

Comparative Evaluation of Filter and Wrapper Feature Selection Methods in Species Distribution Modeling

Yousra Cherif and Ali Idri(✉)

Software Project Management Research Team, ENSIAS, Mohammed V University in Rabat, Rabat, Morocco
ali.idri@um5.ac.ma

Abstract. Species Distribution Models (SDMs) have been extensively used as a tool for exploring the relationship between species occurrence and environmental factors. SDMs can be used for estimating the habitats where species can potentially exist and predicting their geographic distributions. In SDMs, feature selection (FS) is a crucial step in improving the accuracy of models and their interpretability by reducing the dimensionality of the feature space by discarding the least important features. In this study, SDMs were used to model the distributions of seven bird species, and a comparative analysis of FS techniques was conducted to identify the best approach. The FS methods used in this study include six different filter methods with two different thresholds each (40% and 50%). In addition, six wrapper techniques and two different multivariate methods have been used. The performance of the FS methods has been compared using four different classification methods: Extreme Gradient Boosting (XGB), Random Forest (RF), Decision Tree (DT), and Light Gradient Boosting Machine (LGBM). The results have shown that the Sequential Feature Selection (SFS) method has the highest performance. However, when the number of features is taken into consideration, it appears that Fisher Score with Random Forest is the best approach. Fisher Score with Random Forest has the highest performance with only nine features.

Keywords: Species distribution models · Feature selection · Environmental data · Machine learning · Classification

1 Introduction

Given the importance of SDMs in conservation practice, maintaining their accuracy and reliability is crucial. Feature Selection (FS) is commonly applied to improve model performance by selecting the most relevant predictors while removing redundant or irrelevant variables [6–8,13]. In addition to enhancing predictive accuracy [3], FS helps reduce data dimensionality, lower computational complexity, and increase model interpretability [14,19].

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2027
O. Gervasi et al. (Eds.): ICCSA 2026, LNCS 16767, pp. 628–643, 2027.
https://doi.org/10.1007/978-3-032-30491-9_39

FS techniques are generally grouped into three main categories: filters, wrappers, and embedded approaches [15]. Filter methods rely on statistical relationships between features and the target variable, wrapper methods evaluate subsets of features using a predictive model, and embedded methods integrate selection directly into the training process. More recently, hybrid strategies have been proposed to combine the strengths of these approaches while reducing their individual limitations [24]. Filter techniques can further be divided into univariate methods, which rank features individually, and multivariate methods, which consider interactions among features.

For instance, the study [25] compared three filter methods with two embedded approaches using several selection thresholds on seven UCI Machine Learning Repository and FS datasets [17,18]. Their findings showed modest performance gains when ensemble rank aggregation was applied. Similarly, the study [9] evaluated multiple filters, wrappers, embedded, and ensemble strategies on environmental datasets using Random Forest and LightGBM classifiers. Shapley Additive Explanations and Permutation Importance were identified as strong individual methods, while Reciprocal Ranking performed best among ensemble approaches. Other works, such as the study [5], further emphasize the advantages of integrating filter and wrapper techniques, especially in medical applications. Nevertheless, significant gaps persist. Many studies depend on generic benchmark datasets instead of real-world SDM scenarios, which reduces their ecological relevance. Moreover, the interaction between FS methods and different classifiers, especially tree-based models commonly used in SDMs, is still underexplored. There is also a lack of systematic comparisons between univariate, multivariate, and wrapper approaches. These limitations motivate the present study, which aims to provide a more comprehensive and ecologically grounded evaluation of feature selection methods.

To the best of our knowledge, this paper is the first to concentrate on comparing univariate and multivariate filters, along with wrapper techniques. Furthermore, this study utilizes the Scott Knott (SK) statistical analysis, which has shown superior effectiveness compared to other statistical methods [2,6,8,16,29,30], such as those conducted by Calinski and Corsten [4] and Cox and Spjotvoll [28].

In this current paper, we have expanded our analysis by using: (1) six univariate filters (Linear Correlation, Mutual Information, Fisher Score, ANOVA F-value, minimum Redundancy Maximum Relevance, and t-test) with 40% and 50% thresholds as well as (2) two multivariate filters (Consistency-based subset selection and Correlation-based subset selection) and (3) six wrapper methods (Boruta, Permutation Importance, Shapley Additive Explanations, Genetic Algorithm, Recursive Feature Elimination with Cross-Validation, and Sequential Feature Selector), (4) seven bird species datasets (*P.moussieri, P.phoenicurus, P.ochruros, O.leucopyga, O.deserti, O.oenanthe, and O.leucura*) using the SK clustering test and Borda Count (BC) method to assess and contrast how FS affects classifier performance. In total, the current study evaluates 588 variants of

test set-ups: (4 classifiers) * (6 univariate filters * 2 thresholds + 2 multivariate filters + 6 wrapper methods + original feature set) * (7 datasets).

The paper is structured as follows: Sect. 2 describes the study area, detailing the species occurrence data, environmental variables, and the methodological steps followed to conduct the empirical analyses. Section 3 presents and discusses the main findings. Section 4 addresses the potential threats to validity. Finally, Sect. 5 concludes the paper by summarizing the key insights and contributions of the study.

2 Material and Methods

This study was conducted in Morocco using occurrence data from bird species belonging to the genera *Phoenicurus* and *Oenanthe*. Presence-only records were collected and complemented with pseudo-absence data to enable SDM. These datasets were then used to examine spatial patterns and build predictive models.

2.1 Study Area

Morocco, located in northwestern Africa, covers approximately 710,850 km^2. It borders Algeria and Mauritania and is bounded by the Mediterranean Sea to the north and the Atlantic Ocean to the west. The country presents substantial geographical diversity, including coastal areas, the Rif, Middle Atlas, High Atlas, and Anti-Atlas mountain ranges, as well as Saharan regions.

Morocco's climate is influenced by both maritime and desert conditions, resulting in marked spatial and seasonal variability. The average annual precipitation is approximately 318.8 mm, mainly concentrated between October and April, while summers are generally dry. Mean temperatures range from 9.4°C to 26°C.

2.2 Species Occurrence Datasets

The dataset includes occurrence records from three species of the genus *Phoenicurus* and four species of the genus *Oenanthe*. The *Phoenicurus* genus belongs to the Muscicapidae family (Redstarts), and the study focused on *Phoenicurus moussieri*, *Phoenicurus ochruros*, and *Phoenicurus phoenicurus*.

Since true absence data are difficult to obtain, pseudo-absence data were generated following the recommendations of [21]. Such pseudo-absences inherently involve uncertainty; therefore, absence points should not be sampled in ecologically suitable areas or too close to presence records [20]. In this study, pseudo-absence locations were randomly selected within a 250 km radius around presence points, as described in [21] and supported by previous studies such as [27], and applied in works including [1,26]. Buffer zones were further implemented around presence records to reduce environmental similarity and spatial bias. The presence-only dataset for *Phoenicurus* species contains 10,993 valid records [12], obtained after removing four records with missing values from the initial 10,997

entries. The wheatear dataset includes 1,750 records [11], all retrieved from the GBIF global database. Data preprocessing involved removing duplicates, handling missing values, and eliminating outliers to prevent degradation of model performance [22]. The wheatear records were primarily sourced from four repositories, with the EOD – eBird Observation Dataset contributing 91.2% of the data (1,597 records from ebird.org) and the Western Palearctic migratory bird dataset providing 4.4% (78 records from https://macroecology.ku.dk/resources/african_migrants). The observations span from 1965 to 2005.

All species names were validated according to ITIS standards and are associated with permanent Taxonomic Serial Numbers (TSNs) within the GBIF database (https://www.gbif.org/), as shown in Table 1. Ecologically, *Phoenicurus moussieri* is typically found in mountainous and rocky habitats with sparse vegetation. *Phoenicurus ochruros* prefers rocky natural and urban environments for nesting, while *Phoenicurus phoenicurus* is commonly associated with deciduous and mixed forests, forest edges, and open woodland areas. Table 2 summarizes the occurrence records.

Table 1. Description of the four wheatear species in Morocco

Sample Image	Scientific Name	Observations	TSN
	Oenanthe Deserti	400	561647
	Oenanthe Leucopyga	473	561652
	Oenanthe Leucura	469	561653
	Oenanthe Oenanthe	408	179814

Table 2. Description of the three redstart birds in Morocco

Sample Image	Scientific Name	Common Name	Total observations
	Phoenicurus Moussieri	Moussier's Redstart	5223
	Phoenicurus Ochruros	Black Redstart	3364
	Phoenicurus Phoenicurus	Common Redstart	2406

2.3 Environmental Data

Species distributions are strongly shaped by geographic and climatic conditions. Given that the persistence of the seven studied bird species is closely linked to environmental factors, climatic variables were selected as the primary predictors

in the distribution models. These variables enable a detailed representation of environmental conditions and support more reliable habitat suitability predictions.

A total of 19 bioclimatic variables were obtained from the WorldClim global database [10]. These variables describe average climatic conditions for the period 1970–2000. A spatial resolution of 2.5 arc-minutes (approximately 5 km) was adopted, as it provides a suitable compromise between ecological relevance and computational efficiency at the regional scale. This resolution is also consistent with the spatial uncertainty commonly associated with species occurrence records, particularly those derived from observational sources such as GBIF. Using a finer resolution could introduce an artificial level of precision not supported by the occurrence data. In addition, elevation data from the SRTM digital elevation database were incorporated [23]. In total, 20 environmental predictors were used to model the distribution of the seven bird species.

For the four wheatear species, land cover information was also included. Data were derived from NASA's MODIS Land Cover Type Product (MCD12Q1 v6), specifically the LC_type4 (LC4) classification scheme. This dataset categorizes land cover into classes such as broadleaf evergreen forests, non-vegetated areas, and urban zones. Integrating land cover variables allowed the models to better account for habitat characteristics that may influence species distribution patterns.

2.4 Experimental Process

This section outlines the experimental protocol adopted to evaluate the proposed methodology.

To investigate the effectiveness of the FS techniques, we applied a combination of univariate, multivariate, and wrapper approaches. These methods generated multiple reduced feature subsets based on different selection criteria. The impact of each subset was then assessed using four classification algorithms. All resulting variants, including those obtained from FS techniques and the complete feature set, were systematically evaluated across the seven datasets. The evaluation procedure followed an eight-step pipeline, summarized below.

Step 1: For each dataset, feature subsets were generated using six univariate ranking methods with two selection thresholds (40% and 50%), two multivariate methods, and six wrapper approaches. This resulted in 20 variants per dataset, yielding a total of 140 dataset variants (7 datasets $\times$ 20 variants).

Step 2: Four classifiers: DT, LGBM, XGB, and RF were trained on each feature subset generated in Step 1, as well as on the original feature set. A 5-fold cross-validation strategy was employed to compute accuracy, kappa, and f1-score. Overall, 147 model variants were produced per classifier across all datasets ((6 univariate filters $\times$ 2 thresholds + 2 multivariate filters + 6 wrappers + 1 original set) $\times$ 7 datasets).

Step 3: The performance of filter and wrapper methods was compared using the SK test and the BC voting approach to determine statistically significant differences and overall rankings among methods.

3 Results

This section reports and analyzes the performance of the FS techniques in combination with the four classifiers across the seven datasets. To structure the comparison, we applied the SK test based on accuracy and the BC voting method, which integrates Accuracy, Kappa, and F1-score, to rank classifiers within the top SK cluster.

Accordingly, the evaluation focuses exclusively on these three metrics, selected to provide a comprehensive and balanced assessment of model performance in our study context. In particular, we compare and contrast the effectiveness of filter and wrapper techniques using Accuracy, Kappa, and F1-score (RQ1).

– **Evaluation of filters (univariate and multivariate) and wrappers (RQ1)**

In this subsection, we compared the performance of all previously described univariate filters (using 40% and 50% thresholds), multivariate filters, and wrapper methods. Each selected feature subset, as well as the full feature set, was used to train four classifiers. Classifier performance was evaluated using the SK test based on accuracy, and classifiers within the top SK cluster were ranked using the BC voting method.

Figures 2, 3, 4, 5, 6 and 7 and 8 in the Appendix show the SK test findings for univariate and multivariate filters, while Table 3 summarizes the proportion of wrapper and filter methods in the top SK cluster for each dataset.

Table 3. Proportion of wrapper and filter methods in the top SK cluster for each dataset

Dataset	Wrapper (%)	Filter (%)	Interpretation
P.moussieri	54.17 (13/24)	36.67 (22/60)	Wrapper methods perform better than filters
P.ochruros	33.33 (8/24)	36.67 (22/60)	Wrapper and filter methods perform similarly
P.phoenicurus	50 (12/24)	25 (15/60)	Wrapper methods are preferable
O.deserti	75 (18/24)	75 (45/60)	Wrapper and filter methods perform equally
O.leucopyga	75 (18/24)	68.33 (41/60)	Wrapper methods slightly outperform filters
O.leucura	75 (18/24)	33.33 (20/60)	Wrapper methods are preferable
O.oenanthe	58.33 (14/24)	31.67 (19/60)	Wrapper methods perform better

To evaluate how filter and wrapper methods perform overall, we counted how often each feature selection technique appeared in the best SK clusters and then applied the BC using the top 10 ranks across the seven datasets (Table 5). Filter methods can appear up to 56 times per dataset [(6 univariate × 2 thresholds + 2 multivariate) × 4 classifiers] [(6 univariate ×2 thresholds + 2 multivariate)×4 classifiers], and wrapper methods up to 24 times (6 wrappers × 4 classifiers) (6

wrappers × 4 classifiers), resulting in maximum totals of 392 and 168 occurrences, respectively, across all datasets. As shown in Table 4, Boruta obtained the highest overall ranking, followed by the remaining techniques in the order reported in the table. Thus, when the feature-set size is not included in the evaluation criteria, wrapper techniques emerge as the strongest performers.

Table 4. Occurrence rates and BC rankings of the FS methods appearing in the top clusters over the seven datasets

FS technique	Occurrence rates	BC Ranking
Boruta	19 times out of 28 (67.86%)	3rd, 5th, 8th Ranks: 1x 6th Rank: 2x
CON	18 times out of 28 (64.29%)	5th Rank: 4x 6th Rank: 1x
GA	17 times out of 28 (60.71%)	1st, 8th Ranks: 2x 4th, 7th, 9th Ranks: 1x
RFE	17 times out of 28 (60.71%)	2nd, 4th, 10th Ranks: 1x
SFS	16 times out of 28 (57.14%)	1st Rank: 4x 2nd Rank: 2x 6th, 9th Ranks: 1x
PI	16 times out of 28 (57.14%)	1st, 2nd, 5th, 7th Ranks: 1x 3rd, 4th Ranks: 2x
Shap	16 times out of 28 (57.14%)	4th, 5th, 9th, 10th Ranks: 1x 7th Rank: 2x
F	26 times out of 56 (46.43%)	2nd, 6th, 7th, 8th, 9th Ranks: 1x
M	25 times out of 56 (44.64%)	8th, 9th Ranks: 1x 10 Ranks: 2x
CFS	12 times out of 28 (42.86%)	3rd, 6th Rank: 1x
MR	22 times out of 56 (39.29%)	9th Rank: 1x
TT	21 times out of 56 (37.5%)	2nd, 3rd, 10th Ranks: 1x
A	21 times out of 56 (37.5%)	3rd, 7th, 9th, 10th Ranks: 1x
C	21 times out of 56 (37.5%)	8th Rank: 1x

Moreover, to rank the FS techniques, we applied the BC to identify the top 10 methods within the best SK cluster. The best classifierâĂŞfilter (or wrapper) combination is the one that attains the top BC ranking while selecting the smallest number of features.

Table 5 reports the top-ten ranked techniques and their corresponding feature counts; for example, RFORG(20) indicates that RFORG was trained with 20 features, while RFSFS(10) used 10 features. Since all techniques in the best SK cluster show equivalent accuracy, those highlighted in bold represent the most efficient options (highest Borda rank with fewer features).

Overall, these findings highlight the importance of feature selection in improving classifier performance. The full feature set ranked second or third only once and appeared in lower ranks (fourth, sixth, and seventh) when RF was trained with 20âĂŞ21 features, whereas both filters and wrappers achieved comparable performance with far fewer features. When considering feature dimensionality, SFS and Fisher Score showed a clear advantage, consistently delivering strong performance with reduced feature sets.

Table 5. BC top ten rankings of filter and wrapper approaches, including the corresponding feature set size in parentheses

P.mous	P.och	P.pho	O.deserti	O.leucopyga	O.leucura	O.oenanthe
RFSFS(10)	**RFSFS(10)**	**RFSFS(10)**	**RFPI(10)**	RFGA(15)	**RFSFS(10)**	RFGA(13)
RFPI(10)	RFF50(10)	RFRFE(17)	RFSFS(10)	RFTT40(8)	RFORG(21)	**RFSFS(10)**
RFORG(20)	RFCFS(9)	RFPI(10)	**XGBTT50(11)**	RFA50(11)	RFPI(10)	RFBoruta(21)
RFShap(10)	RFORG(20)	RFGA(14)	RFORG(21)	**XGBPI(10)**	RFRFE(21)	RFPI(10)
RFCON(13)	RFShap(10)	RFBoruta(19)	RFCON(16)	RFPI(10)	RFCON(13)	RFCON(18)
XGBSFS(10)	**RFF40(8)**	RFCON(17)	RFBoruta(20)	**RFCFS(6)**	RFBoruta(21)	RFORG(21)
RFA40(8)	RFPI(10)	RFORG(20)	RFShap(10)	RFShap(10)	RFGA(10)	**RFF40(8)**
RFGA(7)	RFM50(10)	XGBORG(20)	RFC50(11)	RFBoruta(20)	**RFF50(11)**	XGBGA(13)
RFM40(8)	RFMR50(10)	XGBGA(14)	RFA50(11)	RFSFS(10)	XGBShap(10)	RFF50(11)
RFM50(10)	RFM40(8)	**RFM40(8)**	RFTT50(11)	RFRFE(11)	RFShap(10)	RFA50(11)

In this subsection, we computed the median accuracy, Kappa, and F1-score across all datasets to determine the most effective FS techniques. For **RF**, the top accuracy was reached using RFSFS (91.79%); for **XGB**, XGBShap (91.54%) and XGBSFS (91.52%); for **LGBM**, LGBMSFS (91.28%); and for **DT**, DTORG and DTBoruta (89.43%) (Table 6). Figure 1 shows classifier performance grouped into 65 SK clusters: techniques sharing a color are statistically indistinguishable, while different colors indicate significant differences. The top clusters for RF contained RFSFS, RFORG/RFPI, and RFRFE, trained with 10, 21, 10, and 14 features, respectively, indicating that RF performs best with SFS. Overall classifier ranking is RF $>$ XGB $>$ LGBM $>$ DT [8]. Considering feature count, XGBShap, LGBMSFS, and DTBoruta were the best combinations.

Without accounting for feature numbers, Boruta ranked highest, followed by Consistency-based subset selection, Genetic Algorithm, RFE with CV, SFS, Permutation Importance, Shap, Fisher Score, Mutual Information, Correlation-based subset selection, mRMR, t-test, ANOVA F-value, and Linear Correlation. When feature count is considered, RF combined with SFS is preferred, as it uses only 10 features and appears in the top SK cluster (see Fig. 1).

Table 6. Median results for the three metrics, computed for the filters over all seven datasets

Classifier	FS technique	Accuracy(%)	F1_score(%)	Kappa(%)
RF	M40	91.53	91.53	83.06
	M50	91.3	91.3	82.6
	C40	91.25	91.25	82.5
	C50	91.36	91.36	82.72
	A40	91.41	91.41	82.82
	A50	91.38	91.38	82.77
	F40	91.36	91.36	82.72
	F50	91.58	91.58	83.16
	TT40	90.9	90.9	81.82
	TT50	91.25	91.25	82.51
	MR40	91.16	91.16	82.33
	MR50	91.32	91.32	82.64
	ORG	91.75	91.75	83.51
	CON	91.63	91.63	83.26
	CFS	91.41	91.41	82.82
	RFE	91.7	91.7	83.41
	PI	91.75	91.75	83.5
	Boruta	91.63	91.63	83.26
	Shap	91.53	91.53	83.06
	SFS	**91.79**	**91.79**	**83.58**
XGB	M40	90.84	90.84	81.69
	M50	90.83	90.83	81.66
	C40	90.49	90.49	80.97
	C50	90.55	90.55	81.1
	A40	90.49	90.49	80.97
	A50	90.55	90.55	81.1
	F40	91.1	91.1	82.21
	F50	91.25	91.25	82.5
	TT40	90.74	90.74	81.48
	TT50	90.88	90.88	81.76
	MR40	90.4	90.4	80.8
	MR50	90.94	90.94	81.88
	CON	91.36	91.36	82.72
	CFS	90.9	90.9	81.81
	ORG	91.26	91.26	82.52
	Boruta	91.15	91.15	82.3
	SFS	**91.52**	**91.52**	**83.04**
	PI	91.33	91.33	82.67
	Shap	**91.54**	**91.54**	**83.09**
	GA	91.01	91.01	82.01
	RFE	91.15	91.15	82.3

(*continued*)

Table 6. (*continued*)

Classifier	FS technique	Accuracy(%)	F1_score(%)	Kappa(%)
LGBM	M40	90.45	90.45	80.9
	M50	90.71	90.71	81.42
	C40	89.93	89.93	79.86
	C50	90.2	90.2	80.41
	A40	89.93	89.93	79.86
	A50	90.2	90.2	80.41
	F40	90.31	90.31	80.64
	F50	90.48	90.48	80.97
	TT40	89.67	89.67	79.34
	TT50	90.41	90.41	80.83
	MR40	89.84	89.84	79.67
	MR50	90.3	90.3	80.6
	ORG	91.13	91.13	82.25
	CON	91.09	91.09	82.18
	CFS	90.6	90.6	81.22
	RFE	91.02	91.02	82.03
	PI	91.03	91.03	82.06
	Boruta	91.02	91.02	82.03
	Shap	91.13	91.13	82.25
	GA	91.01	91.01	82.03
	SFS	**91.28**	**91.28**	**82.56**
DT	M40	89.06	89.06	78.12
	M50	88.94	88.94	77.88
	C40	88.88	88.88	77.75
	C50	88.94	88.94	77.87
	A40	88.85	88.85	77.69
	A50	89.05	89.05	78.1
	F40	88.94	88.94	77.87
	F50	88.81	88.81	77.61
	TT40	87.93	87.93	75.86
	TT50	88.61	88.61	77.21
	MR40	88.69	88.69	77.38
	MR50	88.62	88.62	77.24
	ORG	**89.43**	**89.43**	**78.86**
	CON	88.84	88.84	77.68
	CFS	88.83	88.83	77.65
	RFE	89.21	89.21	78.42
	PI	89.16	89.16	78.32
	Boruta	**89.43**	**89.43**	**78.86**
	Shap	89.12	89.12	78.25
	GA	89.09	89.09	78.18
	SFS	89.25	89.25	78.49

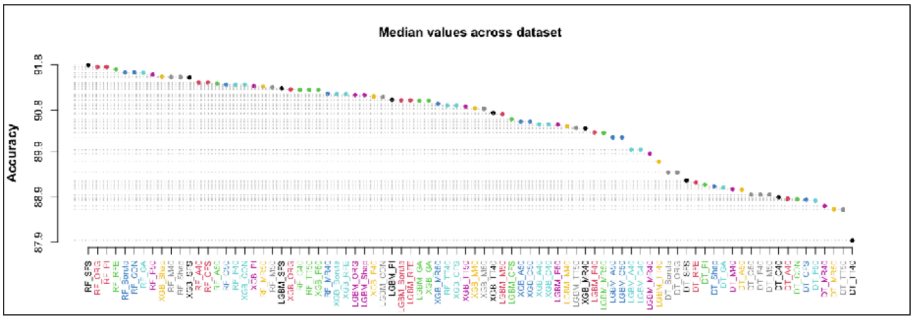

Fig. 1. SK plot of filter (univariate and multivariate) and wrapper techniques over the seven datasets.

4 Threats to Validity

This section summarizes the main validity concerns of the study.

Internal Validity: To reduce overfitting, 5-fold cross-validation was employed. Nevertheless, further hyperparameter tuning (e.g., pruning) could improve model generalization. Although the implementation was carefully verified, minor errors may still exist. The results could also vary if other classifiers (e.g., KNN, Logistic Regression, or Naïve Bayes) were used. In addition, computational cost was not assessed, even though univariate filters are generally less expensive than multivariate and ensemble approaches.

External Validity: The experiments were conducted on seven datasets from the *Phoenicurus* and *Oenanthe* groups. Consequently, findings may differ with other datasets, classifiers or ranking strategies. Replication using additional feature selection techniques and more diverse datasets would enhance generalizability.

Construct Validity: Model performance was evaluated using three widely adopted metrics: accuracy, F1-score, and Cohen's Kappa. Their results were aggregated using the SK test and the BC method with equal weights, ensuring a balanced and robust evaluation.

5 Conclusion

This study empirically compared 588 configurations of four classifiers (LGBM, RF, DT, and XGB) across seven bird species datasets. The comparison included: (1) six univariate filters with two selection thresholds (40% and 50%) plus the original feature set, (2) two multivariate techniques, and (3) six wrapper methods. Performance was assessed using accuracy, Cohen's Kappa, and F1-score, combined through the SK test and BC ranking. The main findings are summarized as follows:

(RQ1): SFS achieved the highest overall performance among feature selection methods. However, considering feature reduction efficiency, the Fisher Score combined with RF was the most economical approach, selecting an average of nine features.

Acknowledgments. This research was supported by the National Center for Scientific and Technical Research (CNRST) of Morocco under the "PhD-Associate Scholarship - PASS" program.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

Appendix

SK test results over the seven datasets for the filter and wrappers FS techniques

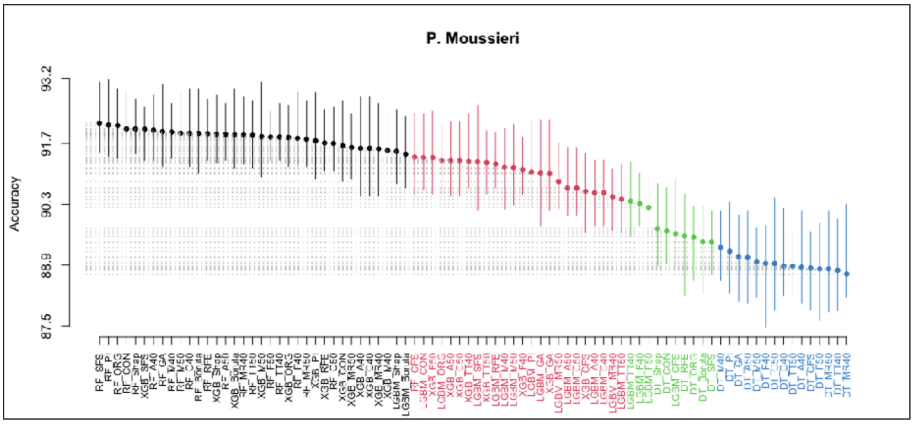

Fig. 2. SK test plot of the filter and wrapper methods across P.moussieri dataset.

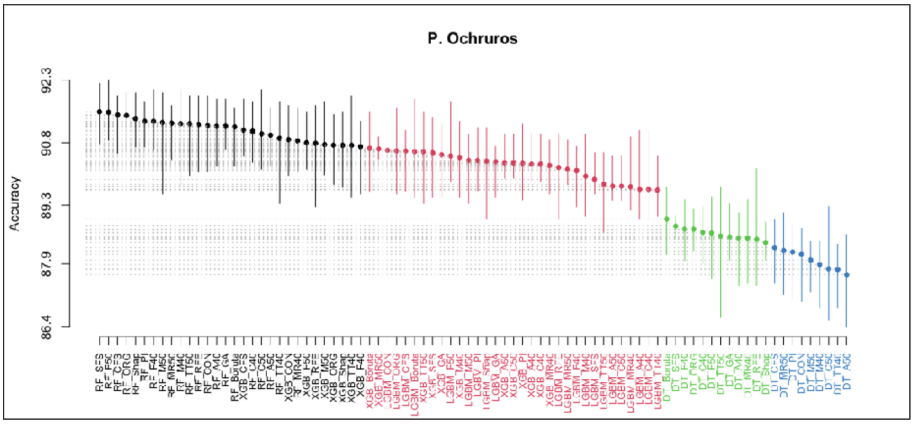

Fig. 3. SK test plot of the filter and wrapper methods across P.ochruros dataset.

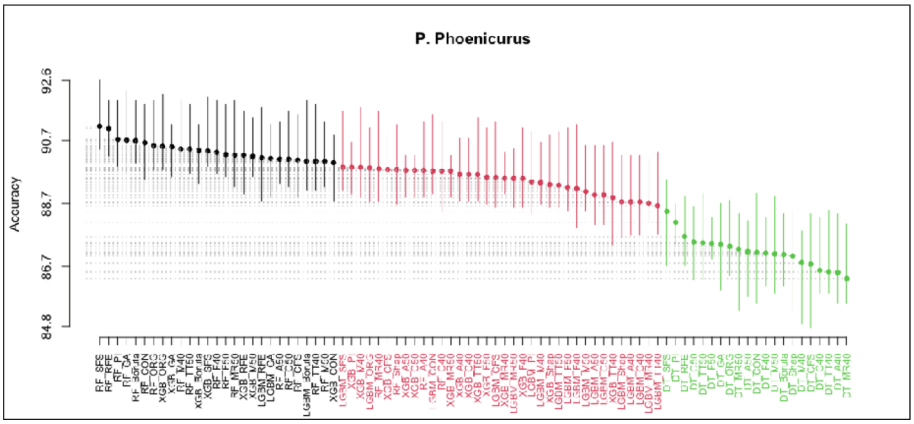

Fig. 4. SK test plot of the filter and wrapper methods across P.phoenicurus dataset.

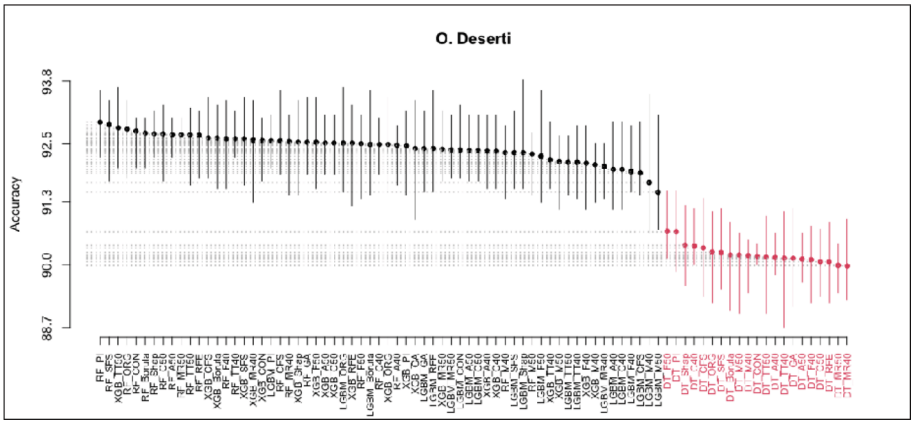

Fig. 5. SK test plot of the filter and wrapper methods across O.deserti dataset.

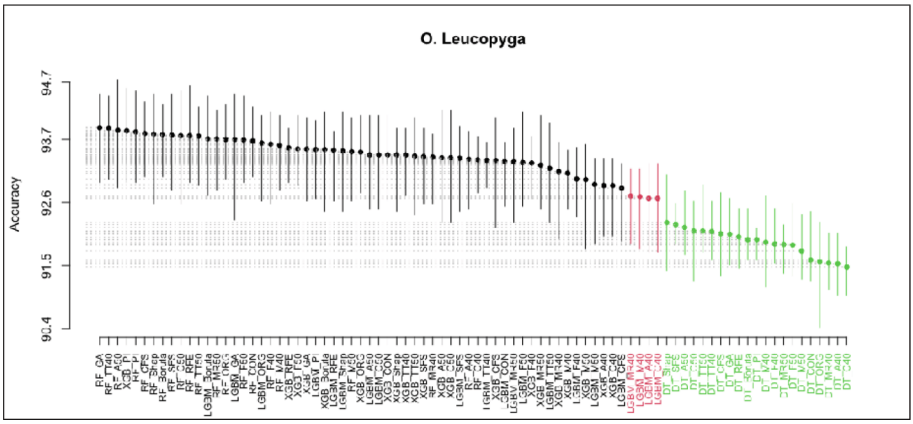

Fig. 6. SK test plot of the filter and wrapper methods across O.leucopyga dataset.

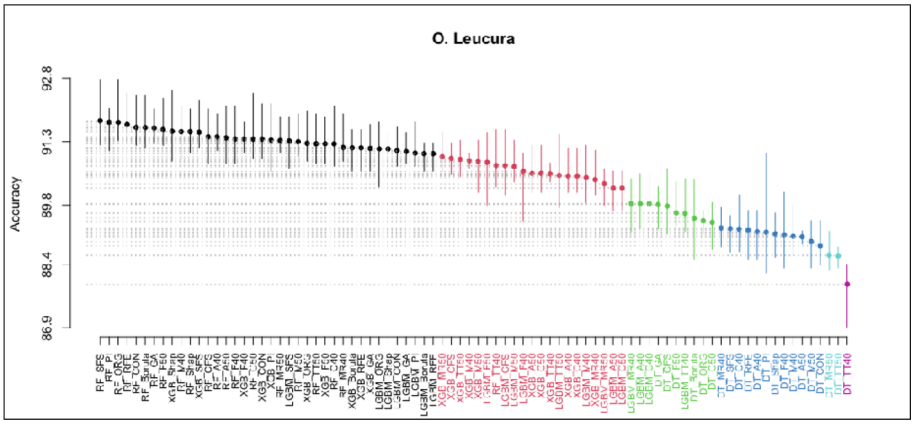

Fig. 7. SK test plot of the filter and wrapper methods across O.leucura dataset.

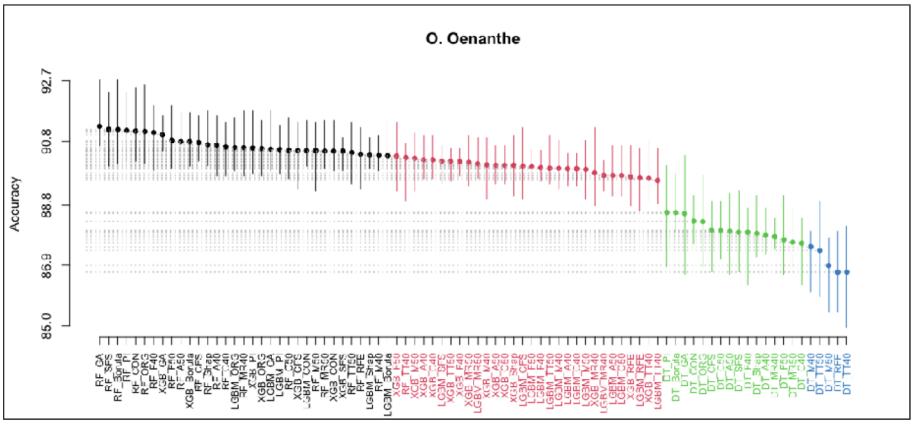

Fig. 8. SK test plot of the filter and wrapper methods across O.oenanthe dataset.

References

1. Barbet-Massin, M., Jiguet, F., Albert, C.H., Thuiller, W.: Selecting pseudo-absences for species distribution models: how, where and how many? Methods Ecol. Evol. **3**(2), 327–338 (2012)
2. Benhar, H., Idri, A., Hosni, M.: Impact of threshold values for filter-based univariate feature selection in heart disease classification. In: HEALTHINF, pp, pp. 391–398 (2020)
3. Bolón-Canedo, V., Sánchez-Maroño, N., Alonso-Betanzos, A.: Recent advances and emerging challenges of feature selection in the context of big data. Knowl.-Based Syst. **86**, 33–45 (2015). https://doi.org/10.1016/j.knosys.2015.05.014, https://www.sciencedirect.com/science/article/pii/S0950705115002002
4. Calinski, T., Corsten, L.: Clustering means in anova by simultaneous testing. Biometrics, 39–48 (1985)
5. Chen, C.W., Tsai, Y.H., Chang, F.R., Lin, W.C.: Ensemble feature selection in medical datasets: combining filter, wrapper, and embedded feature selection results. Expert. Syst. **37**(5), e12553 (2020)
6. CHERIF, Y., IDRI, A.: Evaluating the efficacy of wrappers in predicting species distribution. In: 2024 World Conference on Complex Systems (WCCS), pp. 1–6 (2024). https://doi.org/10.1109/WCCS62745.2024.10765519
7. Cherif, Y., Idri, A.: Single and ensemble based filters in environmental data. Expert. Syst. **42**(7), e70076 (2025)
8. Cherif., Y., Idri., A., El Alaoui., O.: Impact of thresholds of univariate filters for predicting species distribution. In: Proceedings of the 15th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management - KDIR, pp. 86–97. INSTICC, SciTePress (2023). https://doi.org/10.5220/0012203000003598
9. Effrosynidis, D., Arampatzis, A.: An evaluation of feature selection methods for environmental data. Eco. Inform. **61**, 101224 (2021). https://doi.org/10.1016/j.ecoinf.2021.101224, https://www.sciencedirect.com/science/article/pii/S1574954121000157

10. Fick, S., Hijmans, E., R.: Worldclim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol., 1–14 (2017)
11. GBIF, biodiversity, species occurrences (2023). https://doi.org/10.15468/dl.7uxqxk, https://www.gbif.org/occurrence/download/0239931-220831081235567
12. GBIF, biodiversity, species occurrences (2023). https://doi.org/10.15468/dl.vuzsmjhttps://www.gbif.org/occurrence/download/0003099-230530130749713
13. Guyon, I., Gunn, S., Nikravesh, M., Zadeh, L.A.: Feature extraction: foundations and applications, vol. 207. Springer (2008). https://doi.org/10.1007/978-3-540-35488-8
14. Jaganathan, P., Kuppuchamy, R.: A threshold fuzzy entropy based feature selection for medical database classification. Comput. Biol. Med. **43**(12), 2222–2229 (2013)
15. Jovic, A., Brkić, K., Bogunovic, N.: A review of feature selection methods with applications. pp. 1200–1205 (May 2015). https://doi.org/10.1109/MIPRO.2015.7160458
16. Lahmar, C., Idri, A.: On the value of deep learning for diagnosing diabetic retinopathy. Heal. Technol., 1–17 (2022)
17. Li, J., et al.: Feature selection: a data perspective. ACM Comput. Surv. (CSUR) **50**(6), 1–45 (2017)
18. Lichman, M.: Uci machine learning repository, p. 2013 (2013)
19. Liu, H., Yu, L.: Toward integrating feature selection algorithms for classification and clustering. IEEE Trans. Knowl. Data Eng. **17**(4), 491–502 (2005)
20. Lobo, J.M., Jiménez-Valverde, A., Hortal, J.: The uncertain nature of absences and their importance in species distribution modelling. Ecography **33**(1), 103–114 (2010)
21. MacKenzie, D.I., Royle, J.A.: Designing occupancy studies: general advice and allocating survey effort. J. Appl. Ecol. **42**(6), 1105–1114 (2005)
22. Nyitrai, T., Virág, M.: The effects of handling outliers on the performance of bankruptcy prediction models. Socioecon. Plann. Sci. **67**, 34–42 (2019)
23. Orlandi, A.G., Carvalho, O.A., Guimarães, R.F., Bias, E.D.S., Corrêa, D.C., Gomes, R.A.T.: Vertical accuracy assessment of the processed srtm data for the brazilian territory. Boletim de Ciências Geodésicas **25** (2019)
24. Peter, T.J., Somasundaram, K.: Study and development of novel feature selection framework for heart disease prediction. Int. J. Sci. Res. Publ. **2**(10), 1–7 (2012)
25. Seijo-Pardo, B., Porto-Díaz, I., Bolón-Canedo, V., Alonso-Betanzos, A.: Ensemble feature selection: homogeneous and heterogeneous approaches. Knowl.-Based Syst. **118**, 124–139 (2017)
26. Senay, S.D., Worner, S.P., Ikeda, T.: Novel three-step pseudo-absence selection technique for improved species distribution modelling. PLoS ONE **8**(8), e71218 (2013)
27. VanDerWal, J., Shoo, L.P., Graham, C., Williams, S.E.: Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? Ecol. Model. **220**(4), 589–594 (2009)
28. Worsley, K.J.: Confidence regions and tests for a change-point in a sequence of exponential family random variables. Biometrika **73**(1), 91–104 (1986)
29. Zouhri, H., Idri, A.: A comparative assessment of wrappers and filters for detecting cyber intrusions. In: World Conference on Information Systems and Technologies, pp. 118–127. Springer (2024). https://doi.org/10.1007/978-3-031-60221-4_12
30. Zouhri, H., Idri, A., Ratnani, A.: Evaluating the impact of filter-based feature selection in intrusion detection systems. Int. J. Inf. Secur. **23**(2), 759–785 (2024)