Pesquisa Agropecuária Brasileira

Digital Agriculture/ Original Article

ISSN 1678-3921

Journal homepage: www.embrapa.br/pab

For manuscript submission and journal contents, access: www.scielo.br/pab

Remote detection of the invasive grass *Eragrostis plana* using spectral data and machine learning

Abstract – The objective of this work was to evaluate the efficiency of spectral analysis combined with machine learning algorithms, to detect areas infested by *Eragrostis plana* in the Pampa biome. Spectral data were collected using a spectroradiometer across wavelengths from 350 to 2,500 nm and categorized into three classes: *E. plana* only; mixed-class; and other plants. The dataset comprised 42 samples distributed as follows: 5 *E. plana* only; 16 mixed-class; and 21 other plants. To address class imbalance and improve model generalization, Mixup and Adasyn data augmentation techniques were applied. Three datasets were tested: the full spectrum (1,688 bands); 30 bands selected by the Random Forest algorithm; and 28 bands previously reported in the literature. Decision tree, random forest, support vector machine, and artificial neural network classifiers were applied. The best performance (88.5% accuracy, F1-score of 0.73) was achieved in a binary classification scenario using the 30 bands selected.The performance decreased in the multiclass setup, likely due to spectral overlap (Hughes phenomenon). Spectral analysis combined with machine learning algorithms has potential for remote detection of *E. plana*, contributing to the management of invasive species in the Pampa biome.

Index terms: hyperspectral data, invasive species, pasture management, remote sensing.

André Rosa Silveira✉
Universidade Federal do Pampa, Bagé, RS, Brazil. E-mail: andrers.aluno@unipampa.edu.br

Sandro da Silva Camargo
Universidade Federal do Pampa, Bagé, RS, Brazil.
E-mail: sandrocamargo@unipampa.edu.br

Marcos Corrêa Neves
Embrapa Meio Ambiente, Jaguariúna, SP, Brazil. E-mail: marcos.neves@embrapa.br

Fabio Marcelo Breunig
Universidade Federal do Paraná, Departamento de Geografia, Curitiba, PR, Brazil.
E-mail: fabiobreunig@ufpr.br

Naylor Bastiani Perez
Embrapa Pecuária Sul, Bagé, RS, Brazil.
E-mail: naylor.perez@embrapa.br

✉ Corresponding author

Received
July 22, 2025

Accepted
April 9, 2026

How to cite
SILVEIRA, A.R.; CAMARGO, S. da S.; NEVES, M.C.; BREUNIG, F.M.; PEREZ, N.B. Remote detection of the invasive grass *Eragrostis plana* using spectral data and machine learning. **Pesquisa Agropecuária Brasileira**, v.61, e04198, 2026. DOI: https://doi.org/10.1590/S1678-3921.pab2026.v61.04198.

Detecção remota da gramínea invasora *Eragrostis plana* por meio de dados espectrais e aprendizado de máquina

Resumo – O objetivo deste trabalho foi avaliar a eficiência da análise espectral combinada com algoritmos de aprendizado de máquina, para detectar áreas infestadas por *Eragrostis plana* no bioma Pampa. Os dados espectrais foram coletados com um espectrorradiômetro, com abrangência de comprimentos de onda de 350 a 2.500 nm, e foram classificados em três categorias: somente *E. Plana*; misto; e outras plantas. A base de dados foi composta por 42 amostras distribuídas da seguinte forma: 5 somente com *E. Plana*; 16 mistas; e 21 de outras plantas. Para lidar com o desbalanceamento entre classes e melhorar a capacidade de generalização dos modelos, as técnicas de aumento de dados Mixup e Adasyn foram aplicadas. Três conjuntos de dados foram testados: o espectro completo (1.688 bandas), 30 bandas selecionadas pelo algoritmo Random Forest, e 28 bandas previamente recomendadas na literatura. Foram aplicados classificadores baseados em árvores de decisão, florestas aleatórias, máquinas de vetor de suporte e redes neurais artificiais. O melhor desempenho (88,5% de acurácia e F1-score de 0,73) foi obtido em um cenário de classificação binária com as 30 bandas selecionadas. O desempenho diminuiu na configuração multiclasse, possivelmente em razão da sobreposição espectral (fenômeno de Hughes). A análise espectral

This is an open-access article distributed under the Creative Commons Attribution 4.0 International License

combinada com algoritmos de aprendizado de máquina é eficiente para a detecção remota de *E. plana*, o que contribui para o manejo de espécies invasoras no bioma Pampa.

Termos para indexação: dados hiperespectrais, espécies invasoras, manejo de pastagens, sensoriamento remoto.

Introduction

In the state of Rio Grande do Sul (RS), Southern Brazil, *Eragrostis plana* Nees, commonly known as *capim-annoni*, stands out as one of the most aggressive invasive grasses. Originally introduced from Africa in the 1950s as forage, it was later recognized as invasive due to its low nutritional value, mechanical resistance, high reproductive capacity, and adaptation to nutrient-poor soils (Goulart et al., 2009). Its rapid spread has severely affected both the productivity and the floristic diversity of native grasslands in the Pampa biome (Michel & Overbeck, 2024).

The scientific literature reports a number of studies aimed at developing and evaluating methods to monitor *E. plana* infestations, particularly by the use of remote sensing technologies and predictive modeling approaches. A study conducted in shallow basaltic grasslands in RS state, reported that 17% of the mapped area exhibited high levels of *E. plana* infestation, based on Landsat 8 imagery and supervised classification techniques (Cicconet et al., 2015). In another study conducted in areas dominated by Vertisols, the invasion potential of *E. plana* was evaluated using the species distribution model of the Genetic Algorithm for Rule-Set Production (GARP), combined with remote sensing data. Spectral information from Landsat 8 bands (blue, green, red, near infrared, and shortwave infrared) showed that nearly 30% of the municipality of Aceguá, RS, Brazil, was potentially susceptible to invasive plants, with road networks contributing to the expansion of infested areas (Montoya González et al., 2017).

Alternative remote-sensing strategies have also shown promise. Neves et al. (2019) used multispectral sensors mounted on unmanned aerial vehicles (UAVs), to map *E. plana* infestation in natural pastures in the municipality of Bagé, RS, Brazil. Their results showed that remote sensing enabled effective discrimination of infested areas, with correlation coefficients above 0.9, especially when using red edge and near-infrared bands. Strong negative correlations were observed between NDVI and key indicators such as total dry biomass (-0.85), dry green matter of *E. plana* (-0.84), and infestation level (-0.82), reinforcing the method's potential for monitoring. Further, Gomes et al. (2020) reported the feasibility of using Sentinel-2 multispectral data in conjunction with machine-learning algorithms to map *E. plana*. However, as their validation relied partly on destructive sampling, this approach may be limited to areas where infestation is already known, thus constraining its large-scale applicability. Akbarzadeh et al. (2019) tested real-time weed detection using spectral reflectance sensors and machine-learning models, including SVM and convolutional neural networks, achieving over 80% accuracy and high-processing speed, supporting its use under operational field conditions.

These studies underscore the potential of integrating spectral data with machine-learning algorithms to detect invasive species. However, they also show critical gaps, particularly in the application of high-resolution spectral data validated under real-world field conditions for the specific case of *E. plana*. Recent work has shown the successful application of UAV-derived orthomosaics and spectral data, for identifying invasive tree species in protected areas in Southern Brazil (Silva et al., 2023), further highlighting the relevance of this approach.

The objective of this work was to evaluate the efficiency of spectral analysis combined with machine-learning algorithms, to detect areas infested by *E. plana* in Pampa biome.

Materials and Methods

The study was carried out in the Pampa biome, in the state of Rio Grande do Sul, Brazil, under a humid subtropical climate (Cfa) with hot summers (Alvares et al., 2013) and the predominance of Luvisol.

Data were collected in a pasture infested with *capim-annoni* (*Eragrostis plana* Nees), including other native herbaceous species. Sampling was carried out over about 15 acres (at 31°19'18.38"S, 53°59'14.70"W; 31°19'24.53"S, 53°59'18.55"W; 31°19'29.25"S, 53°59'8.59"W; and 31°19'21.83"S, 53°59'5.45"W), by the WGS 84 datum, in April 2016. Spectral signatures were collected through systematic walking, aiming to sample the greatest plant diversity, as well as the greatest diversity offered by the phenotypic plasticity

of *E. plana*. The autumn period, before the first frosts, becomes favorable for capturing the floristic diversity of native vegetation under grazing and the phenological diversity of *E. plana* plants, including flowering plants and plants in the vegetative stage.

Spectral measurements were obtained using a spectroradiometer (FieldSpec 4, Automated Systems Design - ASD Inc., Alpharetta, GA, USA), operating in the 350–2,500 nm range, which includes the visible (VIS), near-infrared (NIR), and shortwave infrared (SWIR) regions. Prior to the acquisition, the instrument was powered on for 30 min, to ensure the detector stabilization. Radiometric calibration was performed using a diffuse reflectance white reference panel, with recalibration every 10 min. The measurements were carried out between 10:00 and 12:00 solar time under clear skies and low-wind conditions. Data acquisition and visualization were performed using RS3 or ViewSpec Pro software.

The sensor was positioned in nadir orientation, at 60 cm above the vegetation canopy, ensuring that the sensor footprint encompassed only the target area while avoiding operator shadow. Canopy-level measurements were performed in situ to capture the composite reflectance of vegetation and exposed soil, which is consistent with realistic remote-sensing scenarios. Metadata recorded for each sample included acquisition date and time and species identification.

A total of 46 spectral samples were initially acquired. After the removal of outliers and noisy spectral bands, based on the wavelength exclusion criteria proposed by Watanabe et al. (2013), the final dataset consisted of 42 samples and 1,688 spectral bands. Because spectral acquisition is nondestructive, the sampled plant material remained available for subsequent laboratory analyses. Three datasets were defined to evaluate the effect of spectral dimensionality on the classification performance, described as follows. The first one is a full-spectrum dataset with 1,688 spectral bands. The second one is a reduced dataset with 30 bands that was selected using the Gini importance index, from the Random Forest algorithm, which prioritizes variables with high-discriminatory power and low covariance (Yano et al., 2016). The third dataset showed 28 bands selected from the literature, identified by Thenkabail et al. (2018) as optimized for vegetation discrimination.

A statistical assessment of spectral band relevance was conducted using the Kruskal–Wallis test, followed by Dunn post hoc test with Bonferroni correction. No spectral band showed significant differences at the 95% confidence level (p-values ranged from 0.201 to 0.222 for the 30 bands with the highest importance). Five classification algorithms were evaluated based on their effectiveness in detecting invasive or weed species in previous studies as following descriptions. The first algorithm is RPART, which was selected for its interpretability and efficiency in agricultural classification (Hastie et al., 2009; Fávero et al., 2023). The second algorithm is the Random Forest (RF), which was successfully applied by Peña et al. (2015), using UAV imagery for weed detection in sunflower and cotton crops. The third algorithm is the Support Vector Machine (SVM), reported as effective in precision agriculture, which was used especially for classifying crops and invasive species (Abouzahir et al., 2018). The fourth algorithm is the artificial neural network (ANN), used by Azeem et al. (2020) to estimate leaf area of the invasive species *Wedelia trilobata*. The fifth algorithm is the Spectral Angle Mapper (SAM), which was employed by Okoye & Hosseini (2024) for the identification of invasive species from hyperspectral airborne imagery under variable lighting conditions.

To address the class imbalance between "annoni only" and "other plants", two oversampling strategies were used: Mixup, which generates synthetic samples by interpolating original samples (Smucny et al., 2022); and Adasyn (adaptive synthetic sampling), which focuses on minority classes located in difficult-to-classify regions (Saleem et al., 2023).

The model validation was performed using two cross-validation approaches: Stratified 10-fold cross validation, which maintains class distribution in each fold (Arlot & Celisse, 2010); and Leave-One-Out Cross Validation (LOOCV), which maximizes training data use and is suitable for small datasets (Haug et al., 2014; Najafabadi et al., 2023).

The model performance was evaluated using the following metrics: overall accuracy, precision, recall (sensitivity), F1-score, and kappa coefficient (Liakos et al., 2018; Cechinel & Camargo, 2020). All metrics were computed for both the binary classification scenario (infested vs. noninfested) and the multiclass scenario (*E. plana* only, mixed-class, and other plants).

Results and Discussion

The evaluation of classification models for the detection of *E. plana* infestation showed that the performance was strongly influenced by spectral dimensionality, the classifier employed, and the complexity of the classification scenario (binary or multiclass). Such sensitivity to the methodological design is consistent with a recent review, which emphasizes that the model choice and feature representation critically determine the success of vegetation identification using spectral or multispectral data (Chanchí Golondrino et al., 2023).

In the full-spectrum dataset (1,688 bands), the SVM model yielded the highest performance (accuracy and F1-score) in the binary classification scenario, while also presenting the lowest false negative rate (Table 1). The superior performance of SVM in high-dimensional hyperspectral contexts aligns with studies showing that the kernel-based methods remain highly competitive, when spectral separability exists and dimensionality is large (Datta et al., 2022), and it has the ability to manage cases of spectral overlap (Breunig et al., 2011). In contrast, both tree-based models − RPART and RF − displayed lower performance and were associated with higher false positive and false negative rates. From a practical standpoint, this result is highly relevant because minimizing false negatives is essential for early detection of invasive weeds, enabling timely management interventions and reducing ecological and economic impacts. Therefore, the present findings reinforce the evidence that machine-learning models applied to proximal spectral sensing can support the precision vegetation monitoring (Chanchí Golondrino et al., 2023).

In contrast, the performance declined markedly in the multiclass scenario, with accuracies ranging from 0.38 to 0.54, and F1-scores were below 0.47. The reduced performance in vegetation discrimination has been widely reported and has been generally attributed to spectral similarity among vegetation classes, phenological variability, and mixed background effects. Thus, the observed decline was expected, particularly because the intermediate mixed-class likely shares biochemical and structural spectral traits with both pure vegetation classes (Fu et al., 2017).

The low accuracy of the spectral angle mapper (SAM) further illustrates this limitation, as angle-based similarity metrics are inherently constrained when interclass spectral separability is weak, or when spectral similarity among vegetation and background components increases the classification ambiguity (Zhao et al., 2022).

Despite these constraints, documenting such limitations is important because it clarifies the operational boundaries of classical hyperspectral classifiers for detecting *E. plana* under realistic pasture conditions, an aspect rarely quantified for South American grassland invasions.

Feature reduction using RF-selected bands (30 bands) substantially improved classification in the binary scenario, particularly for RPART combined with Adasyn oversampling (Table 2). Improvements after band selection are consistent with recent hyperspectral weed-detection studies, showing that noise reduction and removal of redundant wavelengths enhance discrimination and model stability (Datta et al., 2022).

Table 1. Performance of the models using 1,688 wavelengths to identify areas infested with annoni grass.

Model	Oversampling	Scenario	Acc.	Prec.	Sens.	F1	FP	FN
RPART	none	Binary	0.73	0.33	0.40	0.36	0.15	0.12
RF	none	Binary	0.73	0.33	0.40	0.36	0.15	0.12
SVM	none	Binary	0.81	0.50	0.60	0.55	0.12	0.08
RPART	none	Multiclass	0.42	0.33	0.29	0.31	0.24	0.29
RF	none	Multiclass	0.47	0.27	0.23	0.25	0.19	0.24
SVM	none	Multiclass	0.54	0.50	0.44	0.47	0.19	0.24
SAM	none	Binary	0.60	0.25	0.50	0.33	0.30	0.10
SAM	none	Multiclass	0.38	0.33	0.50	0.40	0.12	0.06

Models: RPART, decision tree algorithm; RF, random forest algorithm; SVM, support vector machine algorithm; SAM, spectral angle mapper algorithm. Parameters: Acc., overal accuracy; Prec., precision; Sens., sensitivity; F1, F1 score; FP, false positives; FN, false negatives.

Under certain reduced-feature configurations, the poor performance of SVM and ANN suggests sensitivity to attribute distribution and class imbalance − that are phenomena also reported in modern plant-classification benchmarks, in which minority classes or limited spectral variability reduce the model generalization (Akbarzadeh et al., 2019). Consequently, the present study advances the current knowledge by showing that data-driven spectral selection, combined with imbalance correction, may outperform traditionally robust classifiers, highlighting the importance of preprocessing strategy over algorithm choice alone.

In the multiclass scenario with 30 bands, accuracy ranged from 0.524 to 0.595, indicating that spectral overlap cannot be resolved solely through dimensionality reduction. Similar constraints have been described in recent research on plant identification, in which even advanced models showed a reduced discrimination when species shared physiological or structural traits across growth stages (Akbarzadeh et al., 2019). The particularly weak SVM performance in this context reinforces that classifier robustness in binary classification does not necessarily translate to transitional or mixed-class discrimination, underscoring the need for hierarchical or deep-learning approaches in complex ecological scenarios.

For the dataset containing 28 bands selected from the literature (Table 3), the models exhibited a lower performance overall, suggesting a limited transferability of generic vegetation wavelengths to the specific spectral behavior of *E. plana* in the Pampa biome. This observation is coherent with recent studies reporting that site-specific spectral variability and phenological context strongly influence the optimal feature selection, reducing the effectiveness of universal band sets (Basinger et al., 2020). Therefore, evidence shows that locally optimized spectral features are critical for an accurate detection of invasive species.

Table 2. Performance metrics of classification models using 30 wavelengths selected by random forest.

Model	Oversampling	Scenario	Acc.	Prec.	Sens.	Spec.	F1	FP	FN
RPART	ADASYN	Binary	0.885	0.667	0.800	0.905	0.727	0.077	0.038
RF	ADASYN	Binary	0.808	0.500	0.600	0.857	0.545	0.115	0.077
SVM	ADASYN	Binary	0.808	0.000	0.000	1.000	0.000	0.000	0.192
ANN	ADASYN	Binary	0.769	0.000	0.000	0.952	0.000	0.038	0.192
Rpart	MIXUP	Binary	0.692	0.200	0.200	0.810	0.200	0.154	0.154
RF	MIXUP	Binary	0.731	0.250	0.200	0.857	0.222	0.115	0.154
SVM	MIXUP	Binary	0.808	0.000	0.000	1.000	0.000	0.000	0.192
ANN	MIXUP	Binary	0.808	0.500	0.200	0.952	0.286	0.038	0.154
Rpart	none	Binary	0.846	0.600	0.600	0.905	0.600	0.077	0.077
RF	none	Binary	0.808	0.500	0.600	0.857	0.545	0.115	0.077
SVM	none	Binary	0.808	0.000	0.000	1.000	0.000	0.000	0.192
ANN	none	Binary	0.808	0.000	0.000	1,000	0.000	0.000	0.192
RPART	ADASYN	Multiclass	0.595	0.550	0.630	-	0.565	0.143	0.119
RF	ADASYN	Multiclass	0.571	0.472	0.466	-	0.468	0.119	0.143
SVM	ADASYN	Multiclass	0.190	0.107	0.127	-	0.116	0.310	0.405
ANN	ADASYN	Multiclass	0.524	0.434	0.430	-	0.430	0.143	0.190
RPART	MIXUP	Multiclass	0.429	0.321	0.320	-	0.320	0.238	0.214
RF	MIXUP	Multiclass	0.548	0.474	0.451	-	0.456	0.143	0.190
SVM	MIXUP	Multiclass	0.476	0.304	0.347	-	0.324	0.167	0.286
ANN	MIXUP	Multiclass	0.405	0.269	0.290	-	0.279	0.190	0.262
RPART	none	Multiclass	0.571	0.523	0.609	-	0.539	0.143	0.143
RF	none	Multiclass	0.595	0.480	0.482	-	0.481	0.095	0.119
SVM	none	Multiclass	0.190	0.107	0.127	-	0.116	0.310	0.405
ANN	none	Multiclass	0.595	0.489	0.482	-	0.481	0.095	0.190

Models: RPART, decision tree algorithm; RF, random forest algorithm; SVM, support vector machine algorithm; ANN, artificial neural network algorithm; Acc., accuracy; Prec., precision; Sens. , sensitivity; Spec., specificity; F1-score; FP , false positive rate; FN , false negative rate.

In the binary classification scenario, the ANN model achieved the best results, followed by RPART. Without balancing techniques, the models RF and SVM returned F1-scores of zero. When Adasyn was applied, modest improvements were observed, although these did not surpass the best results obtained with the RF selected bands.

In the multiclass scenario, the maximum accuracy did not exceed 0.595, and F1-scores remained below 0.481 (Table 2). These findings suggest that while literature-based bands are broadly applicable in vegetation analysis (Thenkabail et al., 2018), they may not adequately capture the specific spectral patterns of *E. plana* in the Pampa biome. Similar findings were reported by Basinger et al. (2020) for different experimental contexts.

Overall, the poor performance observed in the multiclass classification, across all datasets, suggests that spectral overlap, particularly involving the mixed class, remains a limiting factor for accurate discrimination. This fact highlights the need for alternative approaches, such as hierarchical classification, prior segmentation, or deep neural networks, to overcome class confusion caused by subtle spectral differences. The decision tree generated from the 30-band dataset, balanced with Adasyn (Figure 1), provided valuable insights into the spectral variables driving the classification decisions. Although RPART did not achieve the top overall performance, its hierarchical structure allowed for clear interpretation of relevant wavelengths.

Initial splits in the tree relied on bands located in the shortwave infrared region (for instance, 2,286 and 2,305 nm), which are often associated with lignin, cellulose, or soil background effects (Thenkabail et al., 2018). Intermediate and terminal nodes frequently employed bands from the visible, red edge, and NIR regions (for instance, 717, 751, and 362 nm), which reflect physiological and structural traits, such as chlorophyll content and leaf area index (Thenkabail et al., 2018). These findings confirm that the selected wavelengths possess high-discriminatory power, particularly for identifying pure *E. plana* patches.

Subsequent splits in VIS, red-edge, and NIR regions reflect chlorophyll content and canopy structure, confirming that the effective identification of pure *E. plana* patches depends on integrated physiological and biochemical spectral responses rather than a single diagnostic band. Similar multispectral dependence has been repeatedly reported in recent plant-sensing

Table 3. Performance of classification models using 28 wavelengths, selected from the literature to identify areas infested with annoni grass.

Model	Oversampling	Scenario	Acc.	Prec.	Sens.	F1	FP	FN
RPART	none	Binary	0.77	0.43	0.60	0.50	0.15	0,08
RF	none	Binary	0.69	0.00	0.00	0.00	0.12	0.19
SVM	none	Binary	0.81	0.00	0.00	0.00	0.00	0.19
ANN_Acc	none	Binary	0.81	0.50	0.60	0.55	0.12	0.08
ANN_Prec	none	Binary	0.77	0.40	0.40	0.40	0.12	0.12
ANN_Sens	none	Binary	0.73	0.25	0.20	0.22	0.12	0.15
ANN_F1	none	Binary	0.81	0.50	0.40	0.44	0.08	0.12
RPART	mixup	Binary	0.77	0.43	0.60	0.50	0.15	0.08
RF	mixup	Binary	0.77	0.40	0.40	0.40	0.12	0.12
SVM	mixup	Binary	0.69	0.38	1.00	0.56	0.31	0.00
ANN	mixup	Binary	0.73	0.38	0.60	0.46	0.19	0.08
RPART	ADASYN	Binary	0.77	0.43	0.60	0.50	0.15	0.08
RF	ADASYN	Binary	0.77	0.40	0.40	0.40	0.12	0.12
SVM	ADASYN	Binary	0.73	0.42	1.00	0.59	0.27	0.00
ANN	ADASYN	Binary	0.73	0.38	0.60	0.46	0.19	0.08

Models: RPART, decision tree algorithm; RF, random forest algorithm; SVM, support vector machine algorithm; ANN, artificial neural network algorithm; Acc., Accuracy; Prec., precision; Sens., sensitivity; F1-score; FP , false positive rate; FN , false negative rate. When specified, ANN models were optimized for different performance metrics. Other models yielded identical results, regardless of the optimization criterion.

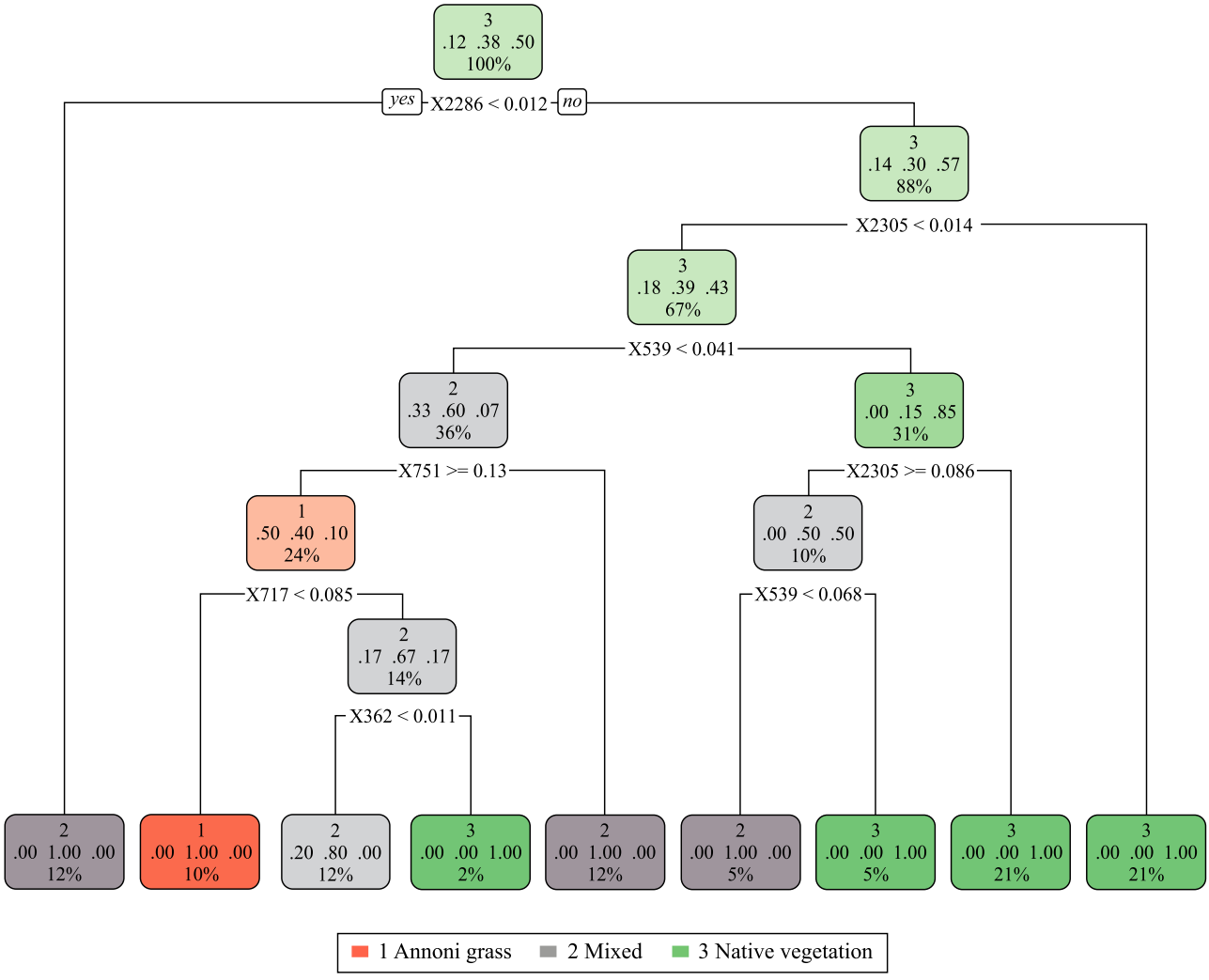

Figure 1. Decision tree generated using the set of 30 wavelengths selected by Random Forest, with data balancing via ADASYN. The numbers on the branches represent attribute thresholds, and the terminal nodes indicate the predicted class: 1 – *E. plana*, 2 – mixed, 3 – native vegetation.

studies (Gámez et al., 2025). Overall, the findings show that class imbalance and interclass spectral similarity remain the dominant barriers to an accurate remote detection of *E. plana*. By quantifying how feature selection, resampling, and classifier choice interact under realistic pasture conditions, this study provides a methodological advance for hyperspectral monitoring of invasive grasses in subtropical grasslands and establishes a foundation for future deep-learning or temporally explicit approaches.

Conclusions

1. Spectral analysis, combined with machine-learning algorithms, enables the detection of areas infested by *Eragrostis plana* in the Pampa biome.

2. The selection of spectral bands using the Random Forest algorithm increases the discriminatory capacity of the models and reduces data dimensionality.

3. The binary classification to distinguish samples with or without *Eragrostis plana* shows a greater efficiency in identifying infested areas than the multiclass classification, due to spectral overlap among vegetation classes.

4. In scenarios with imbalanced class samples, class balancing using Adasyn improves model sensitivity in detecting *Eragrostis plana*.

5. The integration of dimensionality reduction, sample balancing, and supervised classification constitutes a promising approach for the remote monitoring of invasive species in natural pastures.

References

ABOUZAHIR, S.; SADIK, M.; SABIR, E. Enhanced approach for weeds species detection using machine vision. In: INTERNATIONAL CONFERENCE ON ELECTRONICS, CONTROL, OPTIMIZATION AND COMPUTER SCIENCE (ICECOCS), 2018, Kenitra. **Proceedings**. [S.l.]: IEEE, 2018. p.1-6. DOI: https://doi.org/10.1109/ICECOCS.2018.8610505.

AKBARZADEH, S.; AHDEROM, S.; ALAMEH, K. Application of spectral reflectance for increasing plant discrimination speed in precision agriculture. In: IEEE INTERNATIONAL CONFERENCE ON SMART CITIES: IMPROVING QUALITY OF LIFE USING ICT & IOT AND AI (HONET-ICT), 16., 2019, Charlotte. **Proceedings**. [S.l.]: IEEE, 2019. p.140-142. DOI: https://doi.org/10.1109/HONET.2019.8907994.

ALVARES, C.A.; STAPE, J.L.; SENTELHAS, P.C.; GONÇALVES, J.L. de M.; SPAROVEK, G. Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, v.22, p.711-728, 2013. DOI: https://doi.org/10.1127/0941-2948/2013/0507.

ARLOT, S.; CELISSE, A. A survey of cross-validation procedures for model selection. **Statistics Surveys**, v.4, p.40-79, 2010. DOI: https://doi.org/10.1214/09-SS054.

AZEEM, A.; JAVED, Q.; SUN, J.; DU, D. Artificial neural networking to estimate the leaf area for invasive plant *Wedelia trilobata*. **Nordic Journal of Botany**, v.38, art.02768, 2020. DOI: https://doi.org/10.1111/njb.02768.

BASINGER, N.T.; JENNINGS, K.M.; HESTIR, E.L.; MONKS, D.W.; JORDAN, D.L.; EVERMAN, W.J. Phenology affects differentiation of crop and weed species using hyperspectral remote sensing. **Weed Technology**, v.34, p.897-908, 2020. DOI: https://doi.org/10.1017/wet.2020.92.

BREUNIG, F.M.; GALVÃO, L.S.; FORMAGGIO, A.R.; EPIPHANIO, J.C.N. Classification of soybean varieties using different techniques: case study with Hyperion and sensor spectral resolution simulations. **Journal of Applied Remote Sensing**, v.5, art.053533, 2011. DOI: https://doi.org/10.1117/1.3604787.

CECHINEL, C.; CAMARGO, S.S. Mineração de dados educacionais: avaliação e interpretação de modelos de classificação. In: JAQUES, P.A.; SIQUEIRA, S.; BETHANCOURT, I.; PIMENTEL, M. (Org.). **Metodologia de Pesquisa Científica em Informática na Educação**: abordagem quantitativa. Porto Alegre: SBC, 2020. (Metodologia de Pesquisa em Informática na Educação, v.2).

CHANCHÍ GOLONDRINO, G.E.; OSPINA ALARCÓN, M.A.; SABA, M. Vegetation identification in hyperspectral images using distance/correlation metrics. **Atmosphere**, v.14, art.1148, 2023. DOI: https://doi.org/10.3390/atmos14071148.

CICCONET, N.; ALBA, J.M.F.; SPIRONELLO, R.L.; CUNHA, H.M. Mapeamento de *Eragrostis plana* Ness (capim-annoni) por meio de imagens orbitais. **Revista do Departamento de Geografia - USP**, v.29, p.20-33, 2015. DOI: https://doi.org/10.11606/rdg.v29i0.102116.

DATTA, D.; MALLICK, P.K.; REDDY, A.V.N.; MOHAMMED, M.A.; JABER, M.M.; ALGHAWLI, A.S.; AL-QANESS, M.A.A. A hybrid classification of imbalanced hyperspectral images using ADASYN and enhanced deep subsampled multi-grained cascaded forest. **Remote Sensing**, v.14, art.4853, 2022. DOI: https://doi.org/10.3390/rs14194853.

FÁVERO, L.P.; BELFIORE, P.; SOUZA, R. de F. **Data science, analytics and machine learning with R**. London: Academic Press, 2023.

FU, Y.; ZHAO, C.; WANG, J.; JIA, X.; YANG, G.; SONG, X.; FENG, H. An improved combination of spectral and spatial features for vegetation classification in hyperspectral images. **Remote Sensing**, v.9, art.261, 2017. DOI: https://doi.org/10.3390/rs9030261.

GOMES, D.F.; BREUNIG, F.M.; FONSECA, E.L. Identificação de unidades homogêneas do bioma Pampa com uso de imagens Sentinel-2: infestação do capim-annoni-2. **RA'E GA**, v.8, p.18-43, 2020. DOI: https://doi.org/10.5380/raega.v48i0.76941.

GOULART, I.C.G.R.; MEROTTO JUNIOR, A.; PEREZ, N.B.; KALSING, A. Controle de capim-annoni-2 (*Eragrostis plana*) com herbicidas pré-emergentes em associação com diferentes métodos de manejo do campo nativo. **Planta Daninha**, v.27, p.181-190, 2009. DOI: https://doi.org/10.1590/S0100-83582009000100023.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The elements of statistical learning**: data mining, inference, and prediction. 2nd ed. New York: Springer, 2009. 745p. DOI: https://doi.org/10.1007/978-0-387-84858-7.

HAUG, S.; MICHAELS, A.; BIBER, P.; OSTERMANN, J. Plant classification system for crop/weed discrimination without segmentation. In: IEEE WINTER CONFERENCE ON APPLICATIONS OF COMPUTER VISION (WACV), 2014, Steamboat Springs. **Proceedings**. [S.l.]: IEEE, 2014. p.1142-1149. DOI: https://doi.org/10.1109/WACV.2014.6835733.

LIAKOS, K.G.; BUSATO, P.; MOSHOU, D.; PEARSON, S.; BOCHTIS, D. Machine learning in agriculture: a review. **Sensors**, v.18, art.2674, 2018. DOI: https://doi.org/10.3390/s18082674.

MICHEL, T.; OVERBECK, G. Review of ecological restoration in the Brazilian Pampa. **Anais da Academia Brasileira de Ciências**, v.96, e20231283, 2024. DOI: https://doi.org/10.1590/0001-3765202420231283.

MONTOYA GONZÁLEZ, J.D.; FONSECA, E.L. da; PEREZ, N.B. Análise do potencial de invasão do capim-annoni-2 sobre áreas do bioma Pampa no município de Aceguá-RS. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 18., 2017, Santos. **Anais**. Santos: INPE, 2017. p.4197-4204.

NAJAFABADI, M.Y.; HESAMI, M.; ESKANDARI, M. Machine learning-assisted approaches in modernized plant breeding programs. **Genes**, v.14, art.777, 2023. DOI: https://doi.org/10.3390/genes14040777.

NEVES, M.C.; PEREZ, N.B.; LUIZ, A.J.B.; DEIBLER, A.; PINHO, L.B. de. Avaliação de sensor multiespectral em mapeamentos da infestação de pastagens naturais pelo capim-annoni. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 19., 2019, Santos. **Anais**. Santos: INPE, 2019. p.1-4. Available at: <https://proceedings.science/sbsr-2019/trabalhos/avaliacao-de-sensor-multiespectral-em-mapeamentos-da-infestacao-de-pastagens-nat?lang=pt-br>. Accessed on: Nov. 3 2025.

OKOYE, K.; HOSSEINI, S. **R Programming**: Statistical Data Analysis in Research. Singapore: Springer Nature Singapore, 2024. 309p. DOI: https://doi.org/10.1007/978-981-97-3385-9.

PEÑA, J.M.; TORRES-SÁNCHEZ, J.; SERRANO-PÉREZ, A.; CASTRO, A.I. de; LÓPEZ-GRANADOS, F. Quantifying efficacy and limits of unmanned aerial vehicle (UAV) technology for weed seedling detection as affected by sensor resolution. **Sensors**, v.15, p.5609-5626, 2015. DOI: https://doi.org/10.3390/s150305609.

SALEEM, M.; ASLAM, W.; LALI, M.I.U.; RAUF, H.T.; NASR, E.A. Predicting thalassemia using feature selection techniques: a comparative analysis. **Diagnostics**, v.13, art.3441, 2023. DOI: https://doi.org/10.3390/diagnostics13223441.

SILVA, S.D.P. da; EUGENIO, F.C.; FANTINEL, R.A.; AMARAL, L. de P.; SANTOS, A.R. dos; MALLMANN, C.L.; SANTOS, F.D. dos; PEREIRA, R.S.; RUOSO, R. Modeling and detection of invasive trees using UAV image and machine learning in a subtropical forest in Brazil. **Ecological Informatics**, v.74, art.101989, 2023. DOI: https://doi.org/10.1016/j.ecoinf.2023.101989.

SMUCNY, J.; SHI, G.; LESH, T.A.; CARTER, C.S.; DAVIDSON, I. Data augmentation with Mixup: enhancing performance of a functional neuroimaging-based prognostic deep learning classifier in recent onset psychosis. **NeuroImage: Clinical**, v.36, art.103214, 2022. DOI: https://doi.org/10.1016/j.nicl.2022.103214.

THENKABAIL, P.S.; LYON, J.G.; HUETE, A. **Fundamentals, sensor systems, spectral libraries, and data mining for vegetation**. 2nd ed. Boca Raton: CRC Press, 2018. DOI: https://doi.org/10.1201/9781315164151.

WATANABE, F.S.Y.; IMAI, N.N.; ALCÂNTARA, E.H.; ROTTA, L.H. da S.; UTSUMI, A.G. Signal classification of submerged aquatic vegetation based on the hemispherical-conical reflectance factor spectrum shape in the yellow and red regions. **Remote Sensing**, v.5, p.1856-1874, 2013. DOI: https://doi.org/10.3390/rs5041856.

YANO, I.H.; ALVES, J.R.; SANTIAGO, W.E.; MEDEIROS, B.J.T. Identification of weeds in sugarcane fields through images taken by UAV and random forest classifier. **IFAC-PapersOnLine**, v.49, p.415-420, 2016. DOI: https://doi.org/10.1016/j.ifacol.2016.10.076.

ZHAO, H.; MENG, J.; SHI, T.; ZHANG, X.; WANG, Y.; LUO, X.; LIN, Z.; YOU, X. Validating the crop identification capability of the spectral variance at key stages (SVKS) computed via an object self-reference combined algorithm. **Remote Sensing**, v.14, art.6390, 2022. DOI: https://doi.org/10.3390/rs14246390

Author contributions

Andre Rosa Silveira: formal analysis (lead); investigation (lead); project administration (lead); software (lead) writing - original draft (lead). **Sandro da Silva Camargo**: corresponding author; conceptualization (lead); methodology (lead); supervision (equal) writing - review & editing (equal). **Marcos Correa Neves**: validation (equal); writing - review & editing (equal). **Fabio Marcelo Breunig**: validation (equal); writing - review & editing (equal). **Naylor Bastiani Perez**: data curation (lead); funding acquisition (lead); writing - review & editing (equal).

Chief editor: Edemar Corazza
Edited by: Mírian Baptista

Data availability statement

Data available upon request: research data are only available upon reasonable request to the corresponding author.

Declaration of use of AI technologies

No generative artificial intelligence (AI) was used in this study.

Conflict of interest statement

The authors declare no conflicts of interest.

Acknowledgments

To Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), for financing, in part, this study (Finance Code 001).

Disclaimer/Publisher's note

The statements, opinions, and data contained in all texts published in Pesquisa Agropecuária Brasileira (PAB) are solely those of the individual author(s) and not of the journal's publisher, editor, and editorial team, who disclaim responsibility for any injury to people or property resulting from any referred ideas, methods, instructions, or products.

The mention of specific chemical products, machines, and commercial equipment in the texts published in this journal does not imply their recommendation by the publisher.