



Section 2. Computer science

DOI:10.29013/AJT-26-1.2-55-62



A STABILITY-AWARE HYBRID FEATURE SELECTION ALGORITHM BASED ON FILTERED SPARSITY AND WRAPPER REFINEMENT

*Rashidov Khusan Shirinboyevich*¹

¹ Digital Technologies and Artificial Intelligence Development Research Institute

Cite: *Rashidov Kh. Sh. (2026). A Stability-Aware Hybrid Feature Selection Algorithm Based on Filtered Sparsity and Wrapper Refinement. Austrian Journal of Technical and Natural Sciences 2026, No 1 – 2. <https://doi.org/10.29013/AJT-26-1.2-55-62>*

Abstract

The process of choosing relevant features is essential in supervised machine learning, especially when dealing with tabular data of medium to high dimensions where unnecessary, uninformative, and interrelated predictors can negatively impact how well a model generalizes, how easy it is to understand, and how quickly it runs. While well-established techniques – namely filter, wrapper, and embedded methods – have been extensively explored, every category possesses inherent drawbacks concerning unpredictability, high computational cost, or an inability to fully capture dependencies among features. This research introduces the Filtered Sparse Stability Wrapper (FSSW), a novel hybrid feature selection technique. This framework emphasizes stability by combining statistical filtering of relevance, embedded regularization that promotes sparsity, and wrapper-style iterative refinement, all integrated within a bootstrap aggregation procedure. The FSSW hybrid selection method is framed as an optimization challenge involving multiple criteria, aiming to achieve an equilibrium between prediction precision, the conciseness of the selected feature set, and the consistency of the selection outcome. The theoretical underpinning relies upon principles of reducing the combinatorial search space and stability-aware optimization. Performance validation is carried out using established tabular datasets sourced from the UCI Machine Learning Repository, employing nested cross-validation to guarantee an unbiased assessment of efficacy. Experimental outcomes exhibit that the FSSW hybrid method consistently surpasses individual filter, embedded, and wrapper benchmark methodologies across metrics such as discriminant accuracy, F1-score performance, the extent of dimensionality reduction achieved, and stability as quantified by the Kuncheva index.

The conclusions assert that the FSSW hybrid feature selection approach delivers an effective, transparent, and practical methodology for addressing tabular classification tasks.

Keywords: *Filtered Sparse Stability Wrapper (FSSW); hybrid feature selection; stability-aware optimization; L1-regularization; recursive feature elimination; bootstrap aggregation; dimensionality reduction*

Introduction

The mounting complexity in the dimension of contemporary data has significantly elevated the necessity of effective feature selection within supervised learning contexts. Feature spaces with many dimensions often harbor predictors that are superfluous, unhelpful, or intertwined, which can subsequently inflate model variance, diminish how easily the model can be understood, and drive up the resources required for computation. The objective of feature selection is to pinpoint a reduced set of pertinent attributes while ensuring the initial meaning embedded in those attributes remains intact. In contrast to feature extraction techniques, which generate novel versions of the features, feature selection preserves the capacity for interpretation and aids in comprehension pertaining to the specific domain. Established feature selection strategies fall into three primary groupings. Filter techniques assess how relevant features are using statistical measures that are separate from any specific predictive models. Wrapper techniques gauge the worth of feature subsets based on how well the chosen model performs. Embedded techniques incorporate the selection process right into the model fitting phase by employing regularization or structural limitations. While every one of these frameworks has shown utility, no solitary methodology achieves peak performance across predictive precision, feature minimization (sparsity), and consistency (stability) all at once. Wrapper methods demand substantial computational effort, filter methods overlook how features might interact with each other, and embedded methods can become erratic when dealing with features that are highly correlated. Emerging research patterns suggest that composite feature selection structures offer enhanced resilience by merging mechanisms that compensate for one another's weaknesses. Nevertheless, many current composite designs fail to incorporate deliberate control over stability or a rigorous mathematical framework for optimization. This research introduces a feature selection algorithm that is both hybrid and explicitly formalized to consider stability, and we support our proposal with empirical evidence.

Materials and methods

Let the supervised dataset be defined as

$$D = \{(x_i, y_i)\}_{i=1}^n \quad (1)$$

where $x_i \in \mathbb{R}^d$ represents a d -dimensional feature vector and $y_i \in \{0,1\}$ denotes the binary class label.

The goal of feature selection is to determine an optimal subset

$$S^* \subseteq \{1, 2, \dots, d\} \quad (2)$$

that improves predictive performance while reducing dimensionality and maintaining selection stability under data perturbations.

The proposed FSSW hybrid feature selection method is formulated as a multi-criteria optimization problem:

$$S^* = \arg \max_S (\mathcal{A}(S) - \lambda_1 |S| + \lambda_2 \text{Stab}|S|) \quad (3)$$

where: $\mathcal{A}(S)$ denotes predictive accuracy, $|S|$ is the subset cardinality, $\text{Stab}|S|$ is a stability measure, λ_1, λ_2 are trade-off parameters.

Hybrid Feature Selection Procedure

The proposed FSSW method combines statistical filtering, embedded sparsity regularization, wrapper-based refinement, and bootstrap aggregation within a unified framework.

First, feature relevance is estimated using mutual information:

$$I(X_j; Y) = \sum_{x_j, y} p(x_j, y) \log \frac{p(x_j, y)}{p(x_j)p(y)} \quad (4)$$

Features are ranked according to their mutual information scores and the top- k_1 features are retained:

$$S_1 = \text{Top}_{k_1} (I(X_j; Y)) \quad (5)$$

This step substantially reduces combinatorial search complexity from 2^d to 2^{k_1} where $k_1 \ll d$. The filtering stage eliminates clearly irrelevant predictors before model-based optimization.

Next, embedded sparse reduction is performed using L1-regularized logistic regression applied to S_1 :

$$\hat{\beta} = \arg \min_{\beta} \left(\frac{1}{n} \sum_{i=1}^n \ell(y_i, x_i^T \beta) + \alpha \beta_1 \right) \quad (6)$$

where

$$\beta_1 = \sum_{j=1}^{k_1} |\beta_j| \quad (7)$$

The L1 penalty enforces sparsity by shrinking less informative coefficients to zero. The reduced subset becomes

$$S_2 = \{j \in S_1 : \widehat{\beta}_j \neq 0\} \quad (8)$$

This step removes redundant and correlated predictors while preserving discriminative information.

To account for feature interactions, wrapper refinement is applied to S_2 using Recursive Feature Elimination (RFE). At each iteration, the least informative feature is removed according to coefficient magnitude:

$$S^{(t+1)} = S^{(t)} \setminus \arg \min_{j \in S^{(t)}} |\beta_j| \quad (9)$$

The optimal subset is selected using cross-validation:

$$S_3 = \arg \max_{S \subseteq S_2} \mathcal{A}_{CV}(S) \quad (10)$$

Finally, stability aggregation is introduced to enhance reproducibility. Bootstrap resampling is performed B times:

$$D^{(b)} \sim \text{Bootstrap}(D) \quad (11)$$

yielding subsets $S^{(b)}$. Feature frequency is computed as

$$f_j = \frac{1}{B} \sum_{b=1}^B \mathbf{1}(j \in S^{(b)}) \quad (12)$$

The final subset is defined as

$$S^* = \{j : f_j \geq \tau\} \quad (13)$$

where τ is a stability threshold.

This aggregation mechanism reduces sensitivity to sampling variability and improves subset robustness.

Computational Complexity Analysis

The computational efficiency of the proposed FSSW hybrid feature selection algorithm is analyzed by decomposing the procedure into its sequential stages. Let n denote the number of samples, d the original number of features, k_1 the number of features retained after filtering ($k_1 \ll d$), k_2 the number of features after embedded sparsity reduction ($k_2 \ll k_1$), T the number of wrapper iterations, and B the number of bootstrap repetitions.

The asymptotic complexity of each stage is summarized in Table 4.

Table 4. Computational complexity of the proposed FSSW algorithm

Stage	Description	Time Complexity
Stage 1	Mutual Information computation for all features	$O(nd)$
Stage 2	L1-regularized Logistic Regression	$O(nk_1)$
Stage 3	Recursive Feature Elimination (RFE)	$O(T \cdot nk_2)$
Stage 4	Bootstrap stability aggregation	$O(B \cdot (nd + nk_1 + T \cdot nk_2))$

The first stage scales linearly with both the number of samples and the original dimensionality. Since filtering is applied independently per feature, it remains computationally efficient even for moderate d .

The embedded sparse reduction stage operates on the reduced subset k_1 , yielding complexity $O(nk_1)$. Because $k_1 \ll d$, this represents a substantial reduction compared to performing embedded selection on the full feature space.

The wrapper refinement stage has complexity $O(T \cdot nk_2)$, where $k_2 \leq k_1$. In contrast, wrapper-only selection would require

$O(T \cdot nd)$, which becomes considerably larger when d is high.

Bootstrap aggregation multiplies the total cost by B , but this step is fully parallelizable and does not alter the asymptotic scaling behavior.

Since $k_1 \ll d$ and $k_2 \leq k_1$, the proposed hybrid framework significantly reduces the effective search space before wrapper optimization. Consequently, the method achieves improved computational efficiency compared to wrapper-only approaches while preserving model-driven refinement capability.

Datasets and Data Characteristics

We assessed the suggested hybrid feature selection technique, named FSSW, utilizing two well-established benchmark collections sourced from the UCI Machine Learning Repository. Both collections pertain to binary classification tasks and are apt for scrutinizing feature selection algorithms due to their mid-range feature count, presence of interrelated predictors, and lifelike levels of inherent noise.

The Breast Cancer Wisconsin (Diagnostic) collection comprises 569 observations, each described by 30 numerical attributes derived from magnified scans of breast mass aspirates. The variable to be predicted signifies whether the mass is malignant or benign, yielding two categories. A notable aspect of this collection is the presence of highly related descriptive metrics, leading to feature duplication, yet the features themselves are relatively free of error, making it a good candidate for assessing both the sparsity and the consistency of the chosen feature set.

The Ionosphere collection holds 351 observations characterized by 34 continuous variables obtained from radar signals. The goal is to categorize the signal as either “good” or “bad.” This collection is recognized for presenting more pronounced interdependencies among features and greater signal corruption than the Breast Cancer data, thus posing a tougher scenario for achieving a robust selection of optimal feature subsets. Its moderate sample volume coupled with significant feature correlations confirms its suitability for gauging the advantages of iterative refinement and combining results from wrapper-based methods in a hybrid approach.

No missing entries exist in the canonical UCI versions of either dataset. Prior to the model fitting phase, all attributes underwent standardization, with the scaling factors derived solely from the training portions of the data.

Classification Model Used for Evaluation

To ensure that improvements originate from feature selection rather than overly complex predictors, a relatively interpretable and well-established classifier was employed. The evaluation model was **Logistic Regres-**

sion (LR), selected for its suitability in tabular binary classification and its compatibility with both embedded sparsity selection and wrapper refinement.

The classification decision rule is given by:

$$\hat{y} = \mathbb{I}\left(\sigma\left(x^T \beta\right) \geq 0.5\right), \quad \sigma(z) = \frac{1}{1 + e^{-z}} \quad (14)$$

For embedded selection, the LR model was trained with an L1 penalty, enabling coefficient shrinkage and sparsity. For wrapper refinement (RFE), LR was also used as the base estimator because its coefficient magnitudes provide a stable ranking signal for elimination.

Hyperparameters (regularization strength C for LR, and the number of features retained at each stage) were tuned in the inner loop of nested cross-validation to prevent leakage.

Evaluation Metrics and How They Were Computed

To provide a comprehensive evaluation, predictive performance, subset compactness, and stability were assessed. Three primary metrics were used: Accuracy, F1-score, and the Kuncheva stability index. Each metric is described below.

Accuracy. Accuracy measures the proportion of correctly classified instances:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (15)$$

where TP is true positives, TN is true negatives, FP is false positives, and FN is false negatives.

Accuracy is informative when class distribution is not extremely imbalanced. In our experiments, it provides a direct measure of classification quality under reduced feature subsets.

F1-score. To account for possible asymmetries between precision and recall, the F1-score was computed:

$$Precision = \frac{TP}{TP + FP}, \quad (16)$$

$$Recall = \frac{TP}{TP + FN}, \quad (17)$$

$$F1\text{-score} = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (18)$$

F1-score is particularly useful in scenarios where misclassification costs are asymmetric or where the positive class requires stronger detection sensitivity.

Kuncheva Stability Index. Feature selection stability quantifies the consistency of selected subsets under data perturbations. The Kuncheva index is a well-established measure for comparing stability across repeated selections.

Given two selected subsets S_i and S_j , each of size k , drawn from a total pool of d features, the pairwise Kuncheva stability is defined as:

$$\kappa(S_i, S_j) = \frac{|S_i \cap S_j| - \frac{k^2}{d}}{k - \frac{k^2}{d}} \quad (19)$$

The final stability score is computed as the average of the pairwise stability across all subset pairs obtained from bootstrap iterations:

$$\kappa = \frac{2}{B(B-1)} \sum_{1 \leq i < j \leq B} \kappa(S_i, S_j) \quad (20)$$

where B denotes the number of bootstrap repetitions.

A higher Kuncheva index indicates greater reproducibility and robustness of the feature selector, which is a critical requirement for scientific and industrial applications.

Results

The experimental evaluation of the proposed Filtered Sparse Stability Wrapper (FSSW) algorithm was conducted on the Breast Cancer Wisconsin Diagnostic dataset (30 features) and the Ionosphere dataset (34 features). Performance was assessed using nested cross-validation to ensure unbiased estimation.

Nested 5-fold cross-validation was used.

Predictive Performance

Table 1 presents classification Accuracy and F1-score for all compared methods.

Table 1. Predictive performance comparison

Dataset	Method	Accuracy	F1-score
Breast Cancer	Filter-only	0.957	0.956
Breast Cancer	Embedded (L1-LR)	0.963	0.962
Breast Cancer	Wrapper-only (RFE-LR)	0.966	0.965
Breast Cancer	Proposed FSSW	0.971	0.970
Ionosphere	Filter-only	0.889	0.887
Ionosphere	Embedded (L1-LR)	0.903	0.901
Ionosphere	Wrapper-only (RFE-LR)	0.908	0.906
Ionosphere	Proposed FSSW	0.918	0.915

For the Breast Cancer dataset, the proposed hybrid method achieved an accuracy improvement of +1.4% over the filter-only baseline and +0.8% over the embedded-only method. While wrapper-only selection already improves performance compared to filter-only, the addition of filtering and stability aggregation in FSSW further enhances predictive quality.

On the Ionosphere dataset, which is known to exhibit stronger feature interactions and noise, the hybrid method achieved

the largest performance gain, improving accuracy by +2.9% compared to filter-only and +1.5% compared to embedded selection. This suggests that the wrapper refinement stage effectively captures nonlinear dependencies after dimensionality reduction.

Overall, the proposed FSSW method consistently achieved the highest predictive performance across both datasets.

Dimensionality Reduction

Table 2 reports the dimensionality reduction achieved by the proposed method.

Table 2. Dimensionality reduction

Dataset	Original	Proposed
Breast Cancer	30	7
Ionosphere	34	9

The proposed hybrid algorithm reduced the feature space by approximately: 76.7% in Breast Cancer (30 → 7), 73.5% in Ionosphere (34 → 9). Importantly, this reduction was achieved without sacrificing predictive performance. In fact, accuracy increased while dimensionality decreased substantially. This demonstrates that the hybrid strategy effec-

tively eliminates redundant and irrelevant features while preserving discriminative information.

Stability Analysis

Table 3 presents the Kuncheva stability index (κ) for each method.

Table 3. Stability comparison (Kuncheva index κ)

Dataset	Filter-only	Embedded (L1-LR)	Wrapper-only	Proposed FSSW
Breast Cancer	0.74	0.69	0.63	0.81
Ionosphere	0.71	0.67	0.60	0.78

The wrapper-only approach shows the lowest stability values in both datasets. This behavior is expected, as wrapper methods are sensitive to sample perturbations and local search variability.

Embedded L1 selection demonstrates moderate stability but remains sensitive to correlated predictors. Filter-only selection achieves slightly higher stability than embedded methods; however, its predictive performance is lower. The proposed FSSW algorithm achieves the highest stability in both datasets. Compared to wrapper-only selection, stability increases by: +0.18 (Breast Cancer), +0.18 (Ionosphere).

This confirms that bootstrap aggregation significantly improves subset reproducibility.

Discussion

The experimental outcomes show that the suggested hybrid structure offers an improved equilibrium across predictive precision, dimension lessening, and robustness. From a predictive viewpoint, the efficacy improvements noted in both datasets suggest that merging filtering, inherent sparsity, and wrapper refinement successfully seizes mutual facets of feature importance. Filtering eliminates clearly extraneous predictors, embedded L1 regularization lessens overlap among linked variables, and wrap-

per tuning maximizes model-specific interplay. The dimension reduction figures are notably important. Decreasing the feature group by over 70% whilst concurrently enhancing precision implies that the original collections possess considerable needless information. The hybrid design permits gradual compaction without efficacy decline. Stability review uncovers one principal merit of the suggested method. Wrapper-only techniques, though potent, are inconsistent upon resampling. Through incorporating bootstrap aggregation, the FSSW process lessens selection fluctuation and promotes replicability. This is particularly vital in scientific and commercial uses where clarity and resilience are paramount. The Ionosphere set, which displays more intricate feature relationships, profits more from wrapper tuning than the Breast Cancer set. This indicates that hybridization becomes more advantageous as feature links and errors escalate. In sum, the findings verify that organized hybrid feature selection attains multi-standard enhancement: greater accuracy, leaner subsets, and firmer stability versus independent strategies. Subsequent efforts might look into broadening the structure to multi-class issues, incorporating cost-aware aims, or examining scalability in high-dimension omics data.

Conclusions

This study proposed a stability-aware hybrid feature selection algorithm, namely Filtered Sparse Stability Wrapper (FSSW), designed to simultaneously optimize predictive accuracy, dimensionality reduction, and subset stability in tabular classification tasks. The proposed framework integrates three complementary mechanisms: statistical filtering based on mutual information, embedded sparsity reduction via L1-regularized logistic regression, and wrapper-based recursive feature elimination, followed by bootstrap stability aggregation. The theoretical motivation of the method is grounded in combinatorial search reduction and stability-aware optimization principles. By progressively constraining the feature space before performing model-driven refinement, the algorithm reduces computational complexity while preserving the ability to model feature interactions. The inclusion of bootstrap aggregation further enhances reproducibility and mitigates instability commonly observed in wrapper-based approaches. Empirical validation on the Breast Cancer and Ionosphere benchmark datasets demonstrated that the proposed hybrid approach consistently outperforms standalone filter, embedded, and wrapper baselines. The method achieved higher predictive accuracy and F1-score while reducing dimensionality

by more than 70% in both datasets. Moreover, the Kuncheva stability index confirmed improved robustness of selected feature subsets under resampling.

The results indicate that structured hybridization provides a practical and effective solution for moderate-dimensional tabular datasets where redundancy and correlation are present. The integration of filtering, sparsity regularization, wrapper refinement, and stability aggregation enables multi-criteria optimization without excessive computational overhead.

Future research directions include extending the proposed framework to multi-class and multi-label settings, incorporating cost-sensitive objectives for feature acquisition constraints, and evaluating scalability in high-dimensional domains such as genomics and multi-omics data. Additionally, integrating multi-objective optimization strategies that explicitly balance accuracy, sparsity, and stability within a unified mathematical framework represents a promising avenue for further development.

In summary, the proposed FSSW algorithm demonstrates that stability-aware hybrid feature selection is a robust and computationally feasible paradigm for improving generalization performance and interpretability in supervised learning tasks.

References

- Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, – 3. – P. 1157–1182.
- Pirgazi, J., Alimoradi, M., Esmaeili Abharian, T., & Olyae, M. H. (2019). An efficient hybrid filter-wrapper metaheuristic-based gene selection method for high-dimensional datasets. *Scientific Reports*, – 9(1). – 18580 p. URL: <https://doi.org/10.1038/s41598-019-54866-2>
- Nogueira, S., Sechidis, K., & Brown, G. (2018). On the stability of feature selection algorithms. *Journal of Machine Learning Research*, – 18(174). – P. 1–54.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, – 58(1). – P. 267–288.
- Meinshausen, N., & Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, – 72(4). – P. 417–473. URL: <https://doi.org/10.1111/j.1467-9868.2010.00740.x>
- Kuncheva, L.I. (2007, February). A stability index for feature selection. In *Artificial intelligence and applications* (pp. 421–427).
- Brown, G., Pocock, A., Zhao, M. J., & Luján, M. (2012). Conditional likelihood maximisation: a unifying framework for information theoretic feature selection. *The journal of machine learning research*, – 13. – P. 27–66.
- Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine learning*, – 46(1). – P. 389–422.

Asuncion, A., & Newman, D. (2007). UCI machine learning repository. University of California, Irvine. URL: <http://archive.ics.uci.edu/ml>

submitted 12.02.2026;
accepted for publication 26.02.2026;
published 27.02.2026
© Rashidov Kh. Sh.
Contact: rashidovhusan125@gmail.com