429	0.8746
	RelChaNet	0.9259	0.9035	0.8565	0.8773	0.8195
	ReliefF	0.9481	0.9358	0.8981	0.9154	0.8257
	GA	0.8880	0.8654	0.7639	0.8000	0.7727
	MI	0.9629	0.9371	0.9491	0.9429	0.8403
	SA	0.9294	0.9528	0.8909	0.914	0.8547
	MRMR	0.9294	0.9449	0.8957	0.915	0.8469
THCA	SPSA	0.9882	0.9915	0.9818	0.9864	0.8953
	RelChaNet	0.6148	0.5537	0.5424	0.5382	0.5043
	ReliefF	0.9647	0.9692	0.9502	0.9589	0.8459
	GA	0.9882	0.9866	0.9866	0.9866	0.8678
	MI	0.9647	0.9692	0.9502	0.9589	0.8496
	SA	0.8897	0.9378	0.7667	0.8067	0.7934
	MRMR	0.9647	0.9692	0.9502	0.9589	0.8821
UCEC	SPSA	0.9322	0.9483	0.9178	0.9280	0.8497
	RelChaNet	0.7175	0.8377	0.6575	0.6427	0.6195
	ReliefF	0.8248	0.8852	0.7877	0.8004	0.7058
	GA	0.9096	0.9333	0.8904	0.9027	0.7698
	MI	0.8757	0.9127	0.8493	0.8635	0.7596
	SA	0.8814	0.9001	0.7176	0.7652	0.7397
	MRMR	0.9745	0.4873	0.5000	0.4935	0.8596

Table A5. K-Nearest Neighbors with 10% feature selection.

Dataset	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy
COAD	SPSA	1.0000	1.0000	1.0000	1.0000	0.9275
	RelChaNet	0.9936	0.9968	0.8750	0.9269	0.8974
	ReliefF	1.0000	1.0000	1.0000	1.0000	0.9195
	GA	1.0000	1.0000	1.0000	1.0000	0.8678
	MI	1.0000	1.0000	1.0000	1.0000	0.8659
	SA	0.8370	0.7511	0.6898	0.7118	0.7937
	MRMR	0.8975	0.9485	0.5278	0.5255	0.8384
HNSC	SPSA	0.9882	0.9939	0.8889	0.9344	0.8434
	RelChaNet	0.9823	0.9909	0.8333	0.8954	0.8534
	ReliefF	0.9764	0.9879	0.7778	0.8510	0.8436
	GA	0.9823	0.9909	0.8333	0.8954	0.8734
	MI	0.9882	0.9939	0.8889	0.9344	0.8256
	SA	0.9294	0.9449	0.8957	0.9150	0.8075
	MRMR	0.9588	0.9792	0.6111	0.6712	0.8296
KICH	SPSA	0.9285	0.9500	0.9000	0.9181	0.8076
	RelChaNet	0.7857	0.7667	0.7667	0.7667	0.6589
	ReliefF	0.8214	0.8099	0.7944	0.8009	0.7068
	GA	0.8928	0.8918	0.8722	0.8805	0.7824
	MI	0.8571	0.8444	0.8444	0.8444	0.7246
	SA	0.9745	0.4873	0.5000	0.4935	0.8396
	MRMR	0.8556	0.8351	0.7795	0.8005	0.7847

Table A5. *Cont.*

Dataset	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy
KIRP	SPSA	0.8762	0.9008	0.7814	0.8177	0.7329
	RelChaNet	0.8453	0.8541	0.7359	0.7681	0.7267
	ReliefF	0.8144	0.7788	0.7148	0.7355	0.6926
	GA	0.7731	0.7095	0.6988	0.7036	0.6106
	MI	0.8144	0.8025	0.6904	0.7166	0.6967
	SA	0.7319	0.3660	0.5000	0.4226	0.6057
	MRMR	0.9527	0.9709	0.9000	0.9294	0.8295
LIHC	SPSA	0.9212	0.9195	0.8564	0.8824	0.8396
	RelChaNet	0.8503	0.8826	0.6948	0.7340	0.7496
	ReliefF	0.9055	0.9236	0.8115	0.8506	0.7560
	GA	0.8897	0.9126	0.7782	0.8202	0.7649
	MI	0.8503	0.8826	0.6948	0.7340	0.7296
	SA	0.9647	0.9820	0.6667	0.7409	0.8594
	MRMR	0.9647	0.9820	0.6667	0.7409	0.8736
LUSC	SPSA	0.9156	0.8370	0.6599	0.7079	0.7967
	RelChaNet	0.8915	0.6982	0.5244	0.5212	0.7814
	ReliefF	0.9036	0.7714	0.6288	0.6662	0.7745
	GA	0.9156	0.8370	0.6599	0.7079	0.7893
	MI	0.9216	0.8908	0.6633	0.7188	0.7935
	SA	0.9588	0.9792	0.6111	0.6712	0.8691
	MRMR	0.8504	0.9212	0.6834	0.6784	0.7694
PRAD	SPSA	0.9397	0.9010	0.8436	0.8690	0.8315
	RelChaNet	0.9156	0.9551	0.7083	0.7706	0.8064
	ReliefF	0.9036	0.8712	0.7013	0.7508	0.7697
	GA	0.9397	0.9010	0.8436	0.8690	0.7957
	MI	0.9397	0.9010	0.8436	0.8690	0.8069
	SA	0.8975	0.9485	0.5278	0.5255	0.8175
	MRMR	0.9717	0.9718	0.9698	0.9708	0.8936
STAD	SPSA	0.9407	0.8996	0.9213	0.9099	0.8694
	RelChaNet	0.8880	0.8654	0.7639	0.8000	0.7695
	ReliefF	0.9037	0.8815	0.8009	0.8326	0.7946
	GA	0.9407	0.8996	0.9213	0.9099	0.8248
	MI	0.8880	0.8654	0.7639	0.8000	0.7610
	SA	0.8975	0.9485	0.5278	0.5255	0.7395
	MRMR	0.9588	0.9792	0.6111	0.6712	0.8691
THCA	SPSA	0.9941	0.9911	0.9957	0.9933	0.9017
	RelChaNet	0.9647	0.9692	0.9502	0.9589	0.8470
	ReliefF	0.9941	0.9911	0.9957	0.9933	0.8756
	GA	0.9882	0.9915	0.9818	0.9864	0.8678
	MI	0.6148	0.5537	0.5424	0.5382	0.4975
	SA	0.9588	0.9651	0.9411	0.9518	0.6257
	MRMR	0.9470	0.4735	0.5000	0.4864	0.8905
UCEC	SPSA	0.9096	0.9333	0.8904	0.9027	0.8560
	RelChaNet	0.7514	0.8514	0.6986	0.6970	0.6396
	ReliefF	0.8079	0.8768	0.7671	0.7780	0.6849
	GA	0.8022	0.8741	0.7603	0.7703	0.6978
	MI	0.7175	0.8377	0.6575	0.6427	0.5949
	SA	0.9745	0.4873	0.5000	0.4935	0.8492
	MRMR	0.8503	0.9181	0.6833	0.7237	0.7493

Table A6. K-Nearest Neighbors with 15% feature selection.

Dataset	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy
COAD	SPSA	0.9936	0.9968	0.8750	0.9269	0.8974
	RelChaNet	0.9936	0.9968	0.8750	0.9269	0.8974
	ReliefF	0.9936	0.9968	0.8750	0.9269	0.8753
	GA	0.9936	0.9968	0.8750	0.9269	0.8652
	MI	0.9936	0.9968	0.8750	0.9269	0.8935
	SA	0.7142	0.6875	0.6667	0.6725	0.5945
	MRMR	0.9764	0.9693	0.9779	0.9734	0.8417
HNSC	SPSA	0.9823	0.9909	0.8333	0.8954	0.8567
	RelChaNet	0.9705	0.9849	0.7222	0.8000	0.8325
	ReliefF	0.9764	0.9879	0.7778	0.8510	0.8436
	GA	0.9764	0.9879	0.7778	0.8510	0.8236
	MI	0.9823	0.9909	0.8333	0.8954	0.8167
	SA	0.9209	0.9278	0.9102	0.9169	0.8536
	MRMR	0.8915	0.4458	0.5000	0.4713	0.7804
KICH	SPSA	0.8928	0.9286	0.8500	0.8733	0.7549
	RelChaNet	0.7857	0.7667	0.7667	0.7667	0.6395
	ReliefF	0.8571	0.8444	0.8444	0.8444	0.7368
	GA	0.8928	0.8918	0.8722	0.8805	0.7368
	MI	0.7857	0.7667	0.7667	0.7667	0.6634
	SA	0.7216	0.3646	0.4930	0.4192	0.6295
	MRMR	0.7637	0.3819	0.5000	0.4330	0.6349
KIRP	SPSA	0.9175	0.9033	0.8827	0.8922	0.7950
	RelChaNet	0.8144	0.8025	0.6904	0.7166	0.6754
	ReliefF	0.8659	0.8570	0.7866	0.8119	0.7387
	GA	0.7731	0.7095	0.6988	0.7036	0.6495
	MI	0.8762	0.9008	0.7814	0.8177	0.7523
	SA	0.8975	0.9465	0.6458	0.6976	0.7854
	MRMR	0.9764	0.9693	0.9779	0.9734	0.8417
LIHC	SPSA	0.9212	0.9344	0.8448	0.8791	0.8265
	RelChaNet	0.8897	0.9126	0.7782	0.8202	0.7694
	ReliefF	0.8976	0.9181	0.7948	0.8356	0.7694
	GA	0.9212	0.9195	0.8564	0.8824	0.8059
	MI	0.9055	0.9236	0.8115	0.8506	0.7567
	SA	0.8975	0.9485	0.5278	0.5255	0.8275
	MRMR	0.9745	0.4873	0.5000	0.4935	0.8596
LUSC	SPSA	0.9216	0.8908	0.6633	0.7188	0.8192
	RelChaNet	0.8975	0.7534	0.5766	0.6029	0.7943
	ReliefF	0.9096	0.8565	0.6077	0.6496	0.7983
	GA	0.9156	0.8370	0.6599	0.7079	0.7893
	MI	0.9216	0.8908	0.6633	0.7188	0.8084
	SA	0.9527	0.9709	0.9000	0.9294	0.8295
	MRMR	0.7716	0.8849	0.5167	0.4672	0.6493
PRAD	SPSA	0.9397	0.9174	0.8263	0.8637	0.8315
	RelChaNet	0.9216	0.9581	0.7292	0.7924	0.8186
	ReliefF	0.8915	0.8487	0.6596	0.7051	0.7694
	GA	0.9036	0.8712	0.7013	0.7508	0.7594
	MI	0.9156	0.8891	0.7430	0.7920	0.7939
	SA	0.9647	0.9820	0.6667	0.7409	0.8594
	MRMR	0.9529	0.9609	0.9320	0.9446	0.8327

Table A6. *Cont.*

Dataset	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy
STAD	SPSA	0.9259	0.9173	0.8426	0.8733	0.8075
	RelChaNet	0.8880	0.8654	0.7639	0.8000	0.7695
	ReliefF	0.8962	0.8511	0.8102	0.8282	0.7843
	GA	0.9037	0.8815	0.8009	0.8326	0.7850
	MI	0.8962	0.8511	0.8102	0.8282	0.7740
	SA	0.7500	0.8600	0.6500	0.6494	0.6987
	MRMR	0.9291	0.9575	0.8500	0.8896	0.8528
THCA	SPSA	0.9941	0.9911	0.9957	0.9933	0.8794
	RelChaNet	0.9882	0.9866	0.9866	0.9866	0.8653
	ReliefF	0.9823	0.9821	0.9775	0.9797	0.8657
	GA	0.9882	0.9866	0.9866	0.9866	0.8594
	MI	0.9882	0.9866	0.9866	0.9866	0.8693
	SA	0.9527	0.9709	0.9000	0.9294	0.8639
	MRMR	0.8975	0.9465	0.6458	0.6976	0.7854
UCEC	SPSA	0.8757	0.9127	0.8493	0.8635	0.7957
	RelChaNet	0.8022	0.8741	0.7603	0.7703	0.7594
	ReliefF	0.7909	0.8688	0.7466	0.7548	0.6594
	GA	0.8757	0.9127	0.8493	0.8635	0.7496
	MI	0.7909	0.8688	0.7466	0.7548	0.6597
	SA	0.8554	0.4277	0.5000	0.4610	0.7295
	MRMR	0.8614	0.9303	0.5208	0.5025	0.7964

Table A7. LightGBM with 5% feature selection.

Dataset	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy
COAD	SPSA	1.0000	1.0000	1.0000	1.0000	0.9265
	RelChaNet	1.0000	1.0000	1.0000	1.0000	0.8769
	ReliefF	1.0000	1.0000	1.0000	1.0000	0.8976
	GA	1.0000	1.0000	1.0000	1.0000	0.8597
	MI	1.0000	1.0000	1.0000	1.0000	0.8695
	SA	0.7228	0.6406	0.8446	0.6275	0.6289
	MRMR	0.8453	0.8171	0.8944	0.8290	0.7257
HNSC	SPSA	0.9941	0.9969	0.9444	0.9690	0.9056
	RelChaNet	0.9823	0.8969	0.9382	0.9164	0.8658
	ReliefF	0.9882	0.9413	0.9413	0.9413	0.8368
	GA	0.9823	0.8969	0.9382	0.9164	0.8459
	MI	0.9882	0.9413	0.9413	0.9413	0.8495
	SA	0.8433	0.7493	0.9164	0.7739	0.7296
	MRMR	0.8373	0.7450	0.9037	0.7729	0.7504
KICH	SPSA	0.9642	0.9545	0.9722	0.9619	0.9670
	RelChaNet	0.9285	0.9500	0.9000	0.9181	0.8047
	ReliefF	0.9642	0.9545	0.9722	0.9619	0.8486
	GA	0.9285	0.9500	0.9000	0.9181	0.8076
	MI	0.9642	0.9545	0.9722	0.9619	0.8175
	SA	1.0000	1.0000	1.0000	1.0000	0.8691
	MRMR	0.8373	0.7450	0.9037	0.7729	0.7504

Table A7. *Cont.*

Dataset	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy
KIRP	SPSA	0.9690	0.9661	0.9545	0.9601	0.8898
	RelChaNet	0.9381	0.9309	0.9090	0.9192	0.8296
	ReliefF	0.9381	0.9212	0.9212	0.9212	0.8285
	GA	0.9175	0.9033	0.8827	0.8922	0.7986
	MI	0.9793	0.9737	0.9737	0.9737	0.8538
	SA	0.8433	0.7493	0.9164	0.7739	0.7296
	MRMR	0.9717	0.9700	0.9719	0.9709	0.8318
LIHC	SPSA	0.9606	0.9412	0.9512	0.9461	0.9284
	RelChaNet	0.9527	0.9441	0.9230	0.9330	0.8429
	ReliefF	0.9291	0.8931	0.9191	0.9050	0.8158
	GA	0.9685	0.9670	0.9448	0.9553	0.8487
	MI	0.9606	0.9412	0.9512	0.9461	0.8478
	SA	0.7228	0.6406	0.8446	0.6275	0.6289
	MRMR	0.9642	0.9545	0.9722	0.9619	0.8219
LUSC	SPSA	0.9578	0.9774	0.8056	0.8678	0.8743
	RelChaNet	0.9397	0.9212	0.7466	0.8050	0.8157
	ReliefF	0.9397	0.9212	0.7466	0.8050	0.8168
	GA	0.9518	0.9356	0.8022	0.8534	0.8185
	MI	0.9518	0.9744	0.7778	0.8440	0.8276
	SA	0.8453	0.8171	0.8944	0.8290	0.7285
	MRMR	0.9278	0.8953	0.9572	0.9176	0.7968
PRAD	SPSA	0.9819	0.9897	0.9375	0.9614	0.9076
	RelChaNet	0.9759	0.9669	0.9340	0.9495	0.8645
	ReliefF	0.9698	0.9624	0.9131	0.9357	0.8196
	GA	0.9698	0.9460	0.9305	0.9380	0.8594
	MI	0.9759	0.9669	0.9340	0.9495	0.8438
	SA	0.9259	0.8649	0.9537	0.8976	0.8196
	MRMR	0.8373	0.7450	0.9037	0.7729	0.7504
STAD	SPSA	0.9777	0.9716	0.9583	0.9648	0.8896
	RelChaNet	0.9777	0.9716	0.9583	0.9648	0.8494
	ReliefF	0.9777	0.9716	0.9583	0.9648	0.8494
	GA	0.9777	0.9716	0.9583	0.9648	0.8396
	MI	0.9851	0.9769	0.9769	0.9769	0.8754
	SA	0.8433	0.7493	0.9164	0.7739	0.7296
	MRMR	0.9212	0.8750	0.9485	0.9014	0.8659
THCA	SPSA	0.9941	0.9957	0.9909	0.9932	0.9048
	RelChaNet	0.6518	0.6002	0.5618	0.5509	0.5296
	ReliefF	0.9941	0.9957	0.9909	0.9932	0.8694
	GA	0.9941	0.9957	0.9909	0.9932	0.8589
	MI	0.9764	0.9731	0.9731	0.9731	0.8965
	SA	0.8433	0.7493	0.9164	0.7739	0.7296
	MRMR	1.0000	1.0000	1.0000	1.0000	0.8695
UCEC	SPSA	0.9661	0.9650	0.9650	0.9650	0.8959
	RelChaNet	0.9491	0.9505	0.9445	0.9472	0.8195
	ReliefF	0.9491	0.9532	0.9424	0.9470	0.8195
	GA	0.9661	0.9650	0.9650	0.9650	0.8789
	MI	0.9491	0.9532	0.9424	0.9470	0.8478
	SA	0.9887	0.9952	0.9959	0.9963	0.8591
	MRMR	0.8352	0.6256	0.9147	0.6593	0.7089

Table A8. LightGBM with 10% feature selection.

Dataset	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy
COAD	SPSA	1.0000	1.0000	1.0000	1.0000	0.9527
	RelChaNet	1.0000	1.0000	1.0000	1.0000	0.8850
	ReliefF	1.0000	1.0000	1.0000	1.0000	0.8858
	GA	1.0000	1.0000	1.0000	1.0000	0.8965
	MI	1.0000	1.0000	1.0000	1.0000	0.8958
	SA	0.8373	0.7450	0.9037	0.7729	0.7504
	MRMR	0.9259	0.8649	0.9537	0.8976	0.7846
HNSC	SPSA	0.9941	0.9969	0.9444	0.9690	0.9168
	RelChaNet	0.9823	0.8969	0.9382	0.9164	0.8749
	ReliefF	0.9882	0.9413	0.9413	0.9413	0.8068
	GA	0.9882	0.9413	0.9413	0.9413	0.8345
	MI	0.9882	0.9413	0.9413	0.9413	0.8697
	SA	0.8373	0.7450	0.9037	0.7729	0.7504
	MRMR	0.9823	0.9741	0.9870	0.9801	0.8569
KICH	SPSA	0.9642	0.9545	0.9722	0.9619	0.8948
	RelChaNet	0.9285	0.9500	0.9000	0.9181	0.7968
	ReliefF	0.9642	0.9545	0.9722	0.9619	0.8368
	GA	0.9642	0.9545	0.9722	0.9619	0.8486
	MI	0.9285	0.9500	0.9000	0.9181	0.7924
	SA	0.9823	0.9741	0.9870	0.9801	0.8694
	MRMR	1.0000	1.0000	1.0000	1.0000	1.0000
KIRP	SPSA	0.9793	0.9737	0.9737	0.9737	0.8947
	RelChaNet	0.9587	0.9475	0.9475	0.9475	0.8267
	ReliefF	0.9278	0.9122	0.9020	0.9069	0.8176
	GA	0.9690	0.9661	0.9545	0.9601	0.8496
	MI	0.9381	0.9212	0.9212	0.9212	0.8196
	SA	0.9212	0.8750	0.9485	0.9014	0.8659
	MRMR	0.9887	0.9952	0.9959	0.9963	0.8689
LIHC	SPSA	0.9685	0.9670	0.9448	0.9553	0.9059
	RelChaNet	0.9527	0.9270	0.9460	0.9360	0.8385
	ReliefF	0.9606	0.9412	0.9512	0.9461	0.8489
	GA	0.9606	0.9412	0.9512	0.9461	0.8657
	MI	0.9606	0.9502	0.9397	0.9448	0.8594
	SA	0.8373	0.7450	0.9037	0.7729	0.7504
	MRMR	0.8192	0.6875	0.8986	0.7163	0.7286
LUSC	SPSA	0.9518	0.9356	0.8022	0.8534	0.8497
	RelChaNet	0.9337	0.9126	0.7188	0.7783	0.8106
	ReliefF	0.9518	0.9744	0.7778	0.8440	0.8286
	GA	0.9578	0.9774	0.8056	0.8678	0.7945
	MI	0.9397	0.9212	0.7466	0.8050	0.8109
	SA	0.9212	0.8854	0.9525	0.9036	0.8091
	MRMR	0.9212	0.8750	0.9485	0.9014	0.7056
PRAD	SPSA	0.9819	0.9897	0.9375	0.9614	0.9184
	RelChaNet	0.9698	0.9460	0.9305	0.9380	0.8496
	ReliefF	0.9337	0.8703	0.8574	0.8637	0.8286
	GA	0.9337	0.8703	0.8574	0.8637	0.8265
	MI	0.9698	0.9624	0.9131	0.9357	0.8296
	SA	0.9681	0.7234	0.9854	0.8095	0.8494
	MRMR	0.7710	0.6607	0.8716	0.6696	0.7056

Table A8. *Cont.*

Dataset	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy
STAD	SPSA	0.9851	0.9769	0.9769	0.9769	0.8946
	RelChaNet	0.9777	0.9716	0.9583	0.9648	0.8494
	ReliefF	0.9777	0.9716	0.9583	0.9648	0.7956
	GA	0.9703	0.9664	0.9398	0.9524	0.8596
	MI	0.9777	0.9596	0.9722	0.9658	0.8496
	SA	0.8373	0.7450	0.9037	0.7729	0.7504
	MRMR	0.8556	0.8250	0.9014	0.8393	0.7195
THCA	SPSA	0.9941	0.9957	0.9909	0.9932	0.8960
	RelChaNet	0.9764	0.9731	0.9731	0.9731	0.8659
	ReliefF	0.9764	0.9731	0.9731	0.9731	0.8567
	GA	0.9823	0.9821	0.9775	0.9797	0.8675
	MI	0.9941	0.9957	0.9909	0.9932	0.8392
	SA	1.0000	1.0000	1.0000	1.0000	0.9067
	MRMR	0.9642	0.9545	0.9722	0.9619	0.8219
UCEC	SPSA	0.9661	0.9670	0.9630	0.9649	0.8469
	RelChaNet	0.9604	0.9601	0.9582	0.9591	0.8496
	ReliefF	0.9604	0.9601	0.9582	0.9591	0.8754
	GA	0.9604	0.9623	0.9561	0.9589	0.8493
	MI	0.9661	0.9650	0.9650	0.9650	0.8156
	SA	0.7647	0.5983	0.8824	0.5835	0.6496
	MRMR	0.9823	0.9741	0.9870	0.9801	0.8536

Table A9. LightGBM with 15% feature selection.

Dataset	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy
COAD	SPSA	1.0000	1.0000	1.0000	1.0000	0.8796
	RelChaNet	1.0000	1.0000	1.0000	1.0000	0.8694
	ReliefF	1.0000	1.0000	1.0000	1.0000	0.8793
	GA	0.9823	0.8969	0.9382	0.9164	0.8549
	MI	0.9882	0.9413	0.9413	0.9413	0.8437
	SA	0.7228	0.6406	0.8446	0.6275	0.6289
	MRMR	0.9212	0.8854	0.9525	0.9036	0.8091
HNSC	SPSA	0.9882	0.9413	0.9413	0.9413	0.8978
	RelChaNet	0.9823	0.9909	0.8333	0.8954	0.8569
	ReliefF	0.9882	0.9413	0.9413	0.9413	0.8386
	GA	0.9823	0.8969	0.9382	0.9164	0.8695
	MI	0.9882	0.9413	0.9413	0.9413	0.8694
	SA	1.0000	1.0000	1.0000	1.0000	0.8974
	MRMR	0.8373	0.7450	0.9037	0.7729	0.7504
KICH	SPSA	0.9642	0.9545	0.9722	0.9619	0.8948
	RelChaNet	0.9285	0.9500	0.9000	0.9181	0.8285
	ReliefF	0.9642	0.9545	0.9722	0.9619	0.8564
	GA	0.9642	0.9545	0.9722	0.9619	0.8275
	MI	0.9285	0.9500	0.9000	0.9181	0.7850
	SA	0.9604	0.9565	0.9643	0.9596	0.8402
	MRMR	0.7228	0.6406	0.8446	0.6275	0.6289

<div align="center">

Table A9. *Cont.*

</div>

Dataset	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy
KIRP	SPSA	0.9690	0.9661	0.9545	0.9601	0.8865
	RelChaNet	0.9175	0.9033	0.8827	0.8922	0.7952
	ReliefF	0.9587	0.9475	0.9475	0.9475	0.8406
	GA	0.9381	0.9212	0.9212	0.9212	0.8186
	MI	0.9175	0.9033	0.8827	0.8922	0.8058
	SA	0.9887	0.9952	0.9959	0.9963	0.8506
	MRMR	0.8726	0.5833	0.9346	0.6079	0.7495
LIHC	SPSA	0.9685	0.9670	0.9448	0.9553	0.8697
	RelChaNet	0.9606	0.9412	0.9512	0.9461	0.8549
	ReliefF	0.9606	0.9502	0.9397	0.9448	0.8539
	GA	0.9291	0.8931	0.9191	0.9050	0.8076
	MI	0.9606	0.9412	0.9512	0.9461	0.8386
	SA	0.9661	0.9625	0.9691	0.9653	0.8405
	MRMR	0.9681	0.7234	0.9854	0.8095	0.8494
LUSC	SPSA	0.9518	0.9356	0.8022	0.8534	0.8478
	RelChaNet	0.9518	0.9744	0.7778	0.8440	0.8368
	ReliefF	0.9518	0.9744	0.7778	0.8440	0.8205
	GA	0.9337	0.9126	0.7188	0.7783	0.8429
	MI	0.9578	0.9774	0.8056	0.8678	0.8438
	SA	0.7000	0.5750	0.8416	0.5363	0.5876
	MRMR	0.5421	0.5957	0.7432	0.4880	0.4109
PRAD	SPSA	0.9759	0.9669	0.9340	0.9495	0.8869
	RelChaNet	0.9759	0.9669	0.9340	0.9495	0.8329
	ReliefF	0.9698	0.9624	0.9131	0.9357	0.8798
	GA	0.9819	0.9897	0.9375	0.9614	0.8569
	MI	0.9698	0.9624	0.9131	0.9357	0.8769
	SA	0.9212	0.8750	0.9485	0.9014	0.8659
	MRMR	0.5421	0.5957	0.7432	0.4880	0.4076
STAD	SPSA	0.9851	0.9769	0.9769	0.9769	0.8759
	RelChaNet	0.9777	0.9596	0.9722	0.9658	0.8295
	ReliefF	0.9703	0.9664	0.9398	0.9524	0.8459
	GA	0.9703	0.9664	0.9398	0.9524	0.8503
	MI	0.9777	0.9596	0.9722	0.9658	0.8396
	SA	0.5602	0.5989	0.7534	0.5015	0.4892
	MRMR	0.8253	0.7264	0.8979	0.7548	0.7195
THCA	SPSA	0.9941	0.9957	0.9909	0.9932	0.8869
	RelChaNet	0.9823	0.9821	0.9775	0.9797	0.8295
	ReliefF	0.9882	0.9866	0.9866	0.9866	0.8439
	GA	0.9882	0.9866	0.9866	0.9866	0.8439
	MI	0.9882	0.9866	0.9866	0.9866	0.8495
	SA	0.8352	0.6256	0.9147	0.6593	0.7295
	MRMR	0.8882	0.6607	0.9410	0.7119	0.8056
UCEC	SPSA	0.9661	0.9650	0.9650	0.9650	0.8374
	RelChaNet	0.9604	0.9623	0.9561	0.9589	0.8358
	ReliefF	0.9491	0.9468	0.9486	0.9476	0.8295
	GA	0.9491	0.9505	0.9445	0.9472	0.8292
	MI	0.9491	0.9505	0.9445	0.9472	0.8295
	SA	0.9823	0.9741	0.9870	0.9801	0.8749
	MRMR	0.9882	0.9847	0.9928	0.9925	0.8967

Table A10. Logistic Regression with 5% feature selection.

Dataset	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy
COAD	SPSA	0.9808	0.7857	0.9902	0.8587	0.8940
	RelChaNet	0.9681	0.7234	0.9854	0.8095	0.8494
	ReliefF	0.8726	0.5833	0.9346	0.6079	0.7495
	GA	0.9745	0.7564	0.9976	0.8323	0.8539
	MI	0.9808	0.7857	0.9902	0.8587	0.8945
	SA	0.9823	0.9821	0.9775	0.9797	0.8295
	MRMR	0.9578	0.9774	0.8056	0.8678	0.7945
HNSC	SPSA	0.8882	0.6607	0.9410	0.7119	0.8056
	RelChaNet	0.8352	0.6256	0.9147	0.6593	0.7089
	ReliefF	0.7000	0.5750	0.8416	0.5363	0.5849
	GA	0.8352	0.6256	0.9147	0.6593	0.7057
	MI	0.7647	0.5983	0.8824	0.5835	0.6329
	SA	0.9578	0.9774	0.8056	0.8678	0.7945
	MRMR	0.9175	0.9033	0.8827	0.8922	0.8058
KICH	SPSA	1.0000	1.0000	1.0000	1.0000	0.9058
	RelChaNet	0.9642	0.9545	0.9722	0.9619	0.8592
	ReliefF	0.9642	0.9545	0.9722	0.9619	0.8305
	GA	1.0000	1.0000	1.0000	1.0000	0.8596
	MI	1.0000	1.0000	1.0000	1.0000	0.8974
	SA	0.9397	0.9212	0.7466	0.8050	0.8168
	MRMR	0.9606	0.9412	0.9512	0.9461	0.8657
KIRP	SPSA	0.9175	0.8854	0.9487	0.9032	0.8495
	RelChaNet	0.8453	0.8171	0.8944	0.8290	0.7285
	ReliefF	0.9072	0.8714	0.9366	0.8924	0.8694
	GA	0.8556	0.8250	0.9014	0.8393	0.7195
	MI	0.8453	0.8171	0.8944	0.8290	0.7098
	SA	0.9491	0.9505	0.9445	0.9472	0.8195
	MRMR	0.9759	0.9669	0.9340	0.9495	0.8438
LIHC	SPSA	0.9212	0.8854	0.9525	0.9036	0.8596
	RelChaNet	0.8346	0.7941	0.8918	0.8097	0.7290
	ReliefF	0.9212	0.8750	0.9485	0.9014	0.7056
	GA	0.8346	0.7941	0.8918	0.8097	0.7098
	MI	0.8346	0.7941	0.8918	0.8097	0.7109
	SA	0.9642	0.9545	0.9722	0.9619	0.8175
	MRMR	0.9882	0.9413	0.9413	0.9413	0.8068
LUSC	SPSA	0.8192	0.6875	0.8986	0.7163	0.7598
	RelChaNet	0.7710	0.6654	0.8732	0.6720	0.6459
	ReliefF	0.6144	0.5912	0.7350	0.5311	0.4986
	GA	0.7228	0.6406	0.8446	0.6275	0.6289
	MI	0.8192	0.6875	0.8986	0.7163	0.7286
	SA	0.9882	0.9413	0.9413	0.9413	0.8495
	MRMR	0.9642	0.9545	0.9722	0.9619	0.9670
PRAD	SPSA	0.8554	0.7500	0.9155	0.7872	0.7698
	RelChaNet	0.8373	0.7450	0.9037	0.7729	0.7209
	ReliefF	0.7951	0.7069	0.8803	0.7247	0.6792
	GA	0.8433	0.7493	0.9164	0.7739	0.7296
	MI	0.7951	0.7069	0.8803	0.7247	0.6589
	SA	0.9698	0.9624	0.9131	0.9357	0.8296
	MRMR	0.9518	0.9356	0.8022	0.8534	0.8478

Table A10. *Cont.*

Dataset	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy
STAD	SPSA	0.9481	0.8971	0.9676	0.9259	0.8578
	RelChaNet	0.9407	0.8952	0.9693	0.9284	0.8395
	ReliefF	0.9259	0.8649	0.9537	0.8976	0.8196
	GA	0.9259	0.8649	0.9537	0.8976	0.8056
	MI	0.9481	0.9053	0.9739	0.9395	0.8194
	SA	0.9606	0.9502	0.9397	0.9448	0.8594
	MRMR	0.9823	0.9821	0.9775	0.9797	0.8675
THCA	SPSA	0.9882	0.9825	0.9913	0.9867	0.8958
	RelChaNet	0.5777	0.5968	0.6049	0.5727	0.4896
	ReliefF	0.9823	0.9741	0.9870	0.9801	0.8749
	GA	0.9882	0.9847	0.9928	0.9925	0.8695
	MI	0.9882	0.9847	0.9928	0.9925	0.8967
	SA	0.9777	0.9716	0.9583	0.9648	0.8494
	MRMR	0.9777	0.9716	0.9583	0.9648	0.8896
UCEC	SPSA	0.9887	0.9952	0.9959	0.9963	0.8768
	RelChaNet	0.9661	0.9620	0.9712	0.9654	0.8589
	ReliefF	0.9717	0.9700	0.9719	0.9709	0.8318
	GA	0.9661	0.9620	0.9712	0.9654	0.8478
	MI	0.9661	0.9620	0.9712	0.9654	0.8496
	SA	0.9175	0.9033	0.8827	0.8922	0.8058
	MRMR	0.9291	0.8931	0.9191	0.9050	0.8158

Table A11. Logistic Regression with 10% feature selection.

Dataset	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy
COAD	SPSA	0.9808	0.7857	0.9902	0.8587	0.9285
	RelChaNet	0.9681	0.7234	0.9854	0.8095	0.8265
	ReliefF	0.8726	0.5833	0.9346	0.6079	0.6589
	GA	0.9745	0.7564	0.9976	0.8323	0.8156
	MI	0.9745	0.7564	0.9976	0.8323	0.8356
	SA	0.9527	0.9441	0.9230	0.9330	0.8429
	MRMR	0.9606	0.9412	0.9512	0.9461	0.8549
HNSC	SPSA	0.8882	0.6607	0.9410	0.7119	0.7596
	RelChaNet	0.8352	0.6256	0.9147	0.6593	0.6854
	ReliefF	0.7000	0.5750	0.8416	0.5363	0.4596
	GA	0.7000	0.5750	0.8416	0.5363	0.5876
	MI	0.7941	0.6034	0.8923	0.6175	0.6840
	SA	0.9703	0.9664	0.9398	0.9524	0.8503
	MRMR	1.0000	1.0000	1.0000	1.0000	0.8976
KICH	SPSA	1.0000	1.0000	1.0000	1.0000	0.8691
	RelChaNet	0.9642	0.9545	0.9722	0.9619	0.8596
	ReliefF	0.9642	0.9545	0.9722	0.9619	0.8195
	GA	1.0000	1.0000	1.0000	1.0000	0.8695
	MI	1.0000	1.0000	1.0000	1.0000	0.9067
	SA	0.9685	0.9670	0.9448	0.9553	0.8487
	MRMR	0.9397	0.9212	0.7466	0.8050	0.8109

Table A11. *Cont.*

Dataset	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy
KIRP	SPSA	0.9175	0.8854	0.9487	0.9032	0.8267
	RelChaNet	0.8453	0.8171	0.8944	0.8290	0.7285
	ReliefF	0.9072	0.8714	0.9366	0.8924	0.8075
	GA	0.9072	0.8714	0.9366	0.8924	0.7567
	MI	0.9175	0.8854	0.9487	0.9032	0.8056
	SA	0.9764	0.9731	0.9731	0.9731	0.8965
	MRMR	0.9685	0.9670	0.9448	0.9553	0.9059
LIHC	SPSA	0.9212	0.8854	0.9525	0.9036	0.8749
	RelChaNet	0.8346	0.7941	0.8918	0.8097	0.7594
	ReliefF	0.9212	0.8750	0.9485	0.9014	0.7748
	GA	0.9212	0.8750	0.9485	0.9014	0.8295
	MI	0.9212	0.8750	0.9485	0.9014	0.8659
	SA	0.9882	0.9413	0.9413	0.9413	0.8697
	MRMR	1.0000	1.0000	1.0000	1.0000	0.8850
LUSC	SPSA	0.8192	0.6875	0.8986	0.7163	0.7056
	RelChaNet	0.7710	0.6654	0.8732	0.6720	0.5987
	ReliefF	0.6144	0.5912	0.7350	0.5311	0.4892
	GA	0.6144	0.5912	0.7350	0.5311	0.5186
	MI	0.7048	0.6393	0.8394	0.6193	0.5984
	SA	0.9941	0.9969	0.9444	0.9690	0.9056
	MRMR	0.9606	0.9412	0.9512	0.9461	0.8549
PRAD	SPSA	0.8554	0.7500	0.9155	0.7872	0.7504
	RelChaNet	0.8373	0.7450	0.9037	0.7729	0.6594
	ReliefF	0.7951	0.7069	0.8803	0.7247	0.7297
	GA	0.7289	0.6739	0.8415	0.6639	0.6087
	MI	0.8433	0.7493	0.9164	0.7739	0.7285
	SA	0.9642	0.9545	0.9722	0.9619	0.8275
	MRMR	0.9518	0.9356	0.8022	0.8534	0.8478
STAD	SPSA	0.9481	0.8971	0.9676	0.9259	0.8239
	RelChaNet	0.9407	0.8952	0.9693	0.9284	0.8168
	ReliefF	0.9259	0.8649	0.9537	0.8976	0.8372
	GA	0.9481	0.9053	0.9739	0.9395	0.8195
	MI	0.9259	0.8649	0.9537	0.8976	0.7846
	SA	0.9381	0.9212	0.9212	0.9212	0.8196
	MRMR	0.9759	0.9669	0.9340	0.9495	0.8438
THCA	SPSA	0.9882	0.9825	0.9913	0.9867	0.9284
	RelChaNet	0.5777	0.5968	0.6049	0.5727	0.8567
	ReliefF	0.9823	0.9741	0.9870	0.9801	0.8694
	GA	0.5777	0.5968	0.6049	0.5727	0.4489
	MI	0.9823	0.9741	0.9870	0.9801	0.8569
	SA	0.9777	0.9716	0.9583	0.9648	0.8494
	MRMR	0.9397	0.9212	0.7466	0.8050	0.8109
UCEC	SPSA	0.9887	0.9952	0.9959	0.9963	0.8914
	RelChaNet	0.9661	0.9620	0.9712	0.9654	0.8402
	ReliefF	0.9717	0.9700	0.9719	0.9709	0.8405
	GA	0.9548	0.9506	0.9595	0.9539	0.8285
	MI	0.9717	0.9700	0.9719	0.9709	0.8568
	SA	0.9606	0.9412	0.9512	0.9461	0.8549
	MRMR	1.0000	1.0000	1.0000	1.0000	0.8597

Table A12. Logistic Regression with 15% feature selection.

Dataset	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy
COAD	SPSA	0.9745	0.7564	0.9976	0.8323	0.8867
	RelChaNet	0.7643	0.5488	0.8791	0.5201	0.6496
	ReliefF	0.9745	0.7500	0.9869	0.8267	0.8539
	GA	0.7643	0.5488	0.8791	0.5201	0.6696
	MI	0.9745	0.7564	0.9976	0.8323	0.8495
	SA	0.9661	0.9670	0.9630	0.9649	0.8469
	MRMR	0.9397	0.9212	0.7466	0.8050	0.8157
HNSC	SPSA	0.7705	0.5938	0.8789	0.5890	0.6984
	RelChaNet	0.7647	0.5983	0.8824	0.5835	0.6496
	ReliefF	0.5470	0.5523	0.7609	0.4376	0.4285
	GA	0.8352	0.6256	0.9147	0.6593	0.7295
	MI	0.5882	0.5570	0.7826	0.4634	0.4395
	SA	0.9337	0.9126	0.7188	0.7783	0.8106
	MRMR	0.9606	0.9412	0.9512	0.9461	0.8386
KICH	SPSA	0.9642	0.9545	0.9722	0.9619	0.8596
	RelChaNet	0.9642	0.9545	0.9722	0.9619	0.8493
	ReliefF	0.9642	0.9545	0.9722	0.9619	0.8957
	GA	0.9642	0.9545	0.9722	0.9619	0.8219
	MI	0.9642	0.9545	0.9722	0.9619	0.8489
	SA	0.9285	0.9500	0.9000	0.9181	0.7968
	MRMR	0.9604	0.9623	0.9561	0.9589	0.8358
KIRP	SPSA	0.9175	0.8854	0.9487	0.9032	0.8478
	RelChaNet	0.8453	0.8171	0.8944	0.8290	0.7078
	ReliefF	0.9175	0.8824	0.9437	0.9035	0.7984
	GA	0.8453	0.8171	0.8944	0.8290	0.7257
	MI	0.9278	0.8953	0.9572	0.9176	0.7968
	SA	0.9518	0.9744	0.7778	0.8440	0.8276
	MRMR	0.9941	0.9957	0.9909	0.9932	0.8960
LIHC	SPSA	0.9212	0.8854	0.9525	0.9036	0.8296
	RelChaNet	0.8425	0.8000	0.8969	0.8175	0.7694
	ReliefF	0.8976	0.8488	0.9330	0.8750	0.7639
	GA	0.9212	0.8854	0.9525	0.9036	0.8091
	MI	0.8346	0.7941	0.8918	0.8097	0.7257
	SA	0.9759	0.9669	0.9340	0.9495	0.8438
	MRMR	0.9661	0.9650	0.9650	0.9650	0.8789
LUSC	SPSA	0.7228	0.6406	0.8446	0.6275	0.6289
	RelChaNet	0.7048	0.6393	0.8394	0.6193	0.6176
	ReliefF	0.5421	0.5957	0.7432	0.4880	0.4076
	GA	0.7228	0.6406	0.8446	0.6275	0.6096
	MI	0.5421	0.5957	0.7432	0.4880	0.4109
	SA	0.9777	0.9716	0.9583	0.9648	0.8896
	MRMR	0.9882	0.9413	0.9413	0.9413	0.8694
PRAD	SPSA	0.8433	0.7493	0.9164	0.7739	0.7296
	RelChaNet	0.7289	0.6739	0.8415	0.6639	0.6086
	ReliefF	0.8253	0.7264	0.8979	0.7548	0.7195
	GA	0.8554	0.7500	0.9155	0.7872	0.7349
	MI	0.8554	0.7500	0.9155	0.7872	0.7195
	SA	0.9703	0.9664	0.9398	0.9524	0.8596
	MRMR	0.9698	0.9624	0.9131	0.9357	0.8798

Table A12. *Cont.*

Dataset	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy
STAD	SPSA	0.9481	0.9053	0.9739	0.9395	0.8239
	RelChaNet	0.8666	0.8000	0.9167	0.8295	0.7395
	ReliefF	0.9259	0.8649	0.9537	0.8976	0.7946
	GA	0.8666	0.8000	0.9167	0.8295	0.7439
	MI	0.9481	0.9053	0.9739	0.9395	0.8594
	SA	0.9642	0.9545	0.9722	0.9619	0.8948
	MRMR	0.9175	0.9033	0.8827	0.8922	0.7952
THCA	SPSA	0.9882	0.9847	0.9928	0.9925	0.8597
	RelChaNet	0.9705	0.9583	0.9783	0.9671	0.8540
	ReliefF	0.9823	0.9741	0.9870	0.9801	0.8375
	GA	0.9823	0.9741	0.9870	0.9801	0.8536
	MI	0.9823	0.9741	0.9870	0.9801	0.8569
	SA	0.9661	0.9650	0.9650	0.9650	0.8789
	MRMR	1.0000	1.0000	1.0000	1.0000	0.8796
UCEC	SPSA	0.9887	0.9952	0.9959	0.9963	0.8506
	RelChaNet	0.9548	0.9506	0.9595	0.9539	0.8346
	ReliefF	0.9717	0.9687	0.9739	0.9710	0.8589
	GA	0.9887	0.9952	0.9959	0.9963	0.8689
	MI	0.9887	0.9952	0.9959	0.9963	0.8591
	SA	1.0000	1.0000	1.0000	1.0000	0.8796
	MRMR	0.9661	0.9670	0.9630	0.9649	0.8469

Table A13. SVM with 5% feature selection.

Dataset	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy
COAD	SPSA	1.0000	1.0000	1.0000	1.0000	0.9298
	RelChaNet	0.9745	0.4873	0.5000	0.4935	0.8958
	ReliefF	0.9745	0.4873	0.5000	0.4935	0.8386
	GA	0.9808	0.9904	0.6250	0.6951	0.8756
	MI	0.9745	0.4873	0.5000	0.4935	0.8396
	SA	0.8453	0.8541	0.7359	0.7681	0.7267
	MRMR	0.9096	0.9333	0.8904	0.9027	0.8560
HNSC	SPSA	0.9647	0.9820	0.6667	0.7409	0.8594
	RelChaNet	0.9647	0.9820	0.6667	0.7409	0.8417
	ReliefF	0.9470	0.4735	0.5000	0.4864	0.8398
	GA	0.9588	0.9792	0.6111	0.6712	0.8682
	MI	0.9588	0.9792	0.6111	0.6712	0.8692
	SA	0.8022	0.8741	0.7603	0.7703	0.6978
	MRMR	0.9882	0.9866	0.9866	0.9866	0.8594
KICH	SPSA	0.9642	0.9737	0.9500	0.9602	0.8745
	RelChaNet	0.7142	0.6875	0.6667	0.6725	0.6095
	ReliefF	0.7500	0.8600	0.6500	0.6494	0.6285
	GA	0.8928	0.9286	0.8500	0.8733	0.7598
	MI	0.7500	0.8600	0.6500	0.6494	0.6894
	SA	0.9175	0.9033	0.8827	0.8922	0.7950
	MRMR	0.9936	0.9968	0.8750	0.9269	0.8974

<div align="center">Table A13. <i>Cont.</i></div>

Dataset	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy
KIRP	SPSA	0.7422	0.6585	0.5680	0.5642	0.6849
	RelChaNet	0.7216	0.3646	0.4930	0.4192	0.6295
	ReliefF	0.7319	0.5116	0.5305	0.4917	0.6257
	GA	0.7422	0.8698	0.5192	0.4622	0.6295
	MI	0.7422	0.6585	0.5680	0.5642	0.6197
	SA	0.9156	0.8891	0.7430	0.7920	0.7956
	MRMR	0.8762	0.9008	0.7814	0.8177	0.7329
LIHC	SPSA	0.9527	0.9709	0.9000	0.9294	0.8639
	RelChaNet	0.8582	0.6764	0.7000	0.6812	0.7349
	ReliefF	0.7637	0.3819	0.5000	0.4330	0.6349
	GA	0.9527	0.9709	0.9000	0.9294	0.8193
	MI	0.9527	0.9709	0.9000	0.9294	0.8295
	SA	0.7731	0.7095	0.6988	0.7036	0.6495
	MRMR	0.9407	0.8996	0.9213	0.9099	0.8248
LUSC	SPSA	0.8975	0.9485	0.5278	0.5255	0.8275
	RelChaNet	0.8975	0.9485	0.5278	0.5255	0.7594
	ReliefF	0.8915	0.4458	0.5000	0.4713	0.7697
	GA	0.8975	0.9485	0.5278	0.5255	0.7395
	MI	0.8915	0.7057	0.5976	0.6244	0.7941
	SA	0.9036	0.8712	0.7013	0.7508	0.7594
	MRMR	0.9212	0.9344	0.8448	0.8791	0.8265
PRAD	SPSA	0.8915	0.9437	0.6250	0.6702	0.8467
	RelChaNet	0.8674	0.7708	0.5763	0.5969	0.7548
	ReliefF	0.8554	0.4277	0.5000	0.4610	0.7386
	GA	0.8975	0.9465	0.6458	0.6976	0.7632
	MI	0.8795	0.9383	0.5833	0.6100	0.7943
	SA	0.9481	0.9358	0.8981	0.9154	0.8257
	MRMR	0.9291	0.9398	0.8615	0.8927	0.8056
STAD	SPSA	0.9555	0.9306	0.9306	0.9306	0.8943
	RelChaNet	0.9037	0.9463	0.7593	0.8131	0.7689
	ReliefF	0.7481	0.5789	0.5648	0.5693	0.6278
	GA	0.9037	0.9463	0.7593	0.8131	0.7496
	MI	0.7481	0.5789	0.5648	0.5693	0.6295
	SA	0.9936	0.9968	0.8750	0.9269	0.8652
	MRMR	0.9882	0.9939	0.8889	0.9344	0.8256
THCA	SPSA	0.9352	0.9167	0.9522	0.9294	0.8593
	RelChaNet	0.6444	0.3222	0.5000	0.3919	0.5096
	ReliefF	0.9294	0.9449	0.8957	0.9150	0.8469
	GA	0.6444	0.3222	0.5000	0.3919	0.5837
	MI	0.9294	0.9449	0.8957	0.9150	0.7975
	SA	0.8571	0.8444	0.8444	0.8444	0.7246
	MRMR	0.8928	0.9286	0.8500	0.8733	0.7549
UCEC	SPSA	0.9661	0.9650	0.9650	0.9650	0.8823
	RelChaNet	0.9209	0.9278	0.9102	0.9169	0.8536
	ReliefF	0.8474	0.8477	0.8355	0.8401	0.7195
	GA	0.9548	0.9577	0.9493	0.9530	0.8727
	MI	0.9717	0.9718	0.9698	0.9708	0.8273
	SA	0.9175	0.9033	0.8827	0.8922	0.7950
	MRMR	0.9216	0.8908	0.6633	0.7188	0.8084

Table A14. SVM with 10% feature selection.

Dataset	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy
COAD	SPSA	0.9872	0.9935	0.7500	0.8301	0.8750
	RelChaNet	0.9808	0.9904	0.6250	0.6951	0.8697
	ReliefF	0.9745	0.4873	0.5000	0.4935	0.8296
	GA	0.9745	0.4873	0.5000	0.4935	0.8492
	MI	0.9745	0.4873	0.5000	0.4935	0.8290
	SA	0.9055	0.9236	0.8115	0.8506	0.7560
	MRMR	0.8962	0.8511	0.8102	0.8282	0.7740
HNSC	SPSA	0.9647	0.9820	0.6667	0.7409	0.8736
	RelChaNet	0.9588	0.9792	0.6111	0.6712	0.8358
	ReliefF	0.9470	0.4735	0.5000	0.4864	0.8295
	GA	0.9647	0.9820	0.6667	0.7409	0.8385
	MI	0.9470	0.4735	0.5000	0.4864	0.7958
	SA	0.7857	0.7667	0.7667	0.7667	0.6587
	MRMR	0.9882	0.9866	0.9866	0.9866	0.8594
KICH	SPSA	0.9285	0.9500	0.9000	0.9181	0.8594
	RelChaNet	0.7142	0.6875	0.6667	0.6725	0.5945
	ReliefF	0.7500	0.8600	0.6500	0.6494	0.6987
	GA	0.7500	0.8600	0.6500	0.6494	0.6295
	MI	0.8928	0.9286	0.8500	0.8733	0.7256
	SA	0.9156	0.8370	0.6599	0.7079	0.7893
	MRMR	0.9096	0.9333	0.8904	0.9027	0.8560
KIRP	SPSA	0.8556	0.8351	0.7795	0.8005	0.7847
	RelChaNet	0.7422	0.8698	0.5192	0.4622	0.6395
	ReliefF	0.7989	0.8525	0.6494	0.6314	0.6845
	GA	0.8556	0.8351	0.7795	0.8005	0.7395
	MI	0.9042	0.9030	0.8398	0.8650	0.7689
	SA	0.9941	0.9911	0.9957	0.9933	0.8794
	MRMR	0.6148	0.5537	0.5424	0.5382	0.4975
LIHC	SPSA	0.9291	0.9575	0.8500	0.8896	0.8496
	RelChaNet	0.8504	0.9212	0.6834	0.6784	0.7694
	ReliefF	0.7716	0.8849	0.5167	0.4672	0.6493
	GA	0.8504	0.9212	0.6834	0.6784	0.7268
	MI	0.9291	0.9575	0.8500	0.8896	0.8195
	SA	0.8144	0.7788	0.7148	0.7355	0.6595
	MRMR	0.9175	0.9033	0.8827	0.8922	0.8023
LUSC	SPSA	0.8975	0.9485	0.5278	0.5255	0.8843
	RelChaNet	0.8915	0.4458	0.5000	0.4713	0.7804
	ReliefF	0.8915	0.4458	0.5000	0.4713	0.7956
	GA	0.8975	0.9485	0.5278	0.5255	0.8384
	MI	0.8915	0.7057	0.5976	0.6244	0.7304
	SA	0.8928	0.8918	0.8722	0.8805	0.7368
	MRMR	0.9397	0.9010	0.8436	0.8690	0.8069
PRAD	SPSA	0.8975	0.9465	0.6458	0.6976	0.8495
	RelChaNet	0.8975	0.9465	0.6458	0.6976	0.7794
	ReliefF	0.8614	0.9303	0.5208	0.5025	0.7496
	GA	0.8554	0.4277	0.5000	0.4610	0.7295
	MI	0.8614	0.9303	0.5208	0.5025	0.7857
	SA	0.9156	0.8891	0.7430	0.7920	0.7677
	MRMR	0.9156	0.8891	0.7430	0.7920	0.7677

Table A14. *Cont.*

Dataset	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy
STAD	SPSA	0.9407	0.9455	0.8657	0.8986	0.8495
	RelChaNet	0.8814	0.9001	0.7176	0.7652	0.7945
	ReliefF	0.7851	0.6521	0.6296	0.6385	0.6495
	GA	0.7481	0.5789	0.5648	0.5693	0.6829
	MI	0.8370	0.7511	0.6898	0.7118	0.7937
	SA	0.9156	0.8370	0.6599	0.7920	0.7677
	MRMR	0.9156	0.8891	0.7430	0.7920	0.7677
THCA	SPSA	0.9647	0.9692	0.9502	0.9589	0.8821
	RelChaNet	0.9294	0.9528	0.8909	0.9140	0.8547
	ReliefF	0.9529	0.9365	0.9652	0.9481	0.8367
	GA	0.9529	0.9609	0.9320	0.9446	0.8735
	MI	0.9647	0.9692	0.9502	0.9589	0.8628
	SA	0.9096	0.8565	0.6077	0.6496	0.7983
	MRMR	0.8975	0.8236	0.7151	0.7530	0.7649
UCEC	SPSA	0.9717	0.9718	0.9698	0.9708	0.8936
	RelChaNet	0.9322	0.9400	0.9219	0.9287	0.8195
	ReliefF	0.9039	0.9149	0.8897	0.8982	0.8287
	GA	0.9717	0.9718	0.9698	0.9708	0.8472
	MI	0.9661	0.9650	0.9650	0.9650	0.8295
	SA	0.8144	0.7788	0.7148	0.7355	0.6595
	MRMR	0.8880	0.8654	0.7639	0.8000	0.7695

Table A15. SVM with 15% feature selection.

Dataset	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy
COAD	SPSA	0.9745	0.4873	0.5000	0.4935	0.9295
	RelChaNet	0.9745	0.4873	0.5000	0.4935	0.8596
	ReliefF	0.9745	0.4873	0.5000	0.4935	0.8976
	GA	0.9745	0.4873	0.5000	0.4935	0.8597
	MI	0.9808	0.9904	0.6250	0.6951	0.9056
	SA	0.6148	0.5537	0.5424	0.5382	0.4975
	MRMR	0.9212	0.9344	0.8448	0.8791	0.8265
HNSC	SPSA	0.9588	0.9792	0.6111	0.6712	0.8612
	RelChaNet	0.9588	0.9792	0.6111	0.6712	0.8296
	ReliefF	0.9470	0.4735	0.5000	0.4864	0.8905
	GA	0.9470	0.4735	0.5000	0.4864	0.8295
	MI	0.9588	0.9792	0.6111	0.6712	0.8691
	SA	0.9941	0.9911	0.9957	0.9933	0.8794
	MRMR	0.7731	0.7095	0.6988	0.7036	0.6106
KICH	SPSA	0.8928	0.9286	0.8500	0.8733	0.8154
	RelChaNet	0.7142	0.6875	0.6667	0.6725	0.5896
	ReliefF	0.7500	0.8600	0.6500	0.6494	0.6289
	GA	0.7142	0.6875	0.6667	0.6725	0.5986
	MI	0.7500	0.8600	0.6500	0.6494	0.6384
	SA	0.8144	0.7788	0.7148	0.7355	0.6595
	MRMR	0.8975	0.8236	0.7151	0.7530	0.7649

Table A15. *Cont.*

Dataset	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy
KIRP	SPSA	0.7422	0.8698	0.5192	0.4622	0.6891
	RelChaNet	0.7371	0.6179	0.5096	0.4424	0.6086
	ReliefF	0.7319	0.3660	0.5000	0.4226	0.6057
	GA	0.7371	0.6179	0.5096	0.4424	0.6296
	MI	0.7422	0.8698	0.5192	0.4622	0.6285
	SA	0.9941	0.9911	0.9957	0.9933	0.8756
	MRMR	0.9216	0.9581	0.7292	0.7924	0.8186
LIHC	SPSA	0.9291	0.9575	0.8500	0.8896	0.8528
	RelChaNet	0.8897	0.9378	0.7667	0.8067	0.7934
	ReliefF	0.8503	0.9181	0.6833	0.7237	0.7493
	GA	0.8110	0.9015	0.6000	0.5955	0.6967
	MI	0.7716	0.8849	0.5167	0.4672	0.6239
	SA	0.9036	0.8712	0.7013	0.7508	0.7697
	MRMR	0.8915	0.8487	0.6596	0.7051	0.7694
LUSC	SPSA	0.8975	0.9485	0.5278	0.5255	0.8175
	RelChaNet	0.8915	0.7057	0.5976	0.6244	0.8640
	ReliefF	0.8915	0.4458	0.5000	0.4713	0.7834
	GA	0.8915	0.7057	0.5976	0.6244	0.7594
	MI	0.8975	0.9485	0.5278	0.5255	0.7495
	SA	0.8757	0.9127	0.8493	0.8635	0.7596
	MRMR	0.9764	0.9879	0.7778	0.8510	0.8436
PRAD	SPSA	0.8975	0.9465	0.6458	0.6976	0.7854
	RelChaNet	0.8795	0.9383	0.5833	0.6100	0.7478
	ReliefF	0.8614	0.9303	0.5208	0.5025	0.7964
	GA	0.8975	0.9465	0.6458	0.6976	0.7854
	MI	0.8674	0.7708	0.5763	0.5969	0.7395
	SA	0.9941	0.9911	0.9957	0.9933	0.8756
	MRMR	0.7514	0.8514	0.6986	0.6970	0.6396
STAD	SPSA	0.8880	0.9390	0.7222	0.7752	0.7593
	RelChaNet	0.8074	0.9030	0.5185	0.4820	0.6743
	ReliefF	0.8370	0.7511	0.6898	0.7118	0.7937
	GA	0.9555	0.9306	0.9306	0.9306	0.8789
	MI	0.8814	0.9001	0.7176	0.7652	0.7397
	SA	0.9882	0.9866	0.9866	0.9866	0.8594
	MRMR	0.8144	0.7788	0.7148	0.7355	0.6926
THCA	SPSA	0.9764	0.9693	0.9779	0.9734	0.8689
	RelChaNet	0.9529	0.9609	0.9320	0.9446	0.8327
	ReliefF	0.9705	0.9583	0.9783	0.9671	0.8526
	GA	0.9764	0.9693	0.9779	0.9734	0.8417
	MI	0.9294	0.9449	0.8957	0.9150	0.8075
	SA	0.9216	0.9581	0.7292	0.7924	0.8064
	MRMR	0.9055	0.9236	0.8115	0.8506	0.7567
UCEC	SPSA	0.9717	0.9718	0.9698	0.9708	0.8893
	RelChaNet	0.9548	0.9577	0.9493	0.9530	0.8388
	ReliefF	0.9265	0.9322	0.9171	0.9230	0.8495
	GA	0.9548	0.9577	0.9493	0.9530	0.8563
	MI	0.9717	0.9718	0.9698	0.9708	0.8354
	SA	0.9941	0.9911	0.9957	0.9933	0.8794
	MRMR	0.9823	0.9909	0.8333	0.8954	0.8567

Table A16. XGBoost with 5% feature selection.

Dataset	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy
COAD	SPSA	1.0000	1.0000	1.0000	1.0000	1.0000
	RelChaNet	0.9936	0.9968	0.8750	0.9269	0.9256
	ReliefF	1.0000	1.0000	1.0000	1.0000	0.9862
	GA	1.0000	1.0000	1.0000	1.0000	1.0000
	MI	1.0000	1.0000	1.0000	1.0000	0.9589
	SA	0.9705	0.8302	0.9320	0.8731	0.8678
	MRMR	0.9555	0.9216	0.9444	0.9324	0.9166
HNSC	SPSA	0.9941	0.9969	0.9444	0.9690	0.9586
	RelChaNet	0.9823	0.9313	0.8858	0.9071	0.8932
	ReliefF	0.9882	0.9939	0.8889	0.9344	0.9273
	GA	0.9823	0.9313	0.8858	0.9071	0.8596
	MI	0.9882	0.9939	0.8889	0.9344	0.9285
	SA	0.9555	0.9216	0.9444	0.9324	0.8923
	MRMR	0.9291	0.8985	0.9076	0.9029	0.8450
KICH	SPSA	1.0000	1.0000	1.0000	1.0000	1.0000
	RelChaNet	0.9285	0.9500	0.9000	0.9181	0.8495
	ReliefF	1.0000	1.0000	1.0000	1.0000	1.0000
	GA	0.9285	0.9500	0.9000	0.9181	0.8175
	MI	1.0000	1.0000	1.0000	1.0000	0.9486
	SA	0.9481	0.9240	0.9120	0.9178	0.8843
	MRMR	0.9152	0.9120	0.9136	0.9127	0.8462
KIRP	SPSA	0.9690	0.9661	0.9545	0.9601	0.8365
	RelChaNet	0.9278	0.9122	0.9020	0.9069	0.7947
	ReliefF	0.9484	0.9392	0.9282	0.9335	0.8672
	GA	0.9278	0.9122	0.9020	0.9069	0.7648
	MI	0.9587	0.9586	0.9353	0.9461	0.8725
	SA	0.9337	0.8703	0.8574	0.8637	0.8782
	MRMR	0.8865	0.8499	0.8738	0.8604	0.8256
LIHC	SPSA	0.9527	0.9441	0.9230	0.9330	0.8496
	RelChaNet	0.9448	0.9279	0.9179	0.9227	0.8185
	ReliefF	0.9133	0.8727	0.8973	0.8839	0.8564
	GA	0.9448	0.9199	0.9294	0.9245	0.8395
	MI	0.9133	0.8727	0.8973	0.8839	0.7963
	SA	0.9152	0.9120	0.9136	0.9127	0.8086
	MRMR	0.8976	0.8522	0.8754	0.8628	0.8349
LUSC	SPSA	0.9397	0.9212	0.7466	0.8050	0.8946
	RelChaNet	0.9397	0.9212	0.7466	0.8050	0.8175
	ReliefF	0.9397	0.9212	0.7466	0.8050	0.8305
	GA	0.9397	0.9212	0.7466	0.8050	0.8175
	MI	0.9578	0.9774	0.8056	0.8678	0.8283
	SA	0.9647	0.8045	0.9289	0.8542	0.9063
	MRMR	0.9647	0.9752	0.9455	0.9584	0.9600
PRAD	SPSA	0.9759	0.9669	0.9340	0.9495	0.8826
	RelChaNet	0.9518	0.9295	0.8680	0.8952	0.8284
	ReliefF	0.9698	0.9624	0.9131	0.9357	0.8594
	GA	0.9759	0.9863	0.9167	0.9476	0.8485
	MI	0.9759	0.9669	0.9340	0.9495	0.8593
	SA	0.8813	0.8815	0.8725	0.8762	0.7943
	MRMR	0.9555	0.9216	0.9444	0.9324	0.9166

Table A16. *Cont.*

Dataset	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy
STAD	SPSA	0.9851	0.9769	0.9769	0.9769	0.9286
	RelChaNet	0.9555	0.9418	0.9167	0.9285	0.8391
	ReliefF	0.9629	0.9611	0.9213	0.9396	0.8475
	GA	0.9555	0.9306	0.9306	0.9306	0.8238
	MI	0.9481	0.9240	0.9120	0.9178	0.8645
	SA	0.8926	0.8887	0.8903	0.8895	0.8519
	MRMR	0.9156	0.7819	0.7819	0.7819	0.8280
THCA	SPSA	1.0000	1.0000	1.0000	1.0000	0.9893
	RelChaNet	0.6592	0.6143	0.5862	0.5853	0.5973
	ReliefF	0.9823	0.9821	0.9775	0.9797	0.8647
	GA	0.9823	0.9821	0.9775	0.9797	0.8465
	MI	0.9823	0.9778	0.9822	0.9799	0.8392
	SA	0.9277	0.8308	0.7643	0.7925	0.7829
	MRMR	0.9647	0.8119	0.8765	0.8406	0.8130
UCEC	SPSA	0.9717	0.9742	0.9678	0.9707	0.8891
	RelChaNet	0.9435	0.9487	0.9356	0.9409	0.8592
	ReliefF	0.9322	0.9340	0.9260	0.9294	0.8197
	GA	0.9435	0.9487	0.9356	0.9409	0.8625
	MI	0.9717	0.9742	0.9678	0.9707	0.8939
	SA	0.9481	0.9240	0.9120	0.9178	0.8823
	MRMR	0.9529	0.9503	0.9415	0.9457	0.9411

Table A17. XGBoost with 10% feature selection.

Dataset	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy
COAD	SPSA	1.0000	1.0000	1.0000	1.0000	0.9735
	RelChaNet	0.9936	0.9968	0.8750	0.9269	0.9624
	ReliefF	1.0000	1.0000	1.0000	1.0000	0.9625
	GA	1.0000	1.0000	1.0000	1.0000	0.9382
	MI	1.0000	1.0000	1.0000	1.0000	0.9483
	SA	0.9397	0.8782	0.8782	0.8782	0.8538
	MRMR	0.9647	0.8045	0.9289	0.8542	0.8826
HNSC	SPSA	0.9882	0.9939	0.8889	0.9344	0.9629
	RelChaNet	0.9823	0.9313	0.8858	0.9071	0.9074
	ReliefF	0.9882	0.9939	0.8889	0.9344	0.9627
	GA	0.9823	0.9313	0.8858	0.9071	0.8825
	MI	0.9823	0.9313	0.8858	0.9071	0.8836
	SA	0.8976	0.8522	0.8754	0.8628	0.8349
	MRMR	0.8813	0.8815	0.8725	0.8762	0.7943
KICH	SPSA	1.0000	1.0000	1.0000	1.0000	1.0000
	RelChaNet	0.9285	0.9500	0.9000	0.9181	0.8829
	ReliefF	1.0000	1.0000	1.0000	1.0000	0.9572
	GA	1.0000	1.0000	1.0000	1.0000	1.0000
	MI	1.0000	1.0000	1.0000	1.0000	1.0000
	SA	0.9277	0.8308	0.7643	0.7925	0.8459
	MRMR	0.9337	0.8435	0.7920	0.8149	0.6261

Table A17. *Cont.*

Dataset	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy
KIRP	SPSA	0.9484	0.9392	0.9282	0.9335	0.8296
	RelChaNet	0.9381	0.9309	0.9090	0.9192	0.8173
	ReliefF	0.9381	0.9309	0.9090	0.9192	0.8243
	GA	0.9484	0.9237	0.9526	0.9366	0.8395
	MI	0.9690	0.9661	0.9545	0.9601	0.8293
	SA	0.9285	0.9167	0.9444	0.9251	0.9444
	MRMR	0.9647	0.9597	0.9597	0.9597	0.8746
LIHC	SPSA	0.9606	0.9502	0.9397	0.9448	0.8394
	RelChaNet	0.9527	0.9345	0.9345	0.9345	0.8439
	ReliefF	0.9606	0.9502	0.9397	0.9448	0.8475
	GA	0.9606	0.9502	0.9397	0.9448	0.8158
	MI	0.9448	0.9199	0.9294	0.9245	0.8385
	SA	0.9555	0.9216	0.9444	0.9324	0.9166
	MRMR	0.9481	0.9240	0.9120	0.9178	0.8823
LUSC	SPSA	0.9578	0.9774	0.8056	0.8678	0.8720
	RelChaNet	0.9457	0.9002	0.7988	0.8399	0.8285
	ReliefF	0.9337	0.8801	0.7432	0.7922	0.8147
	GA	0.9457	0.9287	0.7744	0.8300	0.8294
	MI	0.9457	0.9002	0.7988	0.8399	0.8167
	SA	0.8976	0.8522	0.8754	0.8628	0.8175
	MRMR	0.9216	0.8387	0.8504	0.8444	0.8711
PRAD	SPSA	0.9819	0.9897	0.9375	0.9614	0.8794
	RelChaNet	0.9698	0.9830	0.8958	0.9332	0.8294
	ReliefF	0.9759	0.9863	0.9167	0.9476	0.8837
	GA	0.9698	0.9624	0.9131	0.9357	0.8836
	MI	0.9698	0.9624	0.9131	0.9357	0.8276
	SA	0.9481	0.9240	0.9120	0.9178	0.8843
	MRMR	1.0000	1.0000	1.0000	1.0000	1.0000
STAD	SPSA	0.9777	0.9716	0.9583	0.9648	0.8936
	RelChaNet	0.9703	0.9664	0.9398	0.9524	0.8594
	ReliefF	0.9481	0.9240	0.9120	0.9178	0.8278
	GA	0.9555	0.9418	0.9167	0.9285	0.8187
	MI	0.9703	0.9664	0.9398	0.9524	0.8468
	SA	0.9337	0.8435	0.7920	0.8149	0.6261
	MRMR	1.0000	1.0000	1.0000	1.0000	1.0000
THCA	SPSA	0.9823	0.9778	0.9822	0.9799	0.8946
	RelChaNet	0.9823	0.9821	0.9775	0.9797	0.8629
	ReliefF	0.9764	0.9777	0.9684	0.9729	0.8593
	GA	0.9823	0.9821	0.9775	0.9797	0.8583
	MI	0.9823	0.9821	0.9775	0.9797	0.8654
	SA	0.8865	0.8499	0.8738	0.8604	0.8256
	MRMR	0.9457	0.9437	0.8298	0.8748	0.8849
UCEC	SPSA	0.9548	0.9577	0.9493	0.9530	0.9381
	RelChaNet	0.9322	0.9318	0.9280	0.9298	0.8754
	ReliefF	0.9548	0.9553	0.9513	0.9532	0.8574
	GA	0.9548	0.9577	0.9493	0.9530	0.8729
	MI	0.9378	0.9387	0.9328	0.9355	0.8692
	SA	0.8865	0.8499	0.8738	0.8604	0.8256
	MRMR	0.9555	0.9216	0.9444	0.9324	0.9166

Table A18. XGBoost with 15% feature selection.

Dataset	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy
COAD	SPSA	1.0000	1.0000	1.0000	1.0000	1.0000
	RelChaNet	0.9936	0.9968	0.8750	0.9269	0.8745
	ReliefF	1.0000	1.0000	1.0000	1.0000	1.0000
	GA	1.0000	1.0000	1.0000	1.0000	0.9018
	MI	1.0000	1.0000	1.0000	1.0000	0.9721
	SA	0.9216	0.8387	0.8504	0.8444	0.8711
	MRMR	0.9481	0.9240	0.9120	0.9178	0.8720
HNSC	SPSA	0.9882	0.9939	0.8889	0.9344	0.9385
	RelChaNet	0.9823	0.9313	0.8858	0.9071	0.8195
	ReliefF	0.9882	0.9939	0.8889	0.9344	0.9656
	GA	0.9941	0.9969	0.9444	0.9690	0.9057
	MI	0.9882	0.9939	0.8889	0.9344	0.8593
	SA	0.9647	0.8045	0.9289	0.8542	0.8826
	MRMR	0.9337	0.8703	0.8574	0.8637	0.8782
KICH	SPSA	1.0000	1.0000	1.0000	1.0000	0.9591
	RelChaNet	0.9285	0.9500	0.9000	0.9181	0.8926
	ReliefF	1.0000	1.0000	1.0000	1.0000	0.9364
	GA	0.9285	0.9500	0.9000	0.9181	0.9053
	MI	1.0000	1.0000	1.0000	1.0000	0.9582
	SA	0.9936	0.9968	0.8750	0.9269	0.9500
	MRMR	0.9555	0.9216	0.9444	0.9324	0.9351
KIRP	SPSA	0.9690	0.9661	0.9545	0.9601	0.8794
	RelChaNet	0.9484	0.9237	0.9526	0.9366	0.8793
	ReliefF	0.9587	0.9586	0.9353	0.9461	0.8729
	GA	0.9484	0.9392	0.9282	0.9335	0.8462
	MI	0.9690	0.9661	0.9545	0.9601	0.8521
	SA	0.9647	0.8045	0.9289	0.8542	0.8742
	MRMR	0.8870	0.8821	0.8875	0.8843	0.7560
LIHC	SPSA	0.9606	0.9502	0.9397	0.9448	0.8295
	RelChaNet	0.9448	0.9199	0.9294	0.9245	0.8517
	ReliefF	0.9527	0.9441	0.9230	0.9330	0.8294
	GA	0.9527	0.9441	0.9230	0.9330	0.8275
	MI	0.9606	0.9502	0.9397	0.9448	0.8395
	SA	0.8870	0.8821	0.8875	0.8843	0.7540
	MRMR	0.9647	0.8045	0.9289	0.8542	0.9257
LUSC	SPSA	0.9638	0.9805	0.8333	0.8901	0.8493
	RelChaNet	0.9457	0.9287	0.7744	0.8300	0.8592
	ReliefF	0.9457	0.9713	0.7500	0.8186	0.8692
	GA	0.9337	0.8801	0.7432	0.7922	0.7849
	MI	0.9337	0.8801	0.7432	0.7922	0.8692
	SA	1.0000	1.0000	1.0000	1.0000	1.0000
	MRMR	0.9764	0.9879	0.7778	0.8510	0.9289
PRAD	SPSA	0.9759	0.9863	0.9167	0.9476	0.8639
	RelChaNet	0.9698	0.9830	0.8958	0.9332	0.8196
	ReliefF	0.9698	0.9624	0.9131	0.9357	0.8526
	GA	0.9518	0.9295	0.8680	0.8952	0.8683
	MI	0.9698	0.9624	0.9131	0.9357	0.8481
	SA	0.9096	0.7651	0.8273	0.7913	0.8374
	MRMR	0.9152	0.9120	0.9136	0.9127	0.8462

Table A18. *Cont.*

Dataset	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Balanced Accuracy
STAD	SPSA	0.9703	0.9664	0.9398	0.9524	0.8605
	RelChaNet	0.9555	0.9306	0.9306	0.9306	0.8364
	ReliefF	0.9703	0.9664	0.9398	0.9524	0.8493
	GA	0.9851	0.9769	0.9769	0.9769	0.8296
	MI	0.9703	0.9664	0.9398	0.9524	0.8468
	SA	0.9216	0.8387	0.8504	0.8444	0.8711
	MRMR	0.9397	0.8782	0.8782	0.8782	0.8644
THCA	SPSA	0.9823	0.9821	0.9775	0.9797	0.9528
	RelChaNet	0.9705	0.9734	0.9593	0.9659	0.8504
	ReliefF	0.9823	0.9778	0.9822	0.9799	0.8749
	GA	0.6592	0.6143	0.5862	0.5853	0.5295
	MI	0.9705	0.9734	0.9593	0.9659	0.8939
	SA	0.8247	0.7803	0.7584	0.7679	0.7920
	MRMR	0.9152	0.9120	0.9136	0.9127	0.8462
UCEC	SPSA	0.9491	0.9505	0.9445	0.9472	0.9017
	RelChaNet	0.9378	0.9387	0.9328	0.9355	0.8429
	ReliefF	0.9491	0.9484	0.9465	0.9474	0.8826
	GA	0.9322	0.9340	0.9260	0.9294	0.8597
	MI	0.9435	0.9487	0.9356	0.9409	0.8197
	SA	1.0000	1.0000	1.0000	1.0000	0.9385
	MRMR	0.9872	0.9935	0.7500	0.8301	0.8949

References

1. Elshawi, R.; Sakr, S.; Talia, D.; Trunfio, P. Big Data Systems Meet Machine Learning Challenges: Towards Big Data Science as a Service. *Big Data Res.* **2018**, *14*, 1–11. [CrossRef]

2. Rawat, D.B.; Doku, R.; Garuba, M. Cybersecurity in Big Data Era: From Securing Big Data to Data-Driven Security. *IEEE Trans. Serv. Comput.* **2021**, *14*, 2055–2072. [CrossRef]

3. Parsajoo, M.; Armaghani, D.J.; Asteris, P.G. A precise neuro-fuzzy model enhanced by artificial bee colony techniques for assessment of rock brittleness index. *Neural Comput. Appl.* **2021**, *34*, 3263–3281. [CrossRef]

4. Hu, H.; Feng, D.; Yang, F. A Promising Nonlinear Dimensionality Reduction Method: Kernel-Based Within Class Collaborative Preserving Discriminant Projection. *IEEE Signal Process. Lett.* **2020**, *27*, 2034–2038. [CrossRef]

5. Parsajoo, M.; Mohammed, A.; Yagiz, S.; Armaghani, D.; Khandelwal, M. An evolutionary adaptive neuro-fuzzy inference system for estimating field penetration index of tunnel boring machine in rock mass. *J. Rock Mech. Geotech. Eng.* **2021**, *13*, 1290–1299. [CrossRef]

6. Wang, J.; Wang, L.; Nie, F.; Li, X. Joint Feature Selection and Extraction With Sparse Unsupervised Projection. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, *34*, 3071–3081. [CrossRef]

7. Wang, S.; Gao, C.; Zhang, Q.; Dakulagi, V.; Zeng, H.; Zheng, G.; Bai, J.; Song, Y.; Cai, J.; Zong, B. Research and Experiment of Radar Signal Support Vector Clustering Sorting Based on Feature Extraction and Feature Selection. *IEEE Access* **2020**, *8*, 93322–93334. [CrossRef]

8. Li, J.; Chen, J.; Qi, F.; Dan, T.; Weng, W.; Zhang, B.; Yuan, H.; Cai, H.; Zhong, C. Two-Dimensional Unsupervised Feature Selection via Sparse Feature Filter. *IEEE Trans. Cybern.* **2023**, *53*, 5605–5617. [CrossRef]

9. Thejas, G.S.; Garg, R.; Iyengar, S.S.; Sunitha, N.R.; Badrinath, P.; Chennupati, S. Metric and Accuracy Ranked Feature Inclusion: Hybrids of Filter and Wrapper Feature Selection Approaches. *IEEE Access* **2021**, *9*, 128687–128701. [CrossRef]

10. Mandal, A.K.; Nadim, M.; Saha, H.; Sultana, T.; Hossain, M.D.; Huh, E.N. Feature Subset Selection for High-Dimensional, Low Sampling Size Data Classification Using Ensemble Feature Selection With a Wrapper-Based Search. *IEEE Access* **2024**, *12*, 62341–62357. [CrossRef]

11. Yoshino, E.; Juarto, B.; Kurniadi, F.I. Hybrid Machine Learning Model for Breast Cancer Classification with K-Means Clustering Feature Selection Techniques. In Proceedings of the 2023 International Seminar on Application for Technology of Information and Communication (iSemantic), Semarang, Indonesia, 16–17 September 2023; pp. 182–186.

12. Nethala, T.R.; Sahoo, B.K.; Srinivasulu, P. GECC-Net: Gene Expression-Based Cancer Classification using Hybrid Fuzzy Ranking Network with Multi-kernel SVM. In Proceedings of the 2022 International Conference on Industry 4.0 Technology (I4Tech), Pune, India, 23–24 September 2022; pp. 1–6.

13. Kohavi, R.; John, G.H. Wrappers for feature subset selection. *Artif. Intell.* **1997**, *97*, 273–324. [CrossRef]

14. Miller, A. *Subset Selection in Regression*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2002.

15. Reunanen, J. Overfitting in making comparisons between variable selection methods. *J. Mach. Learn. Res.* **2003**, *3*, 1371–1382.

16. Bergstra, J.; Bengio, Y. Random search for hyper-parameter optimization. *J. Mach. Learn. Res.* **2012**, *13*, 281–305.

17. Shahriari, B.; Swersky, K.; Wang, Z.; Adams, R.P.; de Freitas, N. Taking the Human Out of the Loop: A Review of Bayesian Optimization. *Proc. IEEE* **2016**, *104*, 148–175. [CrossRef]

18. Kursa, M.B.; Rudnicki, W.R. Feature Selection with the Boruta Package. *J. Stat. Softw.* **2010**, *36*, 1–13. [CrossRef]

19. Degenhardt, F.; Seifert, S.; Szymczak, S. Evaluation of variable selection methods for random forests and omics data sets. *Briefings Bioinform.* **2017**, *20*, 492–503. [CrossRef] [PubMed]

20. Speiser, J.L.; Miller, M.E.; Tooze, J.; Ip, E. A comparison of random forest variable selection methods for classification prediction modeling. *Expert Syst. Appl.* **2019**, *134*, 93–101. [CrossRef] [PubMed]

21. Nilsson, R.; Peña, J.M.; Björkegren, J.; Tegnér, J. Consistent Feature Selection for Pattern Recognition in Polynomial Time. *J. Mach. Learn. Res.* **2007**, *8*, 589–612.

22. Granitto, P.M.; Furlanello, C.; Biasioli, F.; Gasperi, F. Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products. *Chemom. Intell. Lab. Syst.* **2006**, *83*, 83–90. [CrossRef]

23. Meinshausen, N.; Bühlmann, P. Stability Selection. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2010**, *72*, 417–473. [CrossRef]

24. Austin, P.C.; Tu, J.V. Bootstrap Methods for Developing Predictive Models. *Am. Stat.* **2004**, *58*, 131–137. [CrossRef]

25. Kartini, D.; Mazdadi, M.I.; Budiman, I.; Indriani, F.; Hidayat, R. Binary PSO-GWO for Feature Selection in Binary Classification Using K-Nearest Neighbor. In Proceedings of the 2023 Eighth International Conference on Informatics and Computing (ICIC), Manado, Indonesia, 8–9 December 2023; pp. 1–6.

26. Ludwig, S.A. Guided Particle Swarm Optimization for Feature Selection: Application to Cancer Genome Data. *Algorithms* **2025**, *18*, 220. [CrossRef]

27. Singh, S.N.; Mishra, S.; Satapathy, S.K.; Cho, S.B.; Mallick, P.K. Efficient Feature Selection Techniques for Accurate Cancer Analysis Using Krill Heard Optimization. In Proceedings of the 2024 International Conference on Emerging Systems and Intelligent Computing (ESIC), Bandung, Indonesia, 17–18 September 2024; pp. 41–45.

28. Singh, U.K.; Rout, M. Genetic Algorithm based Feature Selection to Enhance Breast Cancer Classification. In Proceedings of the 2023 IEEE International Conference on Contemporary Computing and Communications (InC4), Bangalore, India, 21–22 April 2023; Volume 1, pp. 1–5.

29. Touchanti, K.; Ezzazi, I.; Bekkali, M.E.; Maser, S. A 2-stages feature selection framework for colon cancer classification using SVM. In Proceedings of the 2022 International Conference on Intelligent Systems and Computer Vision (ISCV), Fez, Morocco, 18–20 May 2022; pp. 1–5.

30. Sachdeva, R.K.; Bathla, P.; Rani, P.; Kukreja, V.; Ahuja, R. A Systematic Method for Breast Cancer Classification using RFE Feature Selection. In Proceedings of the 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 28–29 April 2022; pp. 1673–1676.

31. Mohiuddin, T.; Naznin, S.; Upama, P.B. Classification and Performance Analysis of Cancer Microarrays Using Relevant Genes. In Proceedings of the 2021 5th International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), Dhaka, Bangladesh, 18–20 November 2021; pp. 1–6.

32. Si, C.; Zhao, L.; Liu, J. Deep Feature Selection Algorithm for Classification of Gastric Cancer Subtypes. In Proceedings of the 2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Honolulu, HI, USA, 1–4 October 2023; pp. 698–705.

33. Spall, J. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Trans. Autom. Control* **1992**, *37*, 332–341. [CrossRef]

34. Spall, J.C.; Chin, D.C. Traffic-responsive signal timing for system-wide traffic control. *Transp. Res. Part C Emerg. Technol.* **1997**, *5*, 153–163. [CrossRef]

35. Maeda, Y. Real-time control and learning using neuro-controller via simultaneous perturbation for flexible arm system. In Proceedings of the 2002 American Control Conference (IEEE Cat. No.CH37301), Anchorage, AK, USA, 8–10 May 2002; Volume 4, pp. 2583–2588.

36. Johannsen, D.A.; Wegman, E.J.; Solka, J.L.; Priebe, C.E. Simultaneous Selection of Features and Metric for Optimal Nearest Neighbor Classification. *Commun. Stat.—Theory Methods* **2004**, *33*, 2137–2157. [CrossRef]

37. Aksakalli, V.; Malekipirbazari, M. Feature selection via binary simultaneous perturbation stochastic approximation. *Pattern Recognit. Lett.* **2016**, *75*, 41–47. [CrossRef]

38. Yenice, Z.; Adhikari, N.; Wong, Y.; Aksakalli, V.; Taskin, A.; Abbasi, B. SPSA-FSR: Simultaneous Perturbation Stochastic Approximation for Feature Selection and Ranking. *arXiv* **2018**, arXiv:1804.05589. [CrossRef]

39. Aksakalli, V.; Yenice, Z.D.; Malekipirbazari, M.; Kargar, K. Feature selection using stochastic approximation with Barzilai and Borwein non-monotone gains. *Comput. Oper. Res.* **2021**, *132*, 105334. [CrossRef]

40. Algin, R.; Alkaya, A.F.; Agaoglu, M. Performance of Simultaneous Perturbation Stochastic Approximation for Feature Selection. In Proceedings of the Intelligent and Fuzzy Systems, Bornova, Turkey, 19–21 July 2022; pp. 348–354.

41. Akman, D.V.; Malekipirbazari, M.; Yenice, Z.D.; Yeo, A.; Adhikari, N.; Wong, Y.K.; Abbasi, B.; Gumus, A.T. k-best feature selection and ranking via stochastic approximation. *Expert Syst. Appl.* **2023**, *213*, 118864. [CrossRef]

42. Spall, J.C. *Introduction to Stochastic Search and Optimization*, 1st ed.; Wiley-Interscience: New York, NY, USA, 2003.

43. Barzilai, J.; Borwein, J.M. Two-Point Step Size Gradient Methods. *IMA J. Numer. Anal.* **1988**, *8*, 141–148. [CrossRef]

44. Raydan, M. The Barzilai and Borwein Gradient Method for the Large Scale Unconstrained Minimization Problem. *SIAM J. Optim.* **1997**, *7*, 26–33. [CrossRef]

45. Molina, B.; Raydan, M. Preconditioned Barzilai-Borwein method for the numerical solution of partial differential equations. *Numer. Algorithms* **1996**, *13*, 45–60. [CrossRef]

46. Dai, Y.H.; Liao, L.Z. R-linear convergence of the Barzilai and Borwein gradient method. *IMA J. Numer. Anal.* **2002**, *22*, 1–10. [CrossRef]

47. Sietsma; Dow. Neural net pruning-why and how. In Proceedings of the IEEE 1988 International Conference on Neural Networks, San Diego, CA, USA, 24–27 July 1988; Volume 1, pp. 325–333. [CrossRef]

48. Zimmer, F. RelChaNet: Neural Network Feature Selection using Relative Change Scores. *arXiv* **2024**, arXiv:2410.02344.

49. Fan, H.; Xue, L.; Song, Y.; Li, M. A repetitive feature selection method based on improved ReliefF for missing data. *Appl. Intell.* **2022**, *52*, 16265–16280. [CrossRef]

50. Kira, K.; Rendell, L. The feature selection problem: Traditional methods and a new algorithm. In Proceedings of the AAAI-92: Proceedings of the 10th National Conference on Artifical Intelligence, San Jose, CA, USA, 12–16 July 1992; Swartout, W., Ed.; AAAI Press: Menlo Park, CA, USA; The MIT Press: Cambridge, MA, USA, August 1992; pp. 129–134.

51. Kononenko, I.; Šimec, E.; Robnik-Šikonja, M. Overcoming the Myopia of Inductive Learning Algorithms with RELIEFF. *Appl. Intell.* **1997**, *7*, 39–55. [CrossRef]

52. Halim, Z.; Yousaf, M.N.; Waqas, M.; Sulaiman, M.; Abbas, G.; Hussain, M.; Ahmad, I.; Hanif, M. An effective genetic algorithm-based feature selection method for intrusion detection systems. *Comput. Secur.* **2021**, *110*, 102448. [CrossRef]

53. Macedo, F.; Valadas, R.; Carrasquinha, E.; Oliveira, M.R.; Pacheco, A. Feature selection using Decomposed Mutual Information Maximization. *Neurocomputing* **2022**, *513*, 215–232. [CrossRef]

54. Sulaiman, M.A.; Labadin, J. Feature selection based on mutual information. In Proceedings of the 2015 9th International Conference on IT in Asia (CITA), Kuching, Malaysia, 4–5 August 2015; pp. 1–6.

55. Peng, H.; Long, F.; Ding, C. Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1226–1238. [CrossRef]

56. Syaiful, A.; Sartono, B.; Afendi, F.M.; Anisa, R.; Salim, A. Feature Selection using Simulated Annealing with Optimal Neighborhood Approach. *J. Phys. Conf. Ser.* **2021**, *1752*, 012030. [CrossRef]

57. Ding, C.; Peng, H. Minimum redundancy feature selection from microarray gene expression data. In Proceedings of the 2003 IEEE Bioinformatics Conference, Stanford, CA, USA, 11–14 August 2003; pp. 523–528. [CrossRef]

58. Seeja, G.; Doss, A.S.A.; Hency, V.B. A Novel Approach for Disaster Victim Detection Under Debris Environments Using Decision Tree Algorithms with Deep Learning Features. *IEEE Access* **2023**, *11*, 54760–54772. [CrossRef]

59. El Zein, Y.; Lemay, M.; Huguenin, K. PrivaTree: Collaborative Privacy-Preserving Training of Decision Trees on Biomedical Data. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2024**, *21*, 1–13. [CrossRef] [PubMed]

60. Fix, E.; Hodges, J.L., Jr. *Discriminatory Analysis—Nonparametric Discrimination: Consistency Properties*; Technical Report; California Univ Berkeley: Berkeley, CA, USA, 1951.

61. Cover, T.; Hart, P. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **1967**, *13*, 21–27. [CrossRef]

62. Xing, W.; Bei, Y. Medical Health Big Data Classification Based on KNN Classification Algorithm. *IEEE Access* **2020**, *8*, 28808–28819. [CrossRef]

63. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 13–17 August 2016; KDD '16; pp. 785–794.

64. Ashenden, S.K. (Ed.) *The Era of Artificial Intelligence, Machine Learning, and Data Science in the Pharmaceutical Industry*; Elsevier: Amsterdam, The Netherlands, 2021.

65. Li, T.H.S.; Chiu, H.J.; Kuo, P.H. Hepatitis C Virus Detection Model by Using Random Forest, Logistic-Regression and ABC Algorithm. *IEEE Access* **2022**, *10*, 91045–91058. [CrossRef]

66. Teddy, M.; Muhammad, A.; Media, A. Crime Index Based on Text Mining on Social Media using Multi Classifier Neural-Net Algorithm. *TELKOMNIKA Telecommun. Comput. Electron. Control* **2022**, *20*, 570–579.

67. García-Gonzalo, E.; Fernández-Muñiz, Z.; García Nieto, P.J.; Bernardo Sánchez, A.; Menéndez Fernández, M. Hard-Rock Stability Analysis for Span Design in Entry-Type Excavations with Learning Classifiers. *Materials* **2016**, *9*, 531. [CrossRef]

68. Asaly, S.; Gottlieb, L.-A.; Reuveni, Y. Using Support Vector Machine (SVM) and Ionospheric Total Electron Content (TEC) Data for Solar Flare Predictions. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 1469–1481. [CrossRef]

69. Shamreen Ahamed, B. LGBM Classifier based Technique for Predicting Type-2 Diabetes. *Eur. J. Mol. Clin. Med.* **2021**, *8*, 454–467.

70. Algarni, A.; Ahmad, Z.; Alaa Ala'Anzy, M. An Edge Computing-Based and Threat Behavior-Aware Smart Prioritization Framework for Cybersecurity Intrusion Detection and Prevention of IEDs in Smart Grids with Integration of Modified LGBM and One Class-SVM Models. *IEEE Access* **2024**, *12*, 104948–104963. [CrossRef]

71. Alzamzami, F.; Hoda, M.; Saddik, A.E. Light Gradient Boosting Machine for General Sentiment Classification on Short Texts: A Comparative Evaluation. *IEEE Access* **2020**, *8*, 101840–101858. [CrossRef]

72. Weinstein, J.; Collisson, E.; Mills, G.; Shaw, K.; Ozenberger, B.; Ellrott, K.; Shmulevich, I.; Sander, C.; Stuart, J. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **2013**, *45*, 113–1120. [CrossRef]

73. Yin, H. Enhancing Ionospheric Radar Returns Classification with Feature Engineering-Based Light Gradient Boosting Machine Algorithm. In Proceedings of the 2023 3rd International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI), Wuhan, China, 15–17 December 2023; pp. 528–532.