

Article

Simple yet Effective Ensemble Feature Selection Using Hierarchical Binning

Jinho Park, Dohun Kim  and Wonjong Kim * 

Electronics and Telecommunications Research Institute, Daejeon 34129, Republic of Korea; jinho.park@etri.re.kr (J.P.); dohun@etri.re.kr (D.K.)

* Correspondence: wjkim@etri.re.kr

Featured Application

The proposed EFSHB framework provides a robust and efficient feature selection solution for high-dimensional learning tasks. By mitigating model-dependent bias and promoting stable predictive performance through ensemble-based feature aggregation, EFSHB is particularly well-suited for applications in biomedical gene-expression analysis, wearable sensor and physiological signal processing, smart manufacturing systems, and high-dimensional industrial monitoring.

Abstract

Feature selection is essential for improving classification performance and reducing overfitting in high-dimensional learning tasks. However, conventional importance-based methods often suffer from instability, model bias, and sensitivity to threshold settings. To address these limitations, we propose EFSHB (Ensemble Feature Selection using Hierarchical Binning), a hybrid ensemble framework that integrates importance-based sorting, bin-level greedy evaluation, iterative hierarchical refinement, and union-based integration of model-wise selected features. At each iteration, five tree-based models independently perform bin-wise greedy selection, and their selected subsets are merged through a union operation to form the feature set for the next iteration. This iterative process progressively refines the feature space while mitigating model-specific bias and promoting robust predictive performance across heterogeneous models. EFSHB was evaluated on nine high-dimensional benchmark datasets, including biomedical gene-expression, synthetic, proteomics, and speech-feature data. Across all datasets, EFSHB achieved the highest or near-highest classification accuracy, outperforming traditional Greedy Feature Selection (GFS), binning-based GFS (GFSB), and hierarchical binning GFS (GFSHB). On average, EFSHB improved accuracy for all classifiers, achieving mean gains of 14.0% over GFS and 13.3% over GFSHB. EFSHB also provided balanced feature reduction by avoiding excessive feature retention while preserving complementary informative features identified across models. In terms of computational efficiency, EFSHB reduced average feature selection time from 266 min (GFS) to 11 min, corresponding to a 24-fold speed-up. These results demonstrate that EFSHB achieves robust predictive performance and high computational efficiency, making it suitable for diverse high-dimensional applications.

Keywords: ensemble feature selection; AI-driven data processing; machine learning for data management; high-dimensional data analysis; big data analytics; hierarchical binning



Academic Editor: Douglas O'Shaughnessy

Received: 30 November 2025

Revised: 4 January 2026

Accepted: 14 January 2026

Published: 31 March 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Feature selection (FS) is essential for removing irrelevant or redundant features in high-dimensional data while simultaneously improving classification performance and model interpretability [1–7]. As an effective preprocessing step, FS enhances machine learning efficiency, reduces the risk of overfitting, and improves the interpretability of predictive models [4,7,8].

The primary objective of FS is to identify informative and task-relevant features while eliminating unnecessary or duplicated ones, thereby improving classification accuracy [1,5]. In real-world applications—such as sensor signal analysis, physiological signal monitoring, and smart manufacturing—datasets often contain hundreds or thousands of correlated features, from which only a compact subset of features contributing to generalization must be extracted [4,7,9]. This process reduces computational complexity, reduces memory usage, accelerates training and inference, and enhances both model reliability and interpretability. For these reasons, FS has been widely applied in data mining, pattern recognition, and machine learning [4,7,9–11].

Minimum Redundancy Maximum Relevance (mRMR) is a widely used filter-based feature selection method that selects features by maximizing their relevance to the target variable while minimizing redundancy among selected features. By approximating the maximal dependency criterion using pairwise mutual information, mRMR provides an efficient solution for high-dimensional data, particularly in biomedical and gene-expression analysis [12].

ReliefF is an instance-based feature selection algorithm that evaluates feature relevance based on how well feature values distinguish between neighboring instances of different classes. By capturing local data characteristics and exhibiting robustness to noise and multi-class problems, ReliefF has been widely adopted for large-scale and noisy datasets [13].

Lasso introduces ℓ_1 -norm regularization into linear models to achieve simultaneous feature selection and model fitting by enforcing sparsity in feature coefficients. Although effective in producing compact and interpretable models, Lasso is sensitive to feature correlation and often selects only one feature from highly correlated groups [14].

Robust Feature Selection (RFS) based on joint $\ell_{2,1}$ -norm minimization enforces row-wise sparsity to select features consistently across samples while mitigating noise effects. Despite its robustness, RFS typically relies on convex optimization frameworks, which may lead to increased computational cost in large-scale settings [15].

Despite their effectiveness, most classical feature selection methods are optimized within a single learning framework or objective function, which may result in limited robustness to model-dependent bias and instability when applied to complex high-dimensional data.

In addition to importance-based and ensemble-based strategies, a growing body of research has investigated optimization-driven feature selection methods for high-dimensional data. Optimization-based feature selection methods have also been proposed, including hybrid frameworks that combine Elastic Net-based selection with optimization techniques for high-dimensional data classification [16]. Similarly, global optimization-based feature selection methods, such as GSMFO, employ population-based global search strategies and have been applied to machine learning tasks including intrusion detection [17].

Traditional feature importance-based methods leverage the built-in importance metrics of tree-based models such as Random Forest [18], XGBoost [19], and Decision Tree [20] to select features efficiently [7,21,22]. However, these approaches suffer from several fundamental limitations. First, the selected feature set is strongly influenced by the inherent bias of each learning model, causing different models to produce substantially inconsistent feature subsets [4,7,10]. Second, slight variations in importance ranking can directly affect

the selection outcome, thereby increasing sensitivity to ranking fluctuations and reducing robustness [9,23,24]. Third, because these methods depend on predefined thresholds (e.g., selecting the top-K features or features above a certain percentile), they exhibit high sensitivity to threshold choices, leading to reproducibility and robustness issues [10,25–27].

Existing importance-based feature selection methods rely solely on the internal metrics of a single model, which leads to substantial disagreement among models and large variability in the resulting feature subsets [4]. Because each model exhibits a distinct feature importance distribution, it is common for different models to select entirely different subsets of features even when applied to the same dataset. Moreover, single-model approaches are sensitive to noise and easily influenced by local optima, which can ultimately degrade the generalization performance of the downstream classifier.

Since these methods depend more on relative ranking differences rather than the absolute magnitude of feature importance, they may inadvertently exclude truly informative features or include irrelevant ones. Approaches based on a single criterion—such as selecting the top-K features or selecting features above a predefined percentile—also suffer from low reproducibility, as the optimal threshold varies across datasets.

As a result, single-model feature selection methods often struggle to achieve robust generalization and consistent predictive performance and performance consistency. Meanwhile, existing ensemble-based feature selection methods that aggregate feature importance values—through averaging or weighted combinations—attempt to reduce model-specific bias by integrating information from multiple learners. However, these methods still rely on threshold settings, and some require predefined feature groups, which limits their practical flexibility. Therefore, there is a growing need for a hybrid ensemble feature selection framework that can reduce threshold dependency, alleviate model bias, and achieve robust and reliable predictive performance.

In this study, we propose a new hybrid ensemble feature selection framework, named Ensemble Feature Selection using Hierarchical Binning (EFSHB). EFSHB integrates importance-based sorting, bin-level greedy evaluation, and union-based aggregation across heterogeneous tree-based models (AdaBoost [28], Extra Trees [29], Random Forest, XGBoost, and Decision Tree). A bin-based greedy feature selection procedure (GFSB) has been accepted for publication at the ICCE conference. The proposed EFSHB framework extends this work by incorporating hierarchical binning, iterative refinement, and union-based aggregation across multiple models [30].

For each model, features are first sorted by importance and divided into equal-width bins. A cumulative greedy search identifies the bin combination that yields the highest classification accuracy. The features selected by each model are then merged through a union operation, forming the feature set for the next iteration.

This sorting–binning–evaluation cycle repeats iteratively. As the process continues, the feature space becomes progressively refined, and when the remaining number of features is smaller than the number of bins, an adaptive binning mechanism automatically adjusts the bin count. The iterative process terminates when the size of the union feature set no longer decreases between iterations, indicating convergence. The model–feature pair that achieves the highest accuracy across all iterations is selected as the final output.

EFSHB offers several advantages:

- Reduced model bias through union-based integration of multiple learners.
- Lower threshold sensitivity via bin-level cumulative search.
- Deterministic convergence, as the feature space is gradually reduced until it stops shrinking.
- High simplicity and compatibility, requiring only model-provided importance values.

- Joint model–feature selection, enabling automatic identification of the most effective configuration.

2. Ensemble Feature Selection Using Hierarchical Binning (EFSHB)

2.1. Overview of EFSHB

The proposed EFSHB operates by sorting features in descending importance, grouping them into bins, and iteratively refining the feature subset using complementary information from multiple learning models. First, the features are sorted based on importance scores computed from each model, and the sorted feature list is divided into predefined bins. Each model then performs bin-level greedy selection, evaluating cumulative bins to identify the optimal bin range. The feature subsets selected by individual models are merged through a union operation, and the resulting union subset becomes the input for the next iteration. Through this iterative process, the feature space becomes progressively more compact and refined, and once the subset size no longer decreases, the best-performing model–feature pair across all iterations is selected as the final output.

The overall pipeline is illustrated in Figure 1. As shown in Figure 1b, iteration index t denotes the current step, with $U^{(t)}$ and $U^{(t-1)}$ representing the feature subsets at iterations t and previous iteration $t - 1$, respectively, and T indicating the maximum number of iterations. Figure 1a details the bin-based greedy feature selection (GFSB) procedure, including computing importance scores, sorting, binning, cumulative bin expansion, and performance evaluation.

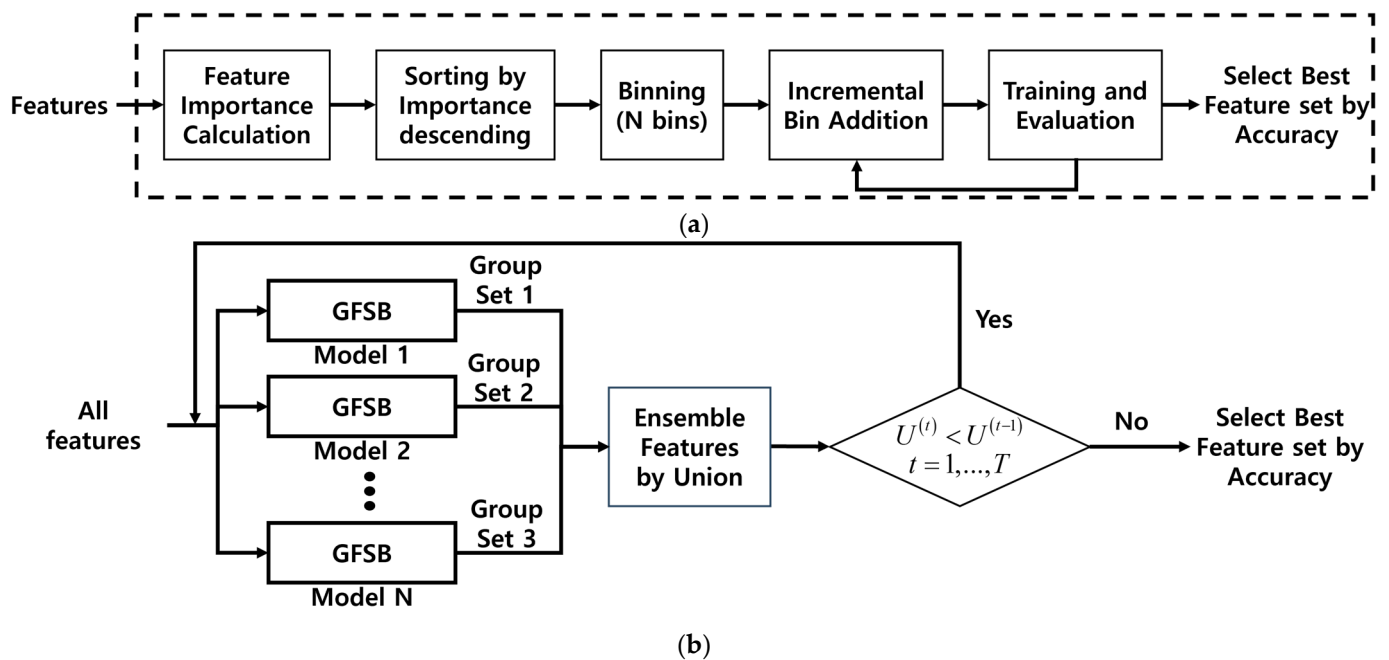


Figure 1. Overview of the proposed EFSHB framework. (a) Bin-based greedy feature selection (GFSB): computing importance scores, sorting, binning, and evaluating cumulative bins to identify the optimal bin range. (b) Iterative hierarchical binning framework of EFSHB: merging model-wise selected features into $U^{(t)}$ and refining them across iterations until convergence. where t denotes the iteration index ($t = 1, \dots, T$), and $U^{(t)}$ and $U^{(t-1)}$ represent the union feature sets at the current and previous iterations, respectively. The process iteratively refines the feature space until convergence.

Accordingly, EFSHB follows a coherent five-stage pipeline that includes importance-based sorting, bin-based grouping, model-wise greedy selection, union-driven iterative refinement, and final model–feature selection, thereby enabling robust and efficient feature selection while effectively leveraging the complementary strengths of multiple models.

2.2. Feature Importance-Based Binning and Bin-Wise Greedy Evaluation

In each iteration, five learning models—Random Forest (RF), Extra Trees (ET), Adaboost (AB), Decision Tree (DT), and XGBoost (XGB)—are used to compute feature importance values. After training, the importance scores are sorted in descending order, and the resulting sorted feature list $F = \{f_1, f_2, \dots, f_m\}$ is evenly divided into N bins according to a user-defined bin number. Each bin corresponds to a continuous group of features arranged by their sorted importance order, and the entire feature space can be expressed as follows:

$$F = \{B_1, B_2, \dots, B_N\}, \tag{1}$$

where B_i denotes the set of features assigned to the i -th bin.

In the second step, a bin-wise greedy evaluation is performed. Starting from the most important bin B_1 , bins are accumulated sequentially to construct progressively expanding feature subsets. The subset used in the k -th evaluation is defined as follows:

$$S_k = \bigcup_{i=1}^k B_i, 1 \leq k \leq N \tag{2}$$

For each value of k , the subset S_k is constructed by cumulatively including the top k bins, meaning that $k = 1$ corresponds to using only the most important bin, while $k = N$ corresponds to using all bins. For each accumulated subset S_k learning model h is trained, and the classification accuracy is computed as follows:

$$Acc(k) = Accuracy(h(S_k)), 1 \leq k \leq N \tag{3}$$

For each base classifier $h \in \{RF, ET, AB, DT, XGB\}$, the classification accuracy $Acc(k)$ is evaluated to identify the optimal bin index k^* . The optimal bin index that yields the highest accuracy, k^* , is obtained as follows:

$$k^* = \operatorname{argmax}_k Acc(k), \tag{4}$$

Accordingly, the optimal bin index k^* may differ across base classifiers $h \in \{RF, ET, AB, DT, XGB\}$, and each model selects its own optimal feature subset $S_h^* = S_{k^*}$

2.3. Union-Based Multi-Model Feature Integration

In the third step, the feature subsets selected by the five models—RF, ET, AB, DT, and XGB—are integrated through a union operation. Let $S_{RF}^*, S_{ET}^*, S_{AB}^*, S_{DT}^*$, and S_{XGB}^* denote the optimal feature subsets obtained from each model. The integrated feature set U is defined as follows:

$$U = S_{RF}^* \cup S_{ET}^* \cup S_{AB}^* \cup S_{DT}^* \cup S_{XGB}^*, \tag{5}$$

The union operation mitigates disagreement among models, recovers informative features that may have been under-selected, and enhances robustness by aggregating evidence from heterogeneous learners. The resulting union set U is then used as the feature space for the next iteration. Through this iterative update, the algorithm incrementally refines the feature space, converging toward a more stable and robust subset over successive iterations.

2.4. Iterative Hierarchical Refinement and Final Model-Feature Selection

FSHB progressively refines the feature space through an iterative hierarchical process. At each iteration t , the union feature set $U^{(t-1)}$ obtained from the previous iteration is

used as the input. Based on this subset, feature importance values are recomputed, sorted in descending order, and partitioned into N bins using equal-width binning. When the number of remaining features is smaller than the predefined bin count—causing potential irregularity in bin structure—an adaptive binning strategy automatically reduces the number of bins, maintaining a stable partitioning scheme.

After binning, the algorithm performs bin-wise greedy evaluation to obtain the selected subset $S_h^{*(t)}$ for each model h . The union of these model-specific selections is defined as follows:

$$U^{(t+1)} = \bigcup_h S_h^{*(t)}, \quad (6)$$

which becomes the input for the next iteration. Through this repeated process, the feature space is incrementally refined, and the distribution of feature importance becomes increasingly consistent as the iterations proceed.

The iterative process terminates when the union feature sets no longer change between consecutive iterations, that is, when $U^{(t)} = U^{(t-1)}$. After convergence, the classification accuracies recorded for all models across all iterations $t = 1, 2, \dots, T$ are compared. The algorithm then selects the iteration–model pair (t^*, h^*) that achieves the highest accuracy. EFSHB adopts h^* as the final classifier and the feature subset selected by that model at iteration t^* , $S_{h^*}^{*(t^*)}$ as the final feature set.

In this manner, EFSHB performs feature selection and model selection simultaneously within a unified iterative framework, enabling the method to identify the most effective and accurate combination of model and feature subset by leveraging performance information accumulated across iterations.

3. Experimental Results

In this study, we evaluate the performance of the proposed EFSHB method using nine representative high-dimensional datasets. To analyze the individual contributions of binning, iterative refinement, and union-based aggregation, EFSHB is compared with three baseline feature selection methods:

- Greedy Feature Selection (GFS), which performs cumulative greedy search without binning;
- binning-based GFS (GFSB), which incorporates equal-width binning to examine the impact of bin partitioning;
- Greedy Feature Selection with Hierarchical Binning (GFSHB), which applies iterative refinement but excludes the union operation.

Through comparative experiments among these four methods, we comprehensively assess the roles of bin grouping, hierarchical refinement, and union-driven feature integration within the proposed EFSHB framework.

3.1. Data Description

The experiments are conducted using nine high-dimensional benchmark datasets: Madelon [31], CLL-SUB-111 [32], Lung [33], TOX-171 [34], Colon [35], GLI-85 [36], Prostate-GE [37], Arcene [38], and Isolet [39]. The characteristics of each dataset, including the number of samples, features, classes, class imbalance and description, are summarized in Table 1.

All datasets are publicly available and can be downloaded from the scikit-feature repository (<https://jundongl.github.io/scikit-feature/datasets.html> and https://jundongl.github.io/scikit-feature/OLD/datasets_old.html). All datasets were last accessed on 27 November 2025).

Table 1. Overview of datasets and class imbalance characteristics (# indicates counts).

Dataset	# Samples	# Features	# Classes	Class Imbalance	Description
Madelon [31]	2600	500	2	No	Artificially generated benchmark dataset from the NIPS 2003 Feature Selection Challenge
CLL-SUB-111 [32]	111	11,340	3	Yes	High-dimensional microarray gene expression dataset derived from leukemia patients
Lung [33]	203	3312	5	Yes	Microarray gene expression benchmark dataset
TOX-171 [34]	171	5748	4	No	Toxicogenomics gene expression dataset collected by the U.S. Environmental Protection Agency and the National Toxicology Program
Colon [35]	62	2000	2	Yes	Early and influential gene expression benchmark dataset originally published in PNAS
GLI-85 [36]	85	22,283	2	Yes	Glioma-related gene expression dataset
Prostate-GE [37]	102	5966	2	No	Gene expression dataset derived from prostate tissue samples
Arcene [38]	200	10,000	2	Yes	Hybrid benchmark dataset from the NIPS 2003 Feature Selection Challenge
Isolet [39]	1560	617	26	No	Speech feature dataset composed of numerical representations of spoken alphabet recordings

3.2. Effectiveness of the Proposed Method

To evaluate the efficiency and effectiveness of the proposed EFSHB method, we constructed several comparison feature selection approaches. This section describes the configuration and roles of these baseline methods and clarifies which components of EFSHB each baseline is designed to verify.

First, we selected five tree-based learning models—RF, ET, AB, DT, and XGB—all of which provide feature importance measures required for importance-based sorting and binning in EFSHB. All experiments were implemented in Python (v3.10) using standard machine learning libraries. Tree-based models, including RF, ET, AB, DT, and XGBoost, were implemented using the scikit-learn (v1.6.1) and XGBoost (v3.0.2) libraries with their default feature importance definitions.

No custom modification was applied to the internal importance calculation of each model, ensuring reproducibility and fair structural comparison among feature selection methods. For all experiments, the initial number of bins was set to $N = 10$, and the maximum number of iterations T for both EFSHB and GFSHB was fixed to 30. Each dataset was randomly split into training and testing sets with a ratio of 8:2. Each dataset was randomly split into training and testing sets with a ratio of 8:2, while preserving the original class distribution. For each base classifier, the same hyperparameter configuration was consistently applied across all datasets.

Feature selection was performed exclusively on the training set, while the test set was used only for performance evaluation. To ensure a fair comparison, identical training

and testing splits were applied to all evaluated algorithms. In addition, all features were normalized using the MinMaxScaler prior to training and testing.

- **GFS:** GFS is a forward selection-based method that sorts all features in descending order according to their importance values and incrementally adds individual features while evaluating classification performance at each step. As a purely greedy approach, it does not employ binning or iterative refinement, and the process continues until all features have been evaluated. This exhaustive behavior enables GFS to serve as a baseline for assessing feature selection performance in the absence of bin-based grouping. The procedure of GFS is illustrated in Figure 2.

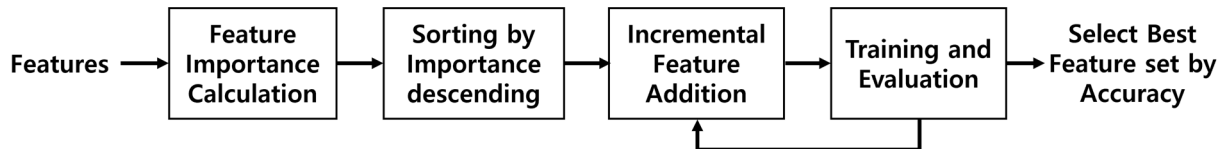


Figure 2. Flowchart of greedy feature selection (GFS).

- **GFSB:** GFSB performs greedy selection after dividing the sorted feature list into a predefined number of bins. Classification performance is evaluated by cumulatively adding bins rather than individual features. Since GFSB operates in a single-pass manner without iterative refinement, it is used to evaluate the effectiveness of introducing bin-based grouping. The algorithmic flow of GFSB is illustrated in Figure 1a.
- **GFSHB:** GFSHB extends GFSB by incorporating iterative hierarchical refinement, where bin-wise greedy selection is repeated at each iteration. The algorithm terminates when the selected feature subset remains unchanged between consecutive iterations. Unlike EFSHB, GFSHB does not apply any union operation across models, allowing us to isolate and examine the contribution of the iterative refinement process. The overall structure of GFSHB is illustrated in Figure 3.

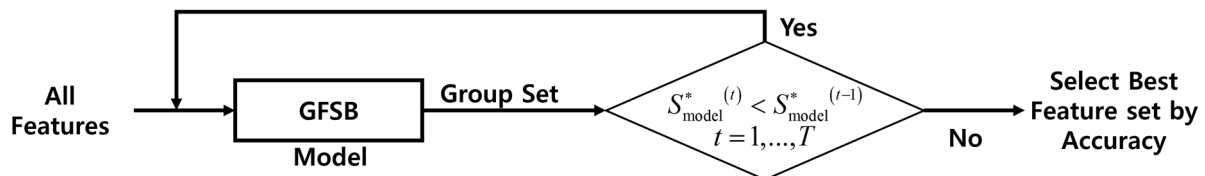


Figure 3. Greedy feature selection using hierarchical binning (GFSHB). S_{model}^* denotes the optimal feature subsets obtained from the classification model.

3.2.1. Overall Performance Comparison

In this subsection, we provide an overall assessment of the proposed EFSHB by summarizing and comparing the results of the four feature selection methods—GFS, GFSB, GFSHB, and EFSHB. Table 2 presents, for each method, the classification accuracy, number of selected features, and feature selection time, based on the classifier that achieved the highest accuracy among the five models. Specifically, the results reported in Table 2 are summarized from Tables 4–6. For each dataset and feature selection method, the classifier–feature selection combination achieving the highest classification accuracy was selected. In cases of identical accuracy, the result with fewer selected features was preferred, and if a tie still remained, the combination with shorter feature selection time was chosen. In addition, Table 3 shows the accuracy of classifiers trained on the full set of original features without applying any feature selection. These results serve as a reference point for evaluating the performance improvements achieved through the feature selection methods.

Table 2. Classification accuracy, number of selected features, and feature selection time for each feature selection method.

Dataset	GFS		GFSB		GFSHB		EFSHB	
	Best Model	Acc.	Best Model	Acc.	Best Model	Acc.	Best Model	Acc.
	Selected Features	CPU Time	Selected Features	CPU Time	Selected Features	CPU Time	Selected Features	CPU Time
Madelon	RF	89.04	RF	85.38	RF	89.04	XGB	90.38
	20	106 min 13 s	50	2 min 27 s	18	11 min 30 s	16	16 min 42 s
CLL-SUB-111	ET	82.61	RF	82.61	RF	82.61	RF	89.96
	21	57 s	9072	24 s	3	1 min 48 s	591	13 min 57 s
Lung	XGB	97.56	RF	97.56	RF	97.56	DT	100
	4	347 min 26 s	332	24 s	6	1 min 59 s	4	9 min 29 s
TOX-171	RF	97.14	RF	97.14	RF	97.14	RF	100
	194	204 min	1150	31 s	133	3 min 51 s	612	18 min 10 s
Colon	XGB	92.31	XGB	84.62	ET	92.31	RF	92.31
	4	4 min 49 s	200	2 s	4	0.2 s	10	3 min 40 s
GLI-85	ET	88.24	RF	82.35	AB	82.35	ET	88.24
	132	2 min 45 s	4458	26 s	3	47 s	4000	6 min 28 s
Prostate-GE	AB	100	XGB	95.24	AB	100	XGB	100
	17	83 min 20 s	597	14 s	17	12 s	6	6 min 35 s
Arcene	XGB	90.00	XGB	87.50	RF	92.50	XGB	92.50
	29	804 min	1000	29 s	24	3 min 5 s	42	15 min 6 s
Isolet	RF	90.71	RF	90.38	RF	91.67	RF	92.31
	195	23 min 57 s	617	1 min 33 s	159	23 min 38 s	369	9 min 43 s

Table 3. Classification accuracy of classifiers trained without feature selection.

Dataset	Random Forest	Extra Trees	AdaBoost	Decision Tree	XGBoost
Madelon	73.08	47.12	61.92	77.31	79.04
CLL-SUB-111	69.57	73.91	69.57	65.22	78.26
Lung	92.68	78.05	80.49	90.24	87.8
TOX-171	94.29	51.43	51.43	71.43	80
Colon	84.62	61.54	76.92	61.54	84.62
GLI-85	82.35	76.47	76.47	70.59	70.59
Prostate-GE	90.48	71.43	85.71	85.71	95.24
Arcene	85	60	80	72.5	80
Isolet	91.35	63.46	25.96	77.24	88.78

Overall, all four feature selection methods achieve higher accuracy than the classifiers trained without feature selection in Table 3 for most datasets. This demonstrates that removing irrelevant or redundant features and identifying a compact subset of informative features improves generalization performance in high-dimensional settings.

A consistent trend is observed in which classification performance improves as the structural complexity of the method increases from GFS to GFSHB and ultimately to EFSHB. However, GFSB—which applies a single-stage bin-based grouping—often yields lower performance. In contrast, GFSHB, with its iterative refinement process, and EFSHB, with its additional union-based integration across models, achieve higher accuracy than both GFS and GFSB for most datasets.

3.2.2. Classification Accuracy Comparison

In this study, classification accuracy was used as the performance evaluation metric.

Accuracy is defined as the proportion of correctly classified samples among all samples and is calculated as follows:

$$\text{Accuracy} = \frac{\text{number of correctly classified samples}}{\text{total number of samples}} \times 100\% \quad (7)$$

Table 4 presents the classification accuracies of GFS, GFSB, GFSHB, and the proposed EFSHB across all datasets. Using the same five classifiers—RF, ET, AB, DT, and XGB—we evaluate the direct impact of each feature selection method on classification performance.

Overall, EFSHB achieves the highest accuracy for most datasets or maintains performance comparable to existing methods. Except for a few cases—specifically, the AB-based EFSHB results for CLL-SUB-111, Lung, and TOX-171, which were slightly lower than those of GFSHB—the proposed method generally preserves or improves accuracy across classifier–dataset combinations. Notably, DT and ET on Lung, RF on TOX-171, and RF, ET, and XGB on Prostate-GE all reached 100% accuracy under EFSHB. Furthermore, for datasets such as Madelon, Arcene, Isolet, and GLI-85, EFSHB consistently outperformed GFS, GFSB, and GFSHB.

Table 4. Classification accuracy of the four feature selection methods across all datasets.

Base	FS Method	Dataset								
		Madelon	CLL-SUB-111	Lung	TOX-171	Colon	GLI-85	Prostate-GE	Arcene	Isolet
RF	Base	73.08	69.57	92.68	94.29	84.62	82.35	90.48	85	91.35
	GFS	89.04	82.61	95.12	97.14	92.31	76.47	95.24	90	90.71
	GFSB	85.38	82.61	97.56	97.14	84.62	82.35	90.48	85	90.38
	GFSHB	89.04	82.61	97.56	97.14	92.31	82.35	95.24	92.5	91.67
	EFSHB	89.23	86.96	97.56	100	92.31	82.35	100	92.5	92.31
ET	Base	47.12	73.91	78.05	51.43	61.54	76.47	71.43	60	63.46
	GFS	72.88	82.61	95.12	77.14	92.31	88.24	95.24	90	70.19
	GFSB	58.65	60.87	85.37	62.86	84.62	76.47	85.71	85	62.18
	GFSHB	75.38	78.26	92.68	71.43	92.31	82.35	95.24	85	67.63
	EFSHB	82.12	82.61	100	77.14	92.31	88.24	100	92.5	75.64
AB	Base	61.92	69.57	80.49	51.43	76.92	76.47	85.71	80	25.96
	GFS	66.35	82.61	82.93	62.86	84.62	76.47	100	87.5	25.96
	GFSB	61.92	73.91	80.49	51.43	76.92	76.47	85.71	80	25.96
	GFSHB	65.96	78.26	87.8	65.71	84.62	82.35	100	82.5	25.96
	EFSHB	66.73	73.91	80.49	57.14	84.62	76.47	100	87.5	25.96
DT	Base	77.31	65.22	90.24	71.43	61.54	70.59	85.71	72.5	77.24
	GFS	81.92	69.57	97.56	74.29	84.62	76.47	95.24	87.5	80.13
	GFSB	78.27	65.22	92.68	71.43	69.23	76.47	95.24	82.5	78.53
	GFSHB	82.31	69.57	97.56	71.43	76.92	76.47	95.24	85	78.85
	EFSHB	83.08	69.57	100	74.29	84.62	76.47	95.24	85	80.13
XGB	Base	79.04	78.26	87.80	80.00	84.62	70.59	95.24	80.00	88.78
	GFS	87.88	78.26	97.56	85.71	92.31	76.47	95.24	90.00	90.38
	GFSB	84.23	78.26	87.80	82.86	84.62	70.59	95.24	87.50	89.42
	GFSHB	88.85	78.26	95.12	82.86	84.62	76.47	95.24	87.50	89.74
	EFSHB	90.38	78.26	95.12	88.57	92.31	76.47	100	92.50	90.06

GFS achieves high accuracy in certain cases—such as AB on CLL-SUB-111 and Prostate-GE—but generally shows lower performance on other datasets. GFSB, which applies single-stage binning, improves accuracy for RF on Lung and GLI-85; however, for most classifier–dataset combinations, its performance tends to remain unchanged or decrease, indicating limited robustness across different classifier–dataset combinations. GFSHB generally outperforms both GFS and GFSB, although its effectiveness varies with dataset characteristics. Taken together, Table 4 shows that EFSHB consistently achieves high accuracy across a wide range of datasets and delivers the best overall performance among the four evaluated feature selection methods.

In addition to dataset-wise comparisons, we further examined classifier-wise average performance to assess how consistently EFSHB improves accuracy across different learning algorithms. For each classifier, the accuracies obtained using GFS and GFSHB across all nine datasets were averaged and compared with the average accuracy of EFSHB. The results show consistent performance gains across all classifiers. When using AB, EFSHB improved accuracy by 27.3% over GFS and 27.7% over GFSHB. With DT, EFSHB achieved 11.7% and 11.3% higher accuracy relative to GFS and GFSHB, respectively. For RF, the improvement was 4.6% over both methods, while ET achieved 20.75% and 18.25% higher accuracy. With XGB, EFSHB outperformed GFS and GFSHB by 5.8% and 4.8%, respectively. These results confirm that EFSHB provides stable and consistent improvements across all classifier architectures.

Figure 4 presents the classification accuracy of five classifiers (RF, ET, AB, DT, and XGB) on the Madelon dataset when applying different feature selection methods (Base, GFS, GFSB, GFSHB, and EFSHB). Overall, all feature selection methods either maintain or improve accuracy compared with the Base model. Although GFSB shows a temporary decrease in accuracy relative to GFS, both GFSHB and the proposed EFSHB consistently recover or surpass performance across most classifiers. Notably, EFSHB achieves the highest accuracy among all classifier–method combinations, demonstrating the effectiveness of hierarchical refinement and model-wise union integration in improving predictive performance.

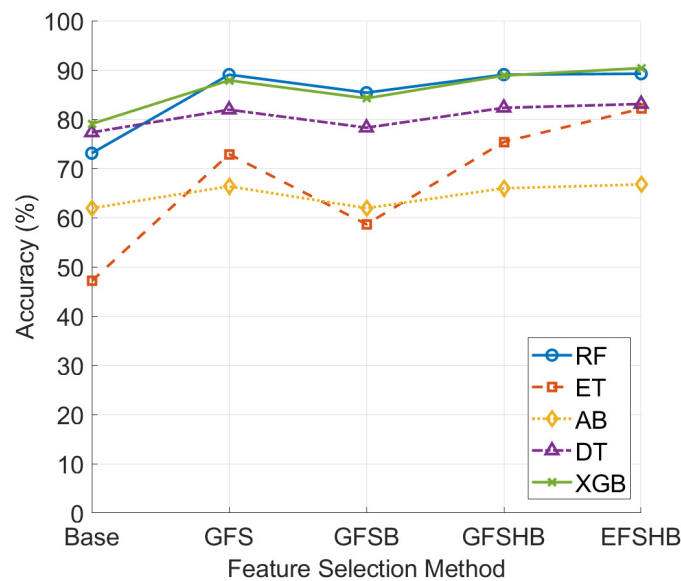


Figure 4. Classification accuracy of five classifiers on the Madelon dataset using different feature selection methods.

3.2.3. Feature Reduction Comparison

Table 5 compares the number of final selected features obtained by GFS, GFSB, GFSHB, and the proposed EFSHB across all datasets. Although all four methods substantially reduced the dimensionality of the original feature space, their selection behaviors differed considerably depending on the method.

Because GFSB determines the feature subset solely through a single-stage binning operation without iterative refinement, low-importance features included within a bin are often selected as part of the final subset. Across all datasets, GFSB consistently selected the largest number of features among all feature selection methods, and in high-dimensional cases such as CLL-SUB-111, GLI-85, Arcene, and TOX-171, it occasionally selected thousands—or even tens of thousands—of features.

In contrast, GFSHB repeatedly reconstructs the bin structure and updates the feature subset through iterative refinement, often yielding markedly fewer features than GFSB. For example, in the Arcene dataset, while GFSB selected up to 10,000 features, GFSHB reduced this number to a range of 3–196 features. Similar reductions were observed for CLL-SUB-111, GLI-85, and Prostate-GE.

Table 5. Number of selected features obtained by the four feature selection methods across all datasets.

Base	FS Method	Dataset								
		Madelon	CLL-SUB-111	Lung	TOX-171	Colon	GLI-85	Prostate-GE	Arcene	Isolet
RF	GFS	20	10	3	194	11	10	2	160	195
	GFSB	50	9072	332	1150	200	4458	597	2000	617
	GFSHB	18	3	6	133	9	1338	2	24	159
	EFSHB	14	591	5	612	10	1000	8	105	369
ET	GFS	8	21	135	154	4	132	47	168	108
	GFSB	50	1134	2650	4600	600	4458	4178	9000	434
	GFSHB	5	228	68	167	4	90	3	196	72
	EFSHB	20	699	56	79	22	4000	15	5633	174
AB	GFS	3	15	3	8	33	4	17	19	19
	GFSB	50	11,340	332	575	200	2229	597	1000	62
	GFSHB	4	10	4	6	3	3	17	3	20
	EFSHB	4	170	12	27	34	1000	8	19	35
DT	GFS	9	41	14	3	69	18	9	33	174
	GFSB	100	2268	1326	2300	1400	17,827	2985	10,000	310
	GFSHB	8	110	5	7	1	14,263	2985	48	124
	EFSHB	20	248	4	103	86	1000	31	450	220
XGB	GFS	8	19	4	117	17	5	5	29	145
	GFSB	50	1134	332	1150	200	2229	597	1000	186
	GFSHB	10	4	4	16	5	201	3	1000	150
	EFSHB	16	85	7	138	36	2456	6	42	122

The proposed EFSHB incorporates a union operation that aggregates features selected by multiple models at each iteration, allowing informative features discarded by one model to be reintroduced by others. As a result, EFSHB typically selects fewer features than GFSB while producing larger subsets than GFSHB, with the exact subset size depending on dataset characteristics.

In summary, Table 5 shows that GFSB tends to retain excessively large feature subsets due to its single-stage binning structure, whereas GFSHB achieves the strongest dimensionality reduction through iterative refinement. Positioned between these two extremes, EFSHB provides a balanced level of feature reduction by combining hierarchical refinement with model-wise aggregation.

3.2.4. FS Execution Time Comparison

Table 6 presents the execution time of GFS, GFSB, GFSHB, and the proposed EFSHB for each dataset. Overall, GFS incurs the highest computational cost because greedy selection is applied repeatedly at the individual-feature level. For high-dimensional datasets, its runtime frequently ranged from several hours to more than a full day. For example, execution time exceeded 2584 min on CLL-SUB-111, 1521 min on GLI-85, and 804 min on Arcene.

In contrast, GFSB performs a single-pass greedy search based on binning without iterative refinement, resulting in very fast execution times—typically from a few seconds to a few minutes across most datasets. For datasets such as Colon, Lung, Madelon, and Prostate-GE, feature selection with GFSB was completed within 1 s to 3 min for nearly all classifiers.

GFSHB employs an iterative binning structure; however, because the feature space is rapidly reduced at each iteration, its overall execution time is significantly lower than that of GFS. Even for high-dimensional datasets such as GLI-85 and Arcene, runtime was limited to a few seconds or minutes, representing one- to two-order-of-magnitude reductions relative to GFS. For instance, on CLL-SUB-111, the XGBoost-based GFS required approximately 2584 min, whereas GFSHB completed feature selection in about one minute. Similarly, on GLI-85, AdaBoost- and XGBoost-based GFS took more than 1436 min, while GFSHB reduced the runtime to under one minute.

The proposed EFSHB aggregates the model-wise feature subsets through a union operation and performs iterative refinement, producing a single integrated execution time per dataset. As shown in Table 6, its total execution time ranged from several minutes to several tens of minutes depending on dataset dimensionality. For example, EFSHB required 6 min 28 s for GLI-85 and up to 18 min 10 s for TOX-171, which showed the longest execution time among the evaluated datasets.

To further quantify computational efficiency, the average execution times of GFS and EFSHB were compared. The mean runtime of GFS, computed by averaging per-classifier runtime across the nine datasets and then averaging across classifiers, was 266 min 21 s. In contrast, because EFSHB produces a single unified runtime per dataset, its average execution time across the nine datasets was 11 min 6 s. This indicates that EFSHB reduces computational cost by more than twenty-fold while maintaining strong accuracy and performance robustness. The most substantial reduction occurred for XGBoost on CLL-SUB-111, where GFS required 2584 min, whereas EFSHB completed the process in only 13 min 57 s—an approximately 185-fold speedup.

In summary, Table 6 shows that GFS incurs the highest computational cost and GFSB provides the shortest execution time, whereas EFSHB offers a balanced and scalable execution profile—drastically reducing computation time compared with GFS while requiring only a modest increase over GFSHB to incorporate multi-model information effectively.

Figure 5 compares the execution time of the four feature selection methods (GFS, GFSB, GFSHB, and EFSHB) using XGBoost on five high-dimensional datasets. Consistent with Table 6, GFS shows the highest computational cost, whereas GFSB achieves the shortest runtime through single-pass binning. GFSHB further reduces execution time by evaluating features at the bin level. Although EFSHB runs slightly longer than GFSHB due to union-based integration, it still provides over a ten-fold speedup compared with GFS while maintaining comparable classification performance.

Table 6. Feature selection execution time for each method across all datasets.

Base	FS Method	Dataset								
		Madelon	CLL-SUB-111	Lung	TOX-171	Colon	GLI-85	Prostate-GE	Arcene	Isolet
RF	GFS	106 min 13 s	349 min 11 s	203 min 58 s	204 min	32 min 14 s	1251 min 18 s	137 min 9 s	411 min 12 s	23 min 57 s
	GFSB	2 min 27 s	24 s	24 s	31 s	10 s	26 s	17 s	24 s	1 min 33 s
	GFSHB	11 min 31 s	1 min 48 s	1 min 59 s	3 min 51 s	1 min 18 s	1 min 32 s	1 min 1 s	3 min 5 s	23 min 38 s
	EFSHB	16 min 42 s	13 min 57 s	9 min 29 s	18 min 10 s	3 min 40 s	6 min 28 s	6 min 35 s	15 min 6 s	9 min 43 s
ET	GFS	6 s	57 s	15 s	21 s	5 s	2 min 45 s	19 s	1 min 8 s	4 s
	GFSB	1 s	3 s	1 s	1 s	1 s	10 s	1 s	2 s	1 s
	GFSHB	1 s	3 s	1 s	2 s	1 s	12 s	2 s	6 s	1 s
	EFSHB	16 min 42 s	13 min 57 s	9 min 29 s	18 min 10 s	3 min 40 s	6 min 28 s	6 min 35 s	15 min 6 s	9 min 43 s
AB	GFS	10 min 58 s	546 min 40 s	86 min 27 s	172 min 27 s	3 min 38 s	1521 min 6 s	83 min 20 s	408 min 39 s	23 min 58 s
	GFSB	11 s	25 s	11 s	22 s	1 s	40 s	7 s	21 s	16 s
	GFSHB	17 s	1 min 20 s	17 s	29 s	5 s	47 s	12 s	26 s	24 s
	EFSHB	16 min 42 s	13 min 57 s	9 min 29 s	18 min 10 s	3 min 40 s	6 min 28 s	6 min 35 s	15 min 6 s	9 min 43 s
DT	GFS	2 min 27 s	25 min 52 s	4 min 59 s	12 min 35 s	9 s	57 min 16 s	3 min 55 min	30 min 19 s	3 min 18 s
	GFSB	3 s	4 s	1 s	2 s	1 s	12 s	1 s	4 s	2 s
	GFSHB	4 s	5 s	1 s	4 s	1 s	39 s	2 s	15 s	5 s
	EFSHB	16 min 42 s	13 min 57 s	9 min 29 s	18 min 10 s	3 min 40 s	6 min 28 s	6 min 35 s	15 min 6 s	9 min 43 s
XGB	GFS	20 min 4 s	2584 min	347 min 26 s	653 min 48 s	4 min 49 s	1436 min 42 s	121 min 56 s	804 min	196 min 28 s
	GFSB	41 s	50 s	56 s	1 min 41 s	2 s	41 s	14 s	29 s	2 min 7 s
	GFSHB	1 min 33 s	1 min 11 s	1 min 13 s	3 min 8 s	11 s	58 s	21 s	46 s	5 min 35 s
	EFSHB	16 min 42 s	13 min 57 s	9 min 29 s	18 min 10 s	3 min 40 s	6 min 28 s	6 min 35 s	15 min 6 s	9 min 43 s

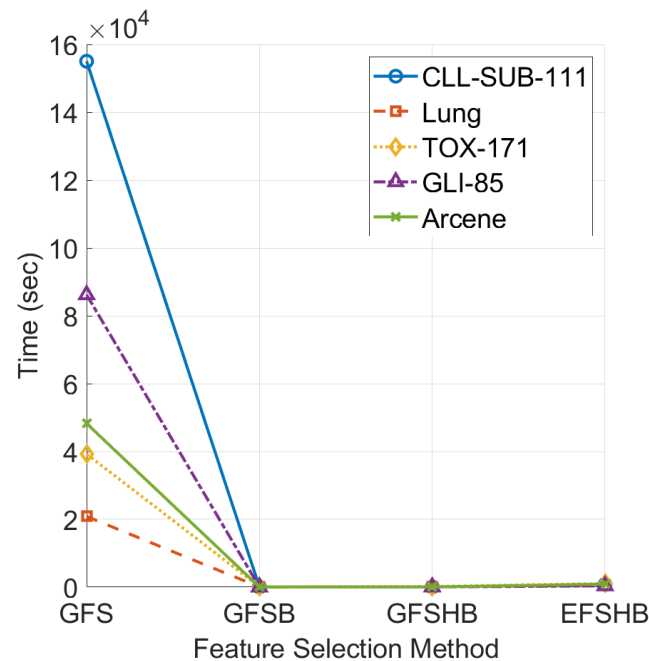


Figure 5. Runtime comparison of GFS, GFSB, GFSHB, and EFSHB using XGBoost.

3.2.5. Comparison with Existing Feature Selection Methods

To evaluate the effectiveness of the proposed EFSHB method, we conducted comparative experiments against several widely used feature selection algorithms, including mRMR [12], ReliefF [13], Lasso [14], and RFS [15]. These methods were selected to represent different feature selection paradigms, covering both filter-based and embedded approaches.

The feature subsets selected by each method (mRMR, ReliefF, Lasso and RFS) were used as input to a common Support Vector Machine (SVM) [40] classifier for performance evaluation. In contrast, the proposed EFSHB method follows its own ensemble learning framework, where feature selection and model construction are jointly determined by multiple base classifiers.

For mRMR and ReliefF, feature importance scores were computed on the training data, and the top-ranked features were selected accordingly. Lasso-based feature selection was performed by training a linear model with ℓ_1 -norm regularization, where features with non-zero coefficients were retained. The regularization parameter was determined via internal cross-validation. RFS is an embedded feature selection method based on $\ell_{2,1}$ -norm sparsity regularization, which selects features directly during model optimization.

In all methods, feature selection was performed exclusively on the training data, while the test data were used only for classification performance evaluation.

Table 7 presents the classification accuracy achieved by each feature selection method across multiple datasets. As shown in the table, the proposed EFSHB method consistently outperforms the existing feature selection methods on all evaluated datasets. In particular, EFSHB demonstrates substantially higher accuracy on high-dimensional datasets such as CLL-SUB-111 and TOX-171, where the performance gap between EFSHB and competing methods is especially pronounced. These results indicate that the superior performance of EFSHB stems from the proposed ensemble feature selection strategy rather than from classifier choice.

Table 7. Classification accuracy (%) comparison between EFSHB and existing feature selection methods.

FS Method	CLL-SUB-111	Lung	TOX-171	Isolet
mRMR-SVM	77.47	94.09	80.12	90.83
ReliefF-SVM	72.07	93.10	83.04	89.10
Lasso-SVM	79.28	93.60	74.27	94.23
RFS-SVM	81.98	94.58	84.80	95.19
EFSHB	89.96	100	100	92.31

3.3. Cross-Validation Performance Analysis

As a complementary evaluation to the fixed train–test split experiments, 5-fold cross-validation was conducted to assess the robustness and stability of the proposed method across different data partitions.

For each fold, feature selection was performed on the training subset, followed by classifier training and evaluation on the corresponding validation subset. The classification performance was then averaged across the five folds, and the mean and standard deviation were reported.

Table 8 presents the cross-validation classification accuracy, while Table 9 reports the balanced accuracy, which accounts for potential class imbalance. Across most datasets, the proposed EFSHB method achieves the highest or near-highest mean accuracy and balanced accuracy compared to competing feature selection methods.

Balanced accuracy is computed independently for each cross-validation fold as the average of class-wise recall values, thereby compensating for skewed class distributions.

For a binary classification task, the balanced accuracy for each fold is defined as follows:

$$BA^{(f)} = \frac{1}{2} \left(\frac{TP^{(f)}}{TP^{(f)} + FN^{(f)}} + \frac{TN^{(f)}}{TN^{(f)} + FP^{(f)}} \right) \quad (8)$$

where $TP^{(f)}$, $TN^{(f)}$, $FP^{(f)}$, and $FN^{(f)}$ denote the numbers of true positives, true negatives, false positives, and false negatives in the f -th cross-validation fold. The balanced accuracy values reported in Table 9 correspond to the mean and standard deviation of $BA^{(f)}$ over the five cross-validation folds.

Notably, while many competing feature selection methods and their corresponding base classifiers exhibit relatively large standard deviations across cross-validation folds, EFSHB consistently shows the lowest or among the lowest performance variance, indicating more stable predictive performance under different data partitions.

These performance advantages are particularly pronounced in datasets with severe class imbalance, such as CLL-SUB-111, Lung, Colon, and GLI-85. In these datasets, EFSHB not only demonstrates comparable or improved balanced accuracy relative to standard accuracy, but also achieves the most substantial reduction in performance variance, consistently yielding the lowest standard deviation across cross-validation folds.

A comparison between Tables 8 and 9 further reveals that the performance trends observed in terms of classification accuracy are largely preserved when balanced accuracy is considered. This indicates that the performance gains of EFSHB are not driven solely by majority-class dominance, but are robustly maintained across class distributions.

Table 8. Mean and standard deviation of classification accuracy obtained by 5-fold cross-validation.

FS Method	Base	Madelon	CLL-SUB-111	Lung	TOX-171	Colon	GLI-85	Prostate-GE	Arcene	Isolet
EFSHB	Union	89.81 ±0.84	91.03 ±3.97	98.04 ±1.10	88.89 ±4.82	98.33 ±3.73	98.82 ±2.63	97.05 ±2.70	92.00 ±2.09	95.45 ±1.16
	Base	62.69 ±2.08	72.17 ±10.35	80.84 ±7.98	66.10 ±6.62	78.97 ±4.65	87.06 ±7.67	92.19 ±5.52	74.50 ±4.11	20.13 ±2.26
Base	DT	74.69 ±2.79	62.21 ±5.55	82.23 ±4.22	64.32 ±9.65	67.82 ±18.34	74.12 ±5.26	80.38 ±3.57	67.00 ±8.91	79.23 ±1.23
	ET	52.54 ±1.99	59.49 ±11.43	78.84 ±4.30	53.75 ±9.42	71.15 ±11.36	76.47 ±11.76	66.76 ±13.45	69.50 ±6.22	61.92 ±5.40
	RF	71.88 ±2.29	79.25 ±6.20	93.12 ±3.98	78.42 ±6.07	77.56 ±10.26	87.06 ±8.72	92.19 ±4.24	82.00 ±6.47	94.68 ±1.23
	XGB	81.08 ±1.14	71.23 ±6.50	92.15 ±4.67	77.19 ±3.21	79.10 ±8.94	83.53 ±4.92	90.14 ±5.98	83.00 ±6.22	92.24 ±1.64
	AB	65.12 ±2.87	79.37 ±7.81	82.80 ±7.10	76.05 ±4.61	85.38 ±3.91	89.41 ±7.67	95.14 ±6.82	81.50 ±5.76	25.19 ±2.96
GFS	DT	82.58 ±1.97	73.91 ±5.63	93.12 ±2.66	73.71 ±7.33	82.18 ±5.95	88.24 ±7.20	91.14 ±4.22	79.00 ±8.40	82.24 ±1.37
	ET	64.54 ±2.95	87.39 ±1.98	94.60 ±2.01	74.27 ±1.27	98.46 ±3.44	95.29 ±4.92	95.14 ±3.37	87.00 ±3.26	68.40 ±1.10
	RF	89.46 ±0.69	91.03 ±4.40	96.09 ±3.27	85.41 ±5.74	88.72 ±4.36	92.94 ±7.67	94.14 ±6.28	84.00 ±7.62	94.87 ±1.22
	XGB	88.92 ±1.22	81.15 ±4.50	94.60 ±2.65	86.55 ±2.59	85.38 ±7.07	91.76 ±5.26	94.10 ±4.07	87.50 ±5.86	92.56 ±1.80
	AB	62.69 ±2.08	73.99 ±11.23	80.84 ±7.98	66.10 ±6.62	78.97 ±4.65	87.06 ±7.67	92.19 ±5.52	75.00 ±3.06	20.13 ±2.26
GFSB	DT	76.42 ±3.50	68.46 ±9.11	89.67 ±1.96	67.87 ±10.25	77.31 ±14.68	87.06 ±4.92	88.19 ±5.79	76.50 ±6.02	81.73 ±0.88
	ET	60.31 ±1.99	71.19 ±9.88	88.17 ±3.72	60.84 ±3.59	85.51 ±8.75	85.88 ±3.22	84.24 ±7.39	81.50 ±3.79	65.32 ±1.70
	RF	85.81 ±1.47	82.85 ±6.02	94.62 ±5.28	84.24 ±3.15	85.38 ±7.07	88.24 ±9.30	93.19 ±5.48	84.50 ±5.70	95.26 ±1.10
	XGB	85.96 ±1.56	78.42 ±3.39	93.13 ±4.66	83.04 ±4.35	79.10 ±8.94	85.88 ±7.89	90.14 ±7.80	86.50 ±5.76	92.56 ±1.49
	AB	64.46 ±3.37	80.24 ±6.59	85.73 ±7.04	74.86 ±1.50	83.72 ±6.14	89.41 ±7.67	96.10 ±5.35	83.50 ±3.35	23.08 ±1.55
GFSHB	DT	82.85 ±1.80	73.00 ±8.34	91.15 ±2.76	71.34 ±5.25	82.18 ±9.95	84.71 ±7.89	90.14 ±5.06	79.50 ±6.47	81.92 ±1.21
	ET	74.77 ±3.90	77.51 ±5.35	91.65 ±2.73	67.88 ±4.59	93.59 ±3.60	89.41 ±7.67	93.19 ±5.48	83.50 ±5.18	66.67 ±2.51
	RF	89.62 ±0.84	89.21 ±6.04	95.60 ±4.00	87.75 ±4.70	85.38 ±8.93	96.47 ±5.26	95.14 ±5.83	88.00 ±6.22	95.32 ±1.29
	XGB	89.38 ±1.09	81.98 ±3.23	94.60 ±2.65	84.79 ±5.28	85.38 ±7.07	90.59 ±7.89	93.10 ±4.40	88.00 ±4.11	92.63 ±1.62

Table 9. Mean and standard deviation of balanced accuracy obtained by 5-fold cross-validation.

FS Method	Base	Madelon	CLL-SUB-111	Lung	TOX-171	Colon	GLI-85	Prostate-GE	Arcene	Isolet
EFSHB	Union	89.81 ±0.84	92.05 ±2.94	96.76 ±4.24	87.27 ±2.81	98.75 ±2.80	98.00 ±4.47	97.09 ±2.66	91.91 ±2.13	95.45 ±1.16
	Base	62.69 ±2.08	77.99 ±10.18	57.71 ±20.77	66.49 ±6.73	78.75 ±6.61	83.00 ±11.68	92.27 ±5.30	74.10 ±3.92	20.13 ±2.26
Base	DT	74.69 ±2.79	61.24 ±9.41	65.03 ±11.85	64.30 ±9.80	65.00 ±22.20	69.76 ±9.41	80.36 ±3.58	67.09 ±8.93	79.23 ±1.23
	ET	52.54 ±1.99	68.18 ±10.19	71.29 ±14.52	54.06 ±8.68	70.25 ±15.55	72.44 ±14.33	66.45 ±13.52	69.17 ±6.80	61.92 ±5.40
	RF	71.88 ±2.29	84.64 ±4.46	81.79 ±13.45	78.64 ±5.89	74.75 ±12.26	80.17 ±12.51	92.27 ±3.95	81.69 ±6.79	94.68 ±1.23
	XGB	81.08 ±1.14	71.02 ±13.26	79.98 ±15.65	77.53 ±2.77	74.25 ±10.52	77.33 ±7.01	90.27 ±5.87	82.56 ±6.22	92.24 ±1.64
	AB	65.12 ±2.87	84.96 ±5.59	70.20 ±13.72	76.65 ±4.50	85.25 ±4.37	88.17 ±9.47	95.27 ±6.60	81.22 ±5.85	25.19 ±2.96
GFS	DT	82.58 ±1.97	79.34 ±6.40	87.35 ±11.27	74.01 ±7.80	80.25 ±20.28	85.67 ±9.04	91.18 ±4.15	78.97 ±8.70	82.24 ±1.37
	ET	64.54 ±2.95	90.85 ±1.44	93.92 ±2.83	74.06 ±1.58	98.75 ±2.80	95.42 ±5.79	95.27 ±3.22	87.52 ±3.35	68.40 ±1.10
	RF	89.46 ±0.69	93.38 ±3.28	92.45 ±7.13	85.62 ±5.59	87.75 ±5.03	89.83 ±10.55	94.27 ±6.06	84.02 ±7.69	94.87 ±1.22
	XGB	88.92 ±1.22	85.56 ±3.41	92.88 ±4.56	87.01 ±2.62	84.00 ±9.90	89.83 ±5.15	94.18 ±3.98	87.39 ±5.83	92.56 ±1.80
	AB	62.69 ±2.08	79.33 ±10.92	57.71 ±20.46	66.49 ±6.73	78.75 ±6.61	83.00 ±11.68	92.27 ±5.30	74.55 ±2.97	20.13 ±2.26
GFSB	DT	76.42 ±3.50	74.67 ±6.49	84.34 ±10.00	68.31 ±10.51	76.00 ±17.49	84.83 ±8.00	88.18 ±5.79	76.25 ±6.51	81.73 ±0.88
	ET	60.31 ±1.99	78.02 ±8.36	84.51 ±5.24	60.98 ±3.52	84.75 ±8.12	85.09 ±5.03	84.36 ±7.42	80.94 ±3.52	65.32 ±1.70
	RF	85.81 ±1.47	87.38 ±4.38	87.79 ±15.29	84.34 ±2.89	85.25 ±4.37	83.33 ±13.27	93.27 ±5.24	84.38 ±5.88	95.26 ±1.10
	XGB	85.96 ±1.56	78.23 ±8.11	85.31 ±13.07	83.15 ±3.85	74.25 ±10.52	81.00 ±10.63	90.27 ±7.71	86.16 ±5.90	92.56 ±1.49
	AB	64.46 ±3.37	85.43 ±4.93	72.79 ±15.89	75.33 ±1.55	80.25 ±10.77	87.00 ±9.29	96.18 ±5.24	82.46 ±3.24	23.08 ±1.55
GFSHB	DT	82.85 ±1.80	78.67 ±7.66	80.92 ±13.54	71.28 ±5.66	77.75 ±22.25	80.50 ±9.37	90.18 ±5.02	79.43 ±6.93	81.92 ±1.21
	ET	74.77 ±3.90	76.63 ±4.04	80.70 ±7.57	68.15 ±4.70	93.75 ±4.42	86.67 ±10.59	93.27 ±5.24	83.24 ±5.02	66.67 ±2.51
	RF	89.62 ±0.84	92.05 ±4.45	90.59 ±11.02	87.69 ±4.40	82.75 ±8.68	95.17 ±6.78	95.27 ±5.57	87.74 ±6.33	95.32 ±1.29
	XGB	89.38 ±1.09	84.11 ±6.47	90.31 ±5.43	84.98 ±4.96	82.75 ±8.68	86.67 ±11.37	93.18 ±4.34	93.61 ±4.17	92.63 ±1.62

3.4. Feature Selection Stability Analysis Using Jaccard Index

In this section, we analyze the stability of feature selection using the Jaccard index from two complementary perspectives. First, we examine the variability in feature selection outcomes across different base classifiers within the same feature selection framework, aiming to quantify model-dependent selection behavior. Second, we evaluate the stability of feature selection with respect to data partitioning by measuring the consistency of selected feature subsets across cross-validation folds.

To quantify the similarity between feature subsets, the Jaccard index is computed in a pairwise manner.

Given two feature subsets S_i and S_j , the Jaccard similarity is defined as follows:

$$J(S_i, S_j) = \frac{|S_i \cap S_j|}{|S_i \cup S_j|} \tag{9}$$

For a collection of M feature subsets obtained either from different base classifiers or from different cross-validation folds, the Jaccard index is computed for all possible pairs (S_i, S_j) with $i < j$.

The overall feature selection stability is then quantified by averaging the pairwise Jaccard values as follows:

$$\bar{J} = \frac{2}{M(M-1)} \sum_{i=1}^{M-1} \sum_{j=i+1}^M J(S_i, S_j) \tag{10}$$

where M denotes the number of feature subsets being compared, obtained either from different base classifiers or from different cross-validation folds.

The mean and standard deviation of the pairwise Jaccard similarities are reported to assess feature selection stability.

3.4.1. Model-Dependent Feature Selection Behavior

To investigate the dependency of feature selection on the underlying learning algorithm, we analyze the inter-model Jaccard similarity within the GFS framework. Specifically, feature subsets selected by different base classifiers (AB, DT, ET, RF, and XGBoost) are compared pairwise using the Jaccard index.

Table 10 reports the mean and standard deviation of the inter-model Jaccard similarity across datasets. The results show that the overlap between feature sets selected by different models is generally low, indicating substantial disagreement among base classifiers regarding feature importance.

Table 10. Mean and standard deviation of inter-model Jaccard similarity for GFS across different base classifiers.

Fold	Madelon	CLL-SUB-111	Lung	TOX-171	Colon	GLI-85	Prostate-GE	Arcene	Isolet
Fold 1	0.059 ±0.060	0.042 ±0.070	0.155 ±0.096	0.091 ±0.082	0.090 ±0.098	0.030 ±0.038	0.024 ±0.037	0.045 ±0.050	0.259 ±0.207
Fold 2	0.041 ±0.043	0.015 ±0.037	0.139 ±0.110	0.200 ±0.332	0.032 ±0.032	0.051 ±0.076	0.133 ±0.293	0.034 ±0.038	0.134 ±0.155
Fold 3	0.021 ±0.033	0.014 ±0.020	0.162 ±0.077	0.090 ±0.085	0.110 ±0.191	0.058 ±0.092	0.101 ±0.300	0.064 ±0.059	0.313 ±0.161
Fold 4	0.072 ±0.142	0.125 ±0.301	0.177 ±0.102	0.051 ±0.101	0.065 ±0.077	0.052 ±0.059	0.077 ±0.037	0.128 ±0.154	0.249 ±0.168
Fold 5	0.073 ±0.136	0.080 ±0.122	0.117 ±0.056	0.045 ±0.058	0.111 ±0.175	0.041 ±0.032	0.026 ±0.032	0.075 ±0.109	0.313 ±0.231

This observation suggests that the feature selection outcome of GFS is strongly influenced by the choice of the underlying classifier, reflecting pronounced model-dependent variability in single-model importance-based selection.

3.4.2. Cross-Validation-Based Feature Selection Stability

In addition to model-dependent variability, we assess the stability of feature selection under different data partitions. For each method, the Jaccard index is computed pairwise between feature subsets obtained from different folds in a 5-fold cross-validation setting.

The resulting values are averaged, and the mean and standard deviation are reported. Table 11 summarizes the cross-validation-based Jaccard similarity for all feature selection methods and datasets. Single-model GFS approaches exhibit noticeable variability in feature overlap across folds, whereas binning-based methods tend to produce more consistent feature subsets.

Table 11. Mean and standard deviation of cross-validation-based Jaccard similarity for feature selection.

FS Method	Base	Madelon	CLL-SUB-111	Lung	TOX-171	Colon	GLI-85	Prostate-GE	Arcene	Isolet
EFSHB	Union	0.425 ±0.279	0.097 ±0.092	0.152 ±0.120	0.180 ±0.108	0.165 ±0.058	0.066 ±0.073	0.117 ±0.102	0.140 ±0.102	0.788 ±0.134
GFS	AB	0.249 ±0.115	0.082 ±0.035	0.147 ±0.099	0.214 ±0.038	0.082 ±0.103	0.022 ±0.032	0.165 ±0.126	0.071 ±0.017	0.535 ±0.053
	DT	0.467 ±0.076	0.081 ±0.188	0.187 ±0.130	0.200 ±0.190	0.012 ±0.027	0.131 ±0.115	0.160 ±0.286	0.117 ±0.171	0.143 ±0.051
	ET	0.079 ±0.018	0.390 ±0.325	0.288 ±0.243	0.435 ±0.050	0.146 ±0.160	0.196 ±0.137	0.122 ±0.244	0.210 ±0.167	0.140 ±0.023
	RF	0.886 ±0.061	0.111 ±0.076	0.218 ±0.083	0.206 ±0.135	0.117 ±0.193	0.065 ±0.052	0.246 ±0.147	0.156 ±0.065	0.667 ±0.119
	XGB	0.368 ±0.084	0.042 ±0.022	0.231 ±0.076	0.071 ±0.018	0.067 ±0.087	0.102 ±0.152	0.280 ±0.285	0.055 ±0.017	0.319 ±0.063
GFSB	AB	0.521 ±0.037	0.693 ±0.300	0.914 ±0.017	0.913 ±0.005	0.748 ±0.016	0.960 ±0.005	0.897 ±0.018	0.651 ±0.333	0.716 ±0.045
	DT	0.295 ±0.147	0.396 ±0.305	0.490 ±0.264	0.595 ±0.215	0.543 ±0.321	0.386 ±0.308	0.482 ±0.237	0.533 ±0.311	0.333 ±0.290
	ET	0.086 ±0.017	0.615 ±0.252	0.725 ±0.137	0.883 ±0.069	0.396 ±0.291	0.672 ±0.137	0.670 ±0.225	0.452 ±0.239	0.475 ±0.176
	RF	0.474 ±0.036	0.187 ±0.022	0.508 ±0.070	0.298 ±0.155	0.370 ±0.032	0.128 ±0.018	0.350 ±0.008	0.262 ±0.030	0.546 ±0.243
	XGB	0.233 ±0.044	0.381 ±0.185	0.460 ±0.207	0.377 ±0.224	0.526 ±0.245	0.523 ±0.322	0.902 ±0.016	0.641 ±0.146	0.561 ±0.179
GFSHB	AB	0.411 ±0.201	0.066 ±0.040	0.099 ±0.083	0.188 ±0.072	0.125 ±0.172	0.063 ±0.077	0.120 ±0.094	0.054 ±0.052	0.437 ±0.172
	DT	0.409 ±0.098	0.037 ±0.099	0.029 ±0.058	0.036 ±0.037	0.033 ±0.100	0.070 ±0.155	0.202 ±0.293	0.016 ±0.030	0.151 ±0.080
	ET	0.284 ±0.131	0.000 ±0.001	0.090 ±0.172	0.121 ±0.185	0.000 ±0.000	0.001 ±0.003	0.030 ±0.034	0.007 ±0.017	0.160 ±0.169
	RF	0.715 ±0.157	0.112 ±0.071	0.198 ±0.182	0.116 ±0.116	0.335 ±0.258	0.287 ±0.285	0.487 ±0.274	0.120 ±0.099	0.522 ±0.206
	XGB	0.785 ±0.056	0.018 ±0.019	0.142 ±0.126	0.074 ±0.089	0.065 ±0.052	0.062 ±0.087	0.242 ±0.259	0.059 ±0.142	0.425 ±0.217

The proposed EFSHB method achieves moderate and consistent overlap across cross-validation folds, indicating stable feature selection behavior without enforcing overly rigid agreement among folds.

4. Discussion

4.1. Interpretation of Key Findings

The experimental results demonstrate that the proposed EFSHB method consistently achieves superior classification performance across a wide range of high-dimensional datasets. In particular, EFSHB attains the highest or near-highest classification accuracy and balanced accuracy in most cases, while exhibiting smaller performance variations under cross-validation compared to competing feature selection methods. These observations

indicate that the performance gains of EFSHB are not dataset-specific artifacts, but rather reflect the robustness of the proposed framework.

The cross-validation results further highlight the stability of EFSHB with respect to data partitioning. Across multiple datasets, EFSHB shows lower standard deviations across folds, suggesting reduced sensitivity to variations in training and validation splits. This behavior is especially evident in datasets with pronounced class imbalance, such as CLL-SUB-111, Lung, Colon, and GLI-85, where both classification accuracy and balanced accuracy remain consistently high. Together, these findings confirm that EFSHB provides robust generalization performance across diverse and challenging high-dimensional scenarios.

4.2. Feature Selection Stability and Predictive Performance

An important observation from the experimental analysis is that EFSHB does not necessarily yield the highest feature overlap as measured by the Jaccard index, particularly when comparisons are made across different classifiers or cross-validation folds. This result highlights a fundamental distinction between feature selection stability and predictive performance. While a high Jaccard similarity reflects consistency in the identities of selected features, it does not inherently guarantee improved classification accuracy.

In contrast, EFSHB prioritizes the selection of complementary and performance-relevant features, even when the exact identities of selected features vary across models or data partitions. By avoiding overly restrictive constraints on feature overlap, the proposed method is able to adapt to variations in data while maintaining stable predictive performance. These results suggest that enforcing strict feature identity consistency may be suboptimal in high-dimensional settings, where multiple feature subsets can provide comparable discriminative power.

4.3. Mitigating Model-Dependent Bias via Union-Based Aggregation

The inter-model Jaccard analysis presented in Table 10 provides further insight into the limitations of single-model feature selection approaches. In particular, methods such as GFS exhibit substantial variability in the selected feature subsets depending on the underlying classifier, resulting in consistently low feature overlap across models. This observation offers empirical evidence of strong model-dependent bias in importance-based feature selection, where the selected features are heavily influenced by the inductive biases of individual learning algorithms.

EFSHB addresses this limitation through union-based aggregation of feature subsets selected by multiple heterogeneous classifiers. By integrating features identified as important by different models, EFSHB preserves diverse yet informative features that may be overlooked by any single classifier. As a result, the proposed framework effectively mitigates model-dependent bias and leverages complementary information across classifiers, leading to improved robustness and generalization performance.

4.4. Structural Trade-Offs: Feature Size and Computational Efficiency

From the perspective of feature subset characteristics, the compared methods exhibit distinct selection behaviors. GFSB tends to retain a large number of low-importance features due to its single-pass binning structure, resulting in overly large feature subsets. In contrast, GFSHB progressively reduces the feature space through iterative refinement, while EFSHB produces intermediate yet balanced feature subsets by reintroducing informative features identified by multiple models through union-based aggregation. This strategy allows EFSHB to maintain feature diversity while avoiding excessive redundancy.

Execution-time analysis reflects these structural differences among the methods. GFS requires the longest runtime because it evaluates features individually, leading to substantial computational overhead in high-dimensional settings. By operating at the bin

level, GFSB and GFSHB significantly reduce runtime. Although EFSHB incurs a modest additional computational cost due to union operations across models, it still achieves considerable efficiency gains compared to GFS, resulting in a favorable trade-off between computational cost and performance.

4.5. Limitations and Future Work

Despite the strong empirical performance of EFSHB, several limitations and directions for future work remain. In the current framework, the number of bins used in the hierarchical binning process is treated as a fixed hyperparameter. Although this setting was shown to be effective across a wide range of datasets, the optimal number of bins may vary depending on data characteristics such as feature dimensionality, sample size, and feature distribution.

Future work will focus on developing adaptive strategies for automatically determining the optimal bin configuration. Possible directions include data-driven criteria based on stability–performance trade-offs, validation-based optimization, or information-theoretic measures. Incorporating such adaptive binning mechanisms could further enhance the robustness and flexibility of the proposed framework while reducing the need for manual parameter tuning.

In addition, the proposed framework does not explicitly incorporate correlation-aware filtering mechanisms to remove highly collinear or redundant features. While the union-based aggregation strategy is designed to preserve complementary features identified by heterogeneous models and to enhance performance stability, it may retain features that convey overlapping information. As discussed in Sections 4.2 and 4.3, the primary focus of EFSHB is performance robustness rather than strict redundancy minimization, and the incorporation of explicit correlation or redundancy-aware filtering strategies remains an important direction for future work.

5. Conclusions

In this study, we proposed EFSHB, a hybrid ensemble feature selection framework that integrates hierarchical binning with union-based aggregation across multiple heterogeneous classifiers. Extensive experiments on high-dimensional datasets demonstrated that EFSHB consistently achieves superior classification accuracy and balanced accuracy while maintaining strong performance stability under cross-validation.

The experimental analysis further revealed that robust predictive performance does not necessarily require strict feature identity consistency. By aggregating complementary features selected by different models, EFSHB effectively mitigates model-dependent bias and achieves stable generalization across diverse data partitions. This characteristic highlights the importance of balancing feature diversity and redundancy in ensemble-based feature selection.

While the proposed framework employs a fixed bin configuration in this study, future work will focus on developing adaptive strategies to automatically determine optimal bin structures based on data characteristics. Such extensions are expected to further enhance the flexibility and robustness of EFSHB for real-world high-dimensional applications.

Author Contributions: Conceptualization, J.P., D.K. and W.K.; methodology, J.P. and W.K.; Software, J.P. and D.K.; validation, J.P. and W.K.; formal analysis, J.P. and W.K.; investigation, J.P. and D.K.; resources, W.K.; data curation, J.P. and D.K.; writing—original draft preparation, J.P.; writing—review and editing, J.P., D.K. and W.K.; visualization, J.P.; supervision, W.K.; project administration, W.K.; funding acquisition, W.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Ministry of Trade, Industry and Energy (MOTIE) and the Korea Evaluation Institute of Industrial Technology (KEIT) under grant number [RS-2024-00487113], the Ministry of SMEs and Startups project “Development of Open Smart Manufacturing Sharing Platform for Enterprise Linkage in Discrete Process Characteristic Industries” [RS-2022-00140586].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets used in this study are publicly available and can be downloaded from the scikit-feature repository at: <https://jundongl.github.io/scikit-feature/datasets.html> and https://jundongl.github.io/scikit-feature/OLD/datasets_old.html. All datasets were last accessed on 27 November 2025.

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

Abbreviations

The following abbreviations are used in this manuscript:

AB	AdaBoost
DT	Decision Tree
EFSHB	Ensemble Feature Selection using Hierarchical Binning
ET	Extra Trees
FS	Feature Selection
GFS	Greedy Feature Selection
GFSB	Greedy Feature Selection with Binning
GFSHB	Greedy Feature Selection using Hierarchical Binning
RF	Random Forest
XGB	XGBoost

References

1. Natarajan, K.; Baskaran, D.; Kamalanathan, S. An adaptive ensemble feature selection technique for model-agnostic diabetes prediction. *Sci. Rep.* **2025**, *15*, 6907. [CrossRef]
2. AlZu'bi, S.; Elbes, M.; Mughaid, A.; Bdair, N.; Abualigah, L.; Forestiero, A.; Zitar, R.A. Diabetes monitoring system in smart health cities based on big data intelligence. *Future Internet* **2023**, *15*, 85. [CrossRef]
3. Lee, H.D.; Mendes, A.I.; Spolaor, N.; Oliva, J.T.; Parmezan, A.R.S.; Wu, F.C.; Fonseca-Pinto, R. Dermoscopic assisted diagnosis in melanoma: Reviewing results, optimizing methodologies and quantifying empirical guidelines. *Knowl. Based Syst.* **2018**, *158*, 9–24. [CrossRef]
4. Gómez-Martínez, V.; Chushig-Muzo, D.; Veierød, M.B.; Granja, C.; Soguero-Ruiz, C. Ensemble feature selection and tabular data augmentation with generative adversarial networks to enhance cutaneous melanoma identification and interpretability. *BioData Min.* **2024**, *17*, 46. [CrossRef] [PubMed]
5. BenSaid, F.; Alimi, A.M. Online feature selection system for big data classification based on multi-objective automated negotiation. *Pattern Recognit.* **2021**, *110*, 107629. [CrossRef]
6. Chandrashekar, G.; Sahin, F. A survey on feature selection methods. *Comput. Electr. Eng.* **2014**, *40*, 16–28. [CrossRef]
7. Pudjihartono, N.; Fadason, T.; Kempa-Liehr, A.W.; O'Sullivan, J.M. A review of feature selection methods for machine learning-based disease risk prediction. *Front. Bioinform.* **2022**, *2*, 927312. [CrossRef]
8. Kumar, A.; Kaur, A.; Singh, P.; Driss, M.; Boulila, W. Efficient multiclass classification using feature selection in high-dimensional datasets. *Electronics* **2023**, *12*, 2290. [CrossRef]
9. Xiang, F.; Zhao, Y.; Zhang, M.; Zuo, Y.; Zou, X.; Tao, F. Ensemble learning-based stability improvement method for feature selection towards performance prediction. *J. Manuf. Syst.* **2024**, *74*, 55–67. [CrossRef]
10. Bolón-Canedo, V.; Alonso-Betanzos, A. Ensembles for feature selection: A review and future trends. *Inf. Fusion* **2019**, *52*, 1–12. [CrossRef]
11. Xu, Z.; Huang, G.; Weinberger, K.Q.; Zheng, A.X. Gradient boosted feature selection. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 24–27 August 2014; pp. 522–531.

12. 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](#)]
13. Kira, K.; Rendell, L.A. A practical approach to feature selection. In *Machine Learning Proceedings 1992*; Elsevier: Amsterdam, The Netherlands, 1992; pp. 249–256.
14. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1996**, *58*, 267–288. [[CrossRef](#)]
15. Nie, F.; Huang, H.; Cai, X.; Ding, C. Efficient and robust feature selection via joint ℓ_2 , 1-norms minimization. *Adv. Neural Inf. Process. Syst.* **2010**, *23*, 1813–1821.
16. Qaraad, M.; Amjad, S.; Manhrawy, I.I.; Fathi, H.; Hassan, B.A.; El Kafrawy, P. A hybrid feature selection optimization model for high dimension data classification. *IEEE Access* **2021**, *9*, 42884–42895. [[CrossRef](#)]
17. Hussein, N.K.; Qaraad, M.; Amjad, S.; Farag, M.; Hassan, S.; Mirjalili, S.; Elhosseini, M.A. Enhancing feature selection with GMSMFO: A global optimization algorithm for machine learning with application to intrusion detection. *J. Comput. Des. Eng.* **2023**, *10*, 1363–1389. [[CrossRef](#)]
18. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
19. 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, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
20. Loh, W.Y. Classification and regression trees. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2011**, *1*, 14–23. [[CrossRef](#)]
21. Htun, H.H.; Biehl, M.; Petkov, N. Survey of feature selection and extraction techniques for stock market prediction. *Financ. Innov.* **2023**, *9*, 26. [[CrossRef](#)]
22. Ma, Z. Ensemble Feature Selection Using Neighbourhood Rough Set-Based Multicriterion Fusion. *J. Appl. Math.* **2024**, *2024*, 5534285. [[CrossRef](#)]
23. Kalousis, A.; Prados, J.; Hilario, M. Stability of feature selection algorithms. In Proceedings of the Fifth IEEE International Conference on Data Mining (ICDM'05), Houston, TX, USA, 27–30 November 2005; p. 8.
24. Zappia, L.; Richter, S.; Ramírez-Suástegui, C.; Kfuri-Rubens, R.; Vornholz, L.; Wang, W.; Dietrich, O.; Frishberg, A.; Luecken, M.D.; Theis, F.J. Feature selection methods affect the performance of scRNA-seq data integration and querying. *Nat. Methods* **2025**, *22*, 834–844. [[CrossRef](#)]
25. Bolón-Canedo, V.; Sánchez-Marroño, N.; Alonso-Betanzos, A. A review of feature selection methods on synthetic data. *Knowl. Inf. Syst.* **2013**, *34*, 483–519. [[CrossRef](#)]
26. Seijo-Pardo, B.; Bolón-Canedo, V.; Alonso-Betanzos, A. On developing an automatic threshold applied to feature selection ensembles. *Inf. Fusion* **2019**, *45*, 227–245. [[CrossRef](#)]
27. Bolón-Canedo, V.; Sánchez-Marroño, N.; Alonso-Betanzos, A. An ensemble of filters and classifiers for microarray data classification. *Pattern Recognit.* **2012**, *45*, 531–539. [[CrossRef](#)]
28. Freund, Y.; Schapire, R.E. A decision-theoretic generalization of on-line learning and an application to boosting. In Proceedings of the European Conference on Computational Learning Theory, Barcelona, Spain, 13–15 March 1995; pp. 23–37.
29. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42. [[CrossRef](#)]
30. Park, J.; Kim, D.; Kim, W. Greedy Feature Selection with Iterative Hierarchical Binning. In Proceedings of the 2026 IEEE International Conference on Consumer Electronics (ICCE), Dubai, United Arab Emirates, 3–5 February 2026.
31. Guyon, I.; Li, J.; Mader, T.; Pletscher, P.A.; Schneider, G.; Uhr, M. Competitive baseline methods set new standards for the NIPS 2003 feature selection benchmark. *Pattern Recognit. Lett.* **2007**, *28*, 1438–1444. [[CrossRef](#)]
32. Haslinger, C.; Schweifer, N.; Stilgenbauer, S.; Dohner, H.; Lichter, P.; Kraut, N.; Stratowa, C.; Abseher, R. Microarray gene expression profiling of B-cell chronic lymphocytic leukemia subgroups defined by genomic aberrations and VH mutation status. *J. Clin. Oncol.* **2004**, *22*, 3937–3949. [[CrossRef](#)]
33. Bhattacharjee, A.; Richards, W.G.; Staunton, J.; Li, C.; Monti, S.; Vasa, P.; Ladd, C.; Beheshti, J.; Bueno, R.; Gillette, M. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 13790–13795. [[CrossRef](#)]
34. Stienstra, R.; Saudale, F.; Duval, C.; Keshtkar, S.; Groener, J.E.; van Rooijen, N.; Staels, B.; Kersten, S.; Müller, M. Kupffer cells promote hepatic steatosis via interleukin-1 β -dependent suppression of peroxisome proliferator-activated receptor α activity. *Hepatology* **2010**, *51*, 511–522. [[CrossRef](#)]
35. Alon, U.; Barkai, N.; Notterman, D.A.; Gish, K.; Ybarra, S.; Mack, D.; Levine, A.J. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 6745–6750. [[CrossRef](#)]
36. Freije, W.A.; Castro-Vargas, F.E.; Fang, Z.; Horvath, S.; Cloughesy, T.; Liaw, L.M.; Mischel, P.S.; Nelson, S.F. Gene expression profiling of gliomas strongly predicts survival. *Cancer Res.* **2004**, *64*, 6503–6510. [[CrossRef](#)]
37. Singh, D.; Febbo, P.G.; Ross, K.; Jackson, D.G.; Manola, J.; Ladd, C.; Tamayo, P.; Renshaw, A.A.; D'Amico, A.V.; Richie, J.P. Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell* **2002**, *1*, 203–209. [[CrossRef](#)]

38. Guyon, I.; Gunn, S.; Ben-Hur, A.; Dror, G. Result analysis of the nips 2003 feature selection challenge. *Adv. Neural Inf. Process. Syst.* **2004**, *17*, 545–552.
39. Fanty, M.; Cole, R. Spoken letter recognition. *Adv. Neural Inf. Process. Syst.* **1990**, *3*, 220–226.
40. Bradley, P.S.; Mangasarian, O.L. Feature selection via concave minimization and support vector machines. *ICML* **1998**, *98*, 82–90.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.