Recibido: 05/07/2024
Aceptado: 18/09/2024
Disponible en línea: 31/10/2024
Publicado: 31/01/2025

REVISTA DE TELEDETECCIÓN
Asociación Española de Teledetección
(2025) 65, 22060
ISSN 1133-0953
EISSN 1988-8740
https://doi.org/10.4995/raet.2025.22060

Optimizing SVM for argan tree classification using Sentinel-2 data: A case study in the Sous-Massa Region, Morocco

Abdelhak El Kharki [1*], Jamila Mechbouh [2], Miriam Wahbi [1], Otmane Yazidi Alaoui [1], Hakim Boulaassal [1], Mustapha Maatouk [1], Omar El Kharki [1]

[1] *Geomatics, Remote Sensing and Cartography Unit, FSTT, Abdelmalek Essaadi University, Tetouan, Morocco.*
[2] *RIFGeomatique ZI Tanger Maroc.*

Abstract: The development of efficient classifiers for land cover remains challenging due to the presence of hyperparameters in the model. Conventional approaches rely on manual tuning, which is both time-consuming and impractical, often leading to suboptimal results. This study aimed to optimize the hyperparameters of the Support Vector Machine (SVM) algorithm using the grid search method to map the distribution of the Argan forest in the Souss-Massa region of Morocco from Sentinel-2 satellite image. To achieve this, we examined the C parameter for the linear function, as well as the C and gamma parameters for the radial RBF and sigmoid functions. Similarly, we explored the C, gamma, and degree parameters for the polynomial function chosen using the grid search method. These parameters are compared with the default hyperparameters of each SVM function. The results are validated using the cross-validation method and by the following scores: accuracy, precision, recall, F1 score, and Cohen's Kappa. The experiments were conducted using the Earth Engine Python API in Google Colab (Google Collaboratory). In addition, experimental results indicate that the hyperparameters selected by grid search yield higher scores than the default hyperparameters. The best results were achieved using the hyperparameters of the polynomial base kernel, specifically with C = 10, degree = 2, and gamma = 10. Accuracy = 96.61%.

Key words: Sentinel-2, Support Vector Machine, hyperparameter, Argan forests, Grid Search.

Optimización de SVM para la clasificación de árboles de argán utilizando datos de Sentinel-2. Un caso de estudio en la región de Sous-Massa, Marruecos

Resumen: El desarrollo de clasificadores eficientes para la cobertura del suelo sigue siendo un desafío debido a la presencia de hiperparámetros en el modelo. Los enfoques convencionales dependen de un ajuste manual, que es tanto lento como poco práctico, lo que a menudo conduce a resultados subóptimos. Este estudio tuvo como objetivo optimizar los hiperparámetros del algoritmo de Máquina de Vectores de Soporte (SVM) utilizando el método de búsqueda en cuadrícula para mapear la distribución del bosque de Argán en la región de Souss-Massa de Marruecos a partir de una imagen de satélite Sentinel-2. Para lograr esto, examinamos el parámetro C para la función lineal, así como los parámetros C y gamma para las funciones radiales RBF y sigmoide. De manera similar, exploramos los parámetros C, gamma y grado para la función polinómica elegida utilizando el método de búsqueda en cuadrícula. Estos parámetros se comparan con los hiperparámetros predeterminados de cada

To cite this article: El Kharki, A., Mechbouh, J., Wahbi, M., Alaoui, O.Y., Boulaassal, H., Maatouk, M., El Kharki, O. 2025. Optimizing SVM for argan tree classification using Sentinel-2 data: A case study in the Sous-Massa Region, Morocco. *Revista de Teledetección*, *65*, 22060. https://doi.org/10.4995/raet.2025.22060

* *Corresponding author: abdelhak.elkharki@etu.uae.ac.ma*

función SVM. Los resultados se validan utilizando el método de validación cruzada y mediante los siguientes puntajes: exactitud, precisión, recall, puntaje F1 y Kappa de Cohen. Los experimentos se realizaron utilizando la API de Python de *Earth Engine* en *Google Colab* (*Google Collaboratory*). Además, los resultados experimentales indican que los hiperparámetros seleccionados por búsqueda en cuadrícula arrojan puntajes más altos que los hiperparámetros predeterminados. Los mejores resultados se lograron utilizando los hiperparámetros del kernel de base polinómica, específicamente con C = 10, grado = 2 y gamma = 10. Exactitud = 96.61%.

Palabras clave: Sentinel-2, Support Vector Machine, hiperparámetro, bosques de Argán, búsqueda de cuadrícula.

1. Introducción

The Argan forest, unique in the world and primarily located in the southwest of Morocco, is of paramount importance both ecologically and economically (El Ghazali *et al.*, 2021). As a unique ecosystem, it hosts rich and endemic biodiversity, with species that rely exclusively on this environment for their survival. The argan tree itself is an endemic species that plays a crucial role in preventing soil erosion and combating desertification (Moussaid & El Jaouhari, 2021). Its deep roots help stabilize the soil and retain moisture, thus contributing to the region's water regulation. Economically, the Argan forest supports local communities by providing a source of income through the production and commercialization of argan oil, known for its numerous beneficial properties and widely used in the cosmetic and food industries (Boukyoud *et al.*, 2021). The production of argan oil often involves women's cooperatives, playing a crucial role in the economic and social empowerment of rural women in the region (Gharby & Charrouf, 2022).

Conservation efforts for the Argan forest are varied and include governmental initiatives, international projects, and local programs aimed at reforestation, improving agricultural practices, and raising awareness about environmental issues (Sinsin *et al.*, 2020).

However, these efforts must be supported by a precise and continuous understanding of the forest's state, which requires effective monitoring and mapping tools. Accurate classification of the Argan forest is crucial for its conservation and sustainable management (Laarıbya *et al.*, 2021). A precise classification would enable monitoring changes in forest cover, identifying high-risk degradation areas, and formulating effective conservation strategies. However, due to its complexity and ecological diversity, classifying this forest is a challenging task.

Recently, there has been considerable interest in applying machine learning algorithms to remote sensing images for Land Use and Land Cover (LULC) mapping (Maxwell *et al.*, 2018). Many studies on land use classification (LULC) have been conducted using different machine learning algorithms. Aryal *et al.* (2023) have examined various algorithms, highlighting that each machine learning technique has different levels of accuracy. Mushtaq *et al.* (2021) compared the SVM (Support Vector Machine) and MLC (Maximum Likelihood Classification) methods for land use classification, finding that non-parametric methods like SVM outperform MLC in terms of accuracy. Rana and Suryanarayana (2020) also demonstrated that the SVM and RF (Random Forest) algorithms produce highly accurate and similar classification results, highlighting the effectiveness of these techniques for LULC classification tasks, found that SVM produces more accurate land use maps. Similarly, Oo *et al.* (2022) reported that SVM and ANN (Artificial Neural Networks) offer the best performance among six supervised classification algorithms, demonstrating the robustness of these methods. Aryal *et al.* (2023) classified land cover by comparing the RF, KNN (k-Nearest Neighbors), and SVM algorithms using Sentinel-2 imagery. They concluded that SVM provided the highest overall accuracy with the least sensitivity to training sample size, followed by the RF and KNN algorithms. Additionally, Yimer *et al.* (2024) used six classification algorithms, including SVM, Naive Bayes, C4.5, RF, Boosted Regression Tree, and KNN, concluding that SVM achieved the best classification performance. Bayas *et al.* (2022) compared Naive Bayes, KNN, SVM, tree ensembles, and artificial neural networks for land cover classification, reporting that SVM and KNN were the best methods for classifying Landsat images. Ballanti *et al.* (2016) compared the performance of SVM and RF classifiers for classifying tree species from hyperspectral images. They found that SVM outperformed RF. Shen and Cao (2017) demonstrated that the combination of hyperspectral and LiDAR data improves the accuracy of tree species classification in subtropical forests. Regarding specific work on Mediterranean trees, Sebbar *et al.* (2022) used time series of the Normalized Difference Vegetation Index (NDVI) derived from Sentinel-2 to classify and map the spatial extent of argan trees (*Argania spinosa*) in the Essaouira province of Morocco. By combining SVM classification techniques with Decision Tree models, they achieved an overall accuracy of 92.60% and an F1-score of 91.27%, demonstrating the effectiveness of this approach in overcoming spectral separability issues. El Moussaoui *et al.* (2024) integrated Sentinel-2 optical images and Sentinel-1 radar data with a Digital Elevation Model (DEM) to map argan trees in the Essaouira province. This study showed that combining multi-source data significantly improves the detection and classification of forest species, particularly using machine learning algorithms like SVM. They achieved an Overall Accuracy (OA) of up to 93.25% and a Kappa coefficient of 91.00%. In general, SVM has shown significant success in classifying remote sensing images for different land cover types, notably forests (Huang *et al.*, 2009). Although effective, SVM faces performance limits, particularly in the selection of appropriate hyperparameters. The selection of SVM hyperparameters has a major impact on the classification results' accuracy (Ahamad *et al.*, 2023).

Hyperparameter optimization for SVM often begins with a manual method, involving adjustments based on experience and trial and error. Although simple, this approach is inefficient for several reasons. It is extremely laborious and time-consuming, and the manual method is likely to miss optimal hyperparameter combinations due to its heuristic and non-systematic nature (Guido *et al.*, 2023).

Numerous studies have demonstrated the effectiveness of hyperparameter optimization in various fields. For example, Victoria and Maragatham (2021) used Bayesian optimization to improve the performance of convolutional neural networks (CNNs) in computer vision tasks, while Al Hindawi *et al*. (2021) applied this method to enhance speech recognition systems. Tian (2020) demonstrated that Grid Search (GS) can significantly improve the accuracy of CNNs in image recognition tasks.

In remote sensing, hyperparameter optimization has shown promising results in various applications. For example, Zhang *et al.* (2021) used Bayesian optimization to adjust the hyperparameters of RF classifiers for urban land cover classification using Sentinel-2 satellite images, finding significant improvements over baseline methods. While advanced methods such as Bayesian optimization, genetic algorithms, and Tree Parzen Estimators (TPE) offer significant advantages in terms of efficiency and performance, GS stands out as a systematic and widely used method for hyperparameter optimization of SVMs. This method is considered the best among those available for SVM optimization for several reasons. Firstly, it is simple and easy to implement, requiring no advanced knowledge in optimization. Secondly, it guarantees finding the best combination of hyperparameters within the provided grid by systematically covering all possible combinations in a predefined space, ensuring that the model is optimized as much as possible within the limits of that space (Belete & Huchaiah, 2021).

GS has proven effective in various fields; however, its specific application to Sentinel-2 data classification in Argan forest environments remains relatively unexplored. To bridge this gap, we utilized a dataset comprising Sentinel-2 satellite image offering detailed geospatial information on forest cover. Using this data, we conducted comprehensive evaluations of different kernel functions and their parameter effects. We validated our approach through k-fold cross-validation ($k = 5$), calculating accuracy, recall, F1 score, and kappa scores.

This article is structured as follows: the first part covers the study area, the data used, and the proposed methods. The second part presents the experimental details and results. The third section analyzes and discusses the obtained results, while the final section provides an in-depth analysis and solid conclusions regarding the effectiveness of our approach.

2. Material

2.1. Study Area

Morocco's Souss-Massa region (Hssaisoune *et al.*, 2017), with its capital Agadir, has a population of approximately 2 676 847 people, making it the ninth-largest region in the country. Known for its arid climate, heavily influenced by both the Atlantic ocean and the desert, the region's primary economic activity is agriculture. Rainfall is notably erratic, with significant spatial and temporal variations ranging between 70 and 350 mm annually, leading to substantial water resource deficiencies in both surface and groundwater.

The study area selected for this research (Figure 1) is located in southwestern Morocco, situated between 30°09'40.74" and 30°25'00.71" North and between 9°06'57.72" and 9°23'34.83" West. The land primarily consists of agriculture, greenhouses, and grassland. Additionally, a large part of the study area is covered by Argan forests. Built-up areas are often discontinuous and considerably less in proportion.

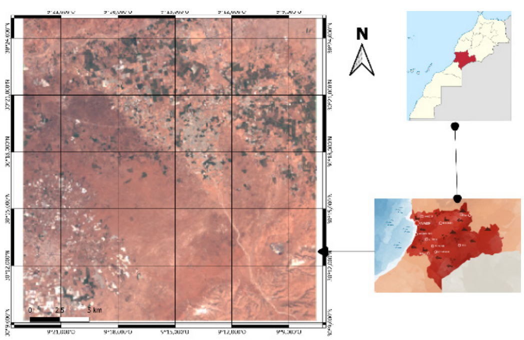


Figure 1. The study area: Arganier, Souss-Massa region.

2.2. Sentinel-2 data

Sentinel-2 (Main-Knorn *et al.*, 2017), a satellite system launched by the European Space Agency (ESA), comprises two identical satellites for Earth observation: Sentinel-2A and Sentinel-2B. Sentinel-2A was launched on June 23, 2015, and Sentinel-2B followed on March 7, 2017. These satellites provide high-resolution optical images and data collection at multiple spatial resolutions, utilizing 13 spectral bands. These bands offer a resolution of 10 meters for blue, green, red, and near-infrared-1; 20 meters for additional bands (red edge 1 to 3, near-infrared-2, shortwave infrared 1 and 2); and 60 meters for three atmospheric correction bands (aerosol, water vapor, cirrus). Sentinel-2 has a temporal resolution of ten days with one satellite and five days when both Sentinel-2A and Sentinel-2B are operational, providing frequent coverage of the Earth's land surface.

3. Methodology

The proposed method, presented in Figure 2 and applied on Sentinel 2 image over the study area, was entirely based on the use of the GEE cloud platform environment.

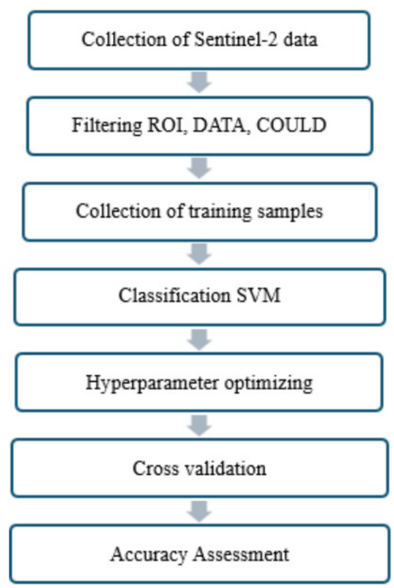


Figure 2. Proposed method.

3.1. Collection of Sentinel-2 data

The Google Earth Engine (GEE) cloud computing platform (https://earthengine.google.com) was employed to gather and process the necessary Sentinel-2 imagery for this study. GEE is an advanced cloud-based platform that facilitates the analysis and visualization of geospatial data. GEE provides an extensive archive of satellite imagery and geospatial datasets. For this study, Sentinel-2A/B Level 2A products were utilized. These products are part of the European Space Agency's

Copernicus program, which provides free and open access to a wide range of environmental data. Sentinel-2 image contain 13 spectral bands, but we focused on bands 2 (blue), 3 (green), 4 (red), and 8 (NIR) for this study because they are particularly effective for classification of the vegetation. The image used is dated September 30, 2021, covering the Souss-Massa region in Morocco. This unique image was chosen for its excellent quality and temporal relevance for our study.

3.2. Preprocessing

Sentinel-2 images, provided by the European Space Agency's Copernicus program, are available in two processing levels: L1C (Top of Atmosphere) and L2A (Bottom of Atmosphere). We opted for L2A images due to their atmospheric correction, making them more suitable for accurate land cover analysis. Only orthorectified images with minimal cloud cover were selected, and pixels affected by clouds were removed using the cloud masking algorithm available on the Google Earth Engine (GEE) platform. To fill gaps caused by cloud cover, temporal aggregation methods such as median, mean, and minimum/maximum values were applied to ensure data quality and consistency.

3.3. Collection of training samples

After the data acquisition stage, we proceeded to the data preparation and feature extraction stage, where we defined six land cover classes described in Table 1.

Training and validation datasets were generated through image observations. In total, 900 samples were used for classification (150 samples for each class). Generally speaking, note that we split the training data into training (70%) and validation (30%) sets to evaluate model performance during training and avoid overfitting.

3.4. SVM

The Support Vector Machines (SVM) classifier is a powerful supervised classification method rooted in statistical learning theory. Although initially developed in the late 1970s, its widespread adoption in remote sensing gained momentum approximately ten years ago (Joshi *et al.*, 2016). Extensive mathematical explanations and the fundamental theory behind SVM have been thoroughly showcased in numerous prior studies (Kurani *et al.*, 2023). SVM's rise in remote sensing reflects its efficacy and relevance in handling complex classification tasks within this field. Although SVM is indeed a classic classifier, it can handle complex datasets using kernel functions. These functions transform the data into a higher-dimensional feature space, facilitating the creation of nonlinear decision boundaries. This method involves discovering the best boundary between different groups by pinpointing key training samples known as support vectors. When the training data cannot be separated linearly, a kernel method is employed to transform the data into a higher-dimensional space, making it possible to separate the groups linearly.

3.5. Hyperparameter of SVM

Table 2 provides a concise description of the key hyperparameters used in (SVM) classification. Mastering these hyperparameters is essential for correctly configuring the SVM model to maximize classification accuracy.

Choosing the appropriate kernel and tuning these hyperparameters is critical to creating an effective SVM model (Czarnecki *et al.*, 2015). Inadequate choices or parameter values can lead to poor classification results. Grid search is a common method used for hyperparameter optimization in SVMs (Yao *et al.*, 2021). It systematically tests different combinations of hyperparameters by evaluating

Table 1. land cover classes.

Class names	Description
Dense Argan forest	Refers to areas where trees are large and dense, typically associated with a high density of the argan tree.
Built-up	Includes buildings, houses, and roads, representing developed or urban areas.
Agricultural	Refers to areas with dense vegetation or arable land, used for agricultural purposes.
Sparse Argan forest	Corresponds to areas where argan trees are small and scattered, indicating low tree density.
Bare soil	Refers to non-arable land devoid of vegetation.
Greenhouse	Covers the surface areas of greenhouses, generally used for cultivation under cover.

Table 2. Overview of SVM Hyperparameters.

Hyperparameter	Description
Kernel	Type of kernel function (e.g., linear, polynomial, RBF) used to transform the data.
C	Regularization parameter balancing margin size and training errors.
Gamma	Determines the influence range of each training point.
Degree	Polynomial kernel parameter specifying polynomial degree.

each unique combination using cross-validation to prevent overfitting. The name «grid search» arises from its representation as a grid-like structure where each point on the grid signifies a unique hyperparameter combination.

This method allows the exploration of various hyperparameter values to find the best configuration for the SVM model in terms of both accuracy and speed. It is an efficient but powerful technique in hyperparameter optimization.

To utilize GS, we must establish a range of specific values for each hyperparameter to be explored throughout the search process. This involves defining these ranges or values for each hyperparameter using a minimum value, a maximum value, and a step. Three different scales can be employed: linear, quadratic, and logarithmic scales.

One of the primary challenges in optimizing SVM parameters involves the absence of defined value ranges for C and gamma. A broader parameter range is often favored as it enhances the GS method's likelihood of discovering the most optimal combination. The ranges of C and gamma values are presented in Table 3 (Mantovani *et al.*, 2015).

We employed k-fold cross-validation (Wong & Yeh, 2019), dividing the data into 5 subsets when K = 5. Among these subsets, one was allocated as the test set, while the remaining four were utilized as the training set. Various metrics, such as accuracy, recall, F1 score, and kappa, were computed for each combination to pinpoint the set of hyperparameters that yielded the best results. The identified parameters were then employed in crafting SVM models, which were subsequently applied to the Sentinel image spanning the study area for the generation of the Land Cover Land Use (LCLU) Map, specifically for mapping the distribution of the Argan forest.

Table 3. The ranges of C and gamma values.

Hyper-parameter	Minimal	Maximal
cost (C)	10^{-2}	10^{15}
gamma	10^{-15}	10^{-2}

To generate training and validation points, we utilized Google Earth Engine (GEE), while Google Colab was employed to execute the classification task and generate the map.

3.6. Metrics

Various metrics can be used to validate a classification model based on the problem's characteristics and available data:

- Accuracy: The percentage of correct predictions out of total predictions.
- Precision: The proportion of true positives out of all positive predictions.
- Recall (Sensitivity): The percentage of true positives out of all actual positive instances.
- F1 Score: The harmonic mean of precision and recall, reflecting both false positives and false negatives.
- Cohen's Kappa: Measures agreement between predictions and ground truth, adjusting for random chance.
- Rate of Increase: The relative growth rate between two values, calculated as the difference divided by the original value.
- Confusion Matrix: Summarizes classification performance and supports the calculation of metrics like precision and recall.

4. Results

Initially, we executed Support Vector Machines (SVMs) with two distinct configurations: SVM using default parameters and SVM employing GS. The default hyperparameters of the scikit-learn library SVM model installed in Google Colab are as follows: {'C': 1.0, 'degree': 3, 'gamma': 'scale', 'kernel': 'rbf'}. In the first configuration, we applied the datasets using default parameters and altered the kernels to four different types: linear, RBF, sigmoid, and polynomial. This approach allowed us to compare the performance of the default parameters across different kernel functions.

For the second configuration, we applied the GS method across four distinct experiments encompassing linear, RBF, sigmoid, and polynomial functions. We examined the range of variation for C and gamma on a logarithmic scale spanning from -2 to 15 for C and from -15 to 2 for gamma (Table 3). For the polynomial kernel, there is a kernel parameter called the degree of the polynomial function. We explored the degree's variation within the range of 1 to 3.

Figure 3 displays a comparison among the four kernels, showcasing the optimal parameters C and γ alongside their respective accuracy rates. Each point on the grid represents the estimated accuracy of the classifier. White points indicate higher

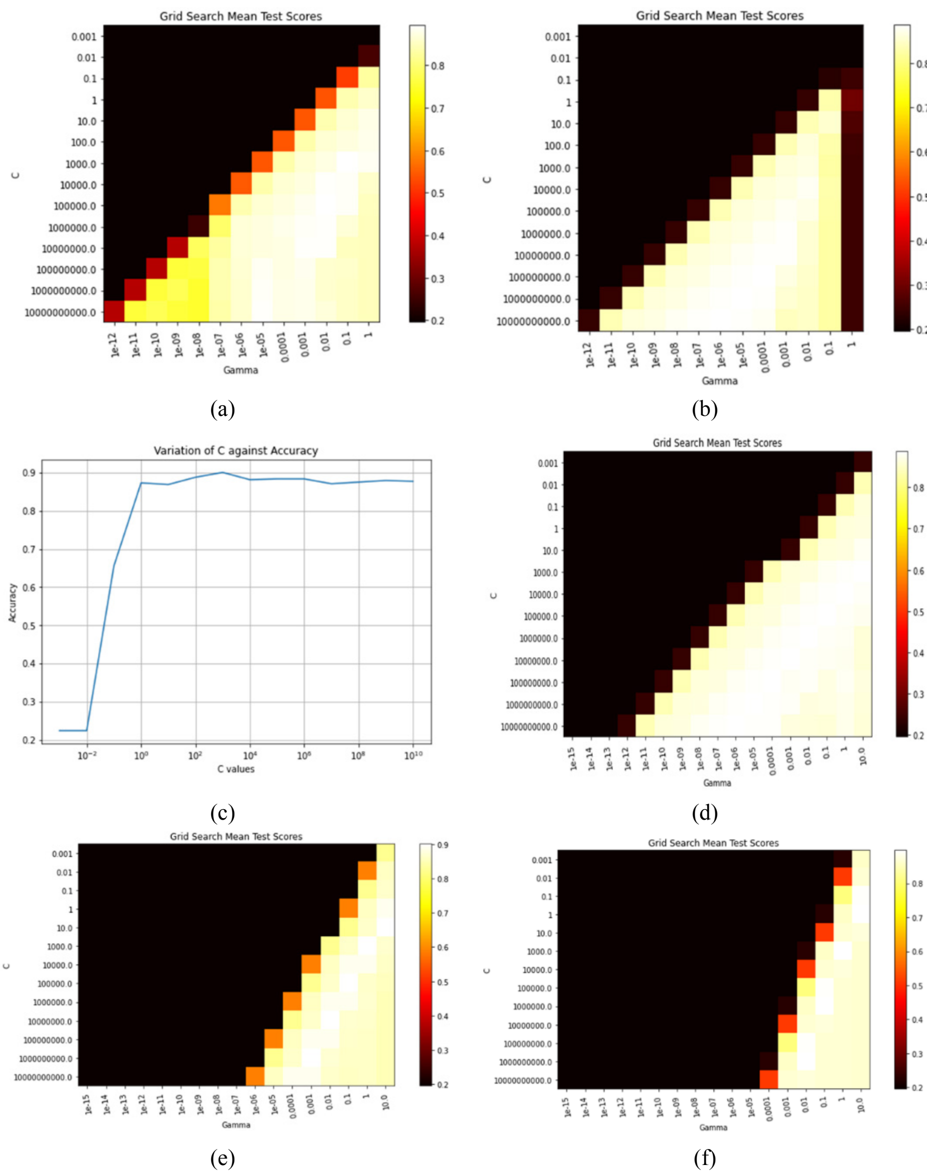


Figure 3. Results obtained by comparing the four kernels: (a) RBF, (b) Sigmoid, (c) Linear, (d) Polynomial degree 1, (e) Polynomial degree 2, (f) Polynomial degree 3.

predictive accuracy, while black points signify lower accuracy. Across the four kernels, the impact of different combinations of (C, γ) is notably significant. Results indicate varying accuracy levels among the kernels, ranging from 90.91% to 96.61%. The linear kernel yields the lowest accuracy, while the Polynomial kernel, specifically with parameters (degree = 2, C = 10, γ = 10), achieves the highest accuracy, as highlighted in Table 4.

Table 4. Provides the values of the optimal hyperparameters for each kernel.

	Linear	RBF	Sigmoid	Polynomial
C	10^3	10^5	10^8	10^3
Gamma	-	10^{-2}	10^{-4}	10
Degree	-	-	-	2
Accuracy (%)	90.91	93.65	93.15	96.61

After testing the performance of the trained SVMs with different kernel functions and verifying their effectiveness, the best SVM models were compared to the default models. The results are presented in Table 5 and Figure 4.

The SVM parameter optimization using GS provides a significant improvement for all kernels compared to the default parameters. To effectively illustrate the results, we calculated the rate of increase between the default values (D) and the grid search (GS). The results are presented in Table 6.

Table 6. The rate of increase between the default values (D) and the grid search (GS).

	Linear	RBF	Sigmoid	Polynomial
Accuracy (%)	1.80	4.24	457.49	4.82
Precision (%)	1.80	3.72	3276.98	4.59
Recall (%)	1.80	4.24	457.49	4.82
F1 score (%)	1.86	4.33	1843.93	4.84
Cohen's Kappa (%)	2.35	5.22	-	5.88

The results indicate that the SVM's performance improved significantly across all metrics, particularly in the sigmoid kernel, where default parameters yielded very low scores, but GS optimization led to significant improvements. For further analysis, we generated a confusion matrix for the SVM classifier using the polynomial kernel Table 7.

Table 5. Results obtained from comparing the performance measurements of the. parameters of each default kernel (D) with those optimized via grid search (GS).

	Linear		RBF		Sigmoid		Polynomial (d = 2)	
	D	GS	D	GS	D	GS	D	GS
Accuracy (%)	89.21	90.91	89.85	93.65	16.70	93.15	92.17	96.61
Precision (%)	89.02	90.72	90.30	93.66	2.78	93.79	92.40	96.66
Recall (%)	89.21	90.91	89.85	93.65	16.70	93.15	92.17	96.61
F1 score (%)	89.05	90.71	89.69	93.58	4.78	93.07	92.14	96.59
Kappa (%)	87.04	89.08	87.80	92.38	0.0	91.78	90.60	95.93

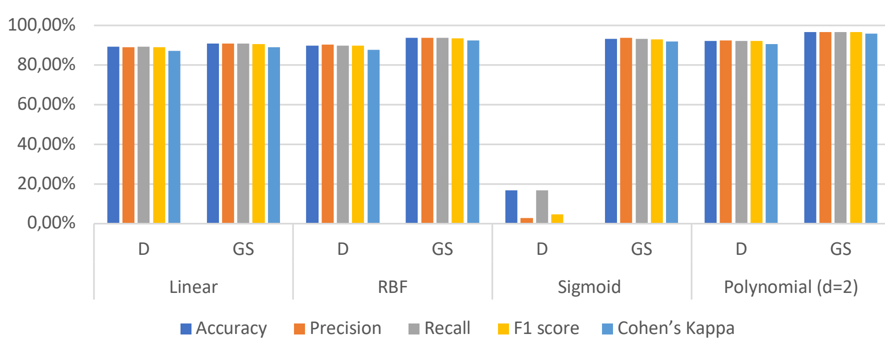


Figure 4. the performance measurements of the parameters of each default kernel (D) with those optimized via grid search (GS).

Table 7. Confusion matrix for SVM classifier with Polynomial kernel.

		Ground truth (%)					
		Dense Argan	Built-up	Agriculture	Sparse Argan	Bare Soil	Greenhouse
Classes	Dense Argan	100.00	0.00	0.00	0.00	0.00	0.00
	Built-up	0.00	97.30	0.00	0.00	10.13	0.00
	Agriculture	0.00	0.00	100.00	0.00	0.00	0.00
	Sparse Argan	0.00	0.00	0.00	97.44	3.80	0.00
	Bare Soil	0.00	2.70	0.00	2.56	86.08	1.41
	Greenhouse	0.00	0.00	0.00	0.00	0.00	98.59
	Total	100	100	100	100	100	100
	Overall accuracy (%)	96.61					

The classifier achieved high overall accuracy, especially in distinguishing classes like Dense Argan, Agriculture, and Greenhouse. However, some confusion was observed between Bare Soil and Built-up classes, which could be due to their spectral similarities.

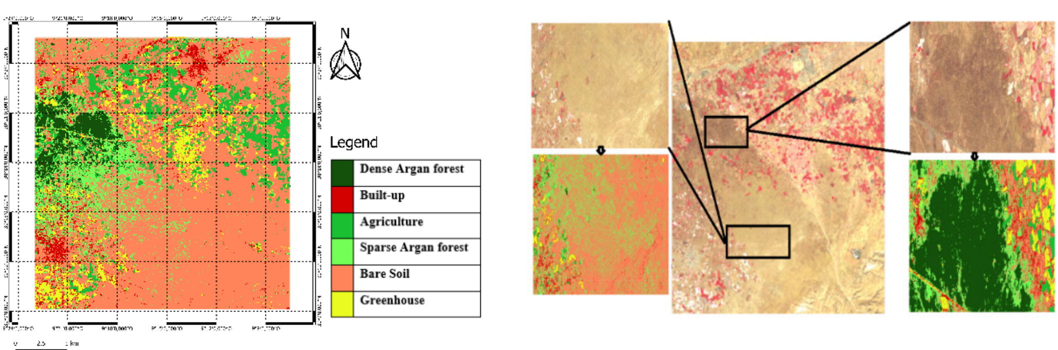
In addition to the classification map (Figure 5.a), we zoomed in on two regions, Sparse Argan and Dense Argan, to visualize the spatial distribution of these classes more clearly (Figure 5.b).

5. Discussion

The results indicate that GS optimization significantly enhances the performance of SVM classifiers across different kernel functions. For the linear kernel, modest yet consistent improvements in accuracy, precision, recall, F1 score, and Cohen's Kappa were observed, suggesting that even simple models benefit from hyperparameter tuning.

The RBF kernel exhibited more substantial gains, with improvements in precision, accuracy, and recall surpassing 4%. This indicates the sensitivity of the RBF kernel to hyperparameter adjustments and the value of GS in identifying optimal configurations. The most dramatic enhancements were seen in the sigmoid kernel, which initially performed poorly with default parameters but showed remarkable improvements after GS. This suggests that the sigmoid kernel's performance is highly dependent on the correct setting of parameters, underscoring the importance of optimization in scenarios where default settings lead to suboptimal results.

The polynomial kernel, despite already having a relatively high default performance, benefited from GS as well. Fine-tuning resulted in a noticeable accuracy increase, emphasizing that even well-performing models can achieve incremental gains through careful hyperparameter optimization.



(a) (b)

Figure 5. (a) Thematic map derived from the classification of Sentinel-2 image using a polynomial kernel and (b) Zoom on the regions Dense and Sparse Argan Forest.

The confusion matrix revealed areas of class misclassification, primarily due to spectral similarities between certain classes. The Built-up and Bare Soil classes exhibited significant overlap, which can be attributed to the spectral characteristics of rural areas with sparse construction. Similarly, Sparse Argan was often confused with Bare Soil, reflecting the challenge of distinguishing between sparse vegetation and exposed soil surfaces. This confusion is particularly evident in the fragmented forest regions of southwest Morocco, where the spectral signature of the argan tree can be influenced by the underlying soil.

The thematic map and its zoomed-in sections provided a detailed visualization of the classification results. The classifier effectively delineated different land cover classes, showcasing its ability to capture spatial distribution accurately.

These outcomes underscore the importance of hyperparameter tuning in achieving superior model efficacy in machine learning and remote sensing applications. The findings also highlight the potential of SVMs, particularly with polynomial kernels, in land cover classification tasks. Future research could explore the integration of additional spectral indices and texture features to further enhance classification accuracy, especially in distinguishing between spectrally similar classes.

6. Conclusion

This study addressed the crucial challenge of optimizing support vector machine (SVM) hyperparameters to ensure accurate classification of the valuable Argan forest in the Sous-Massa region of Morocco using Sentinel-2 satellite imagery data. It involved a comparison between default SVM hyperparameters and those optimized using the Grid Search method. Through experimentation involving different kernel functions (linear, RBF, sigmoid, and polynomial) and their respective parameters (C, gamma, and degree), it becames apparent that the Grid Search method yielded significantly better results than the default parameters.

Comparative analyses revealed notable improvements in all evaluated metrics (accuracy, precision, recall, F1 score, and Cohen's Kappa). The remarkable improvement in performance for the sigmoid function is particularly noteworthy, showing a significant increase despite initially mediocre performance. The most remarkable results were obtained with the basic polynomial kernel, specifically with hyperparameters set to C = 10, degree = 2, and gamma = 10, demonstrating superior performance. This configuration achieved an accuracy of 96.61%, precision of 96.66%, recall of 96.61%, F1 score of 96.59%, and Cohen's Kappa of 95.93%.

This research highlights the vital significance of hyperparameter optimization in machine learning models. It presents a guide for researchers and practitioners to enhance the accuracy and reliability of classification algorithms in similar remote sensing applications. The optimized SVM model serves as a robust tool for effectively monitoring the Argan forest in the Sous-Massa region and beyond.

The improved precision in mapping helps identify areas of high argan tree density and degraded zones needing intervention. These accurate maps can aid forest managers in planning conservation actions like reforestation and protecting critical areas. Additionally, distinguishing between sparse argan three areas and bare soil facilitates the prioritization of restoration efforts, enhancing ecosystem resilience.

Although based on a single image, our methodology can be applied to time series of images to monitor changes over time, evaluate conservation efforts, and adapt management strategies. The use of remote sensing imagery, especially Sentinel-2, proves effective for large-scale, continuous forest monitoring at reduced costs. Our tools and techniques can be transferred to other regions and vegetation types, promoting sustainable forest management globally.

Furthermore, the study's findings can inform policymakers about the state of Argan forests and conservation needs, supporting data-driven decisions that garner local community support. Effective forest management also has socio-economic benefits, as Argan forests provide essential resources like argan oil, contributing to local livelihoods and sustainable regional development.

7. Acknowledgements

We would like to thank the European Space Agency (ESA) for providing the Sentinel-2 images free of charge.

References

Ahamad, G.N., Shafiullah, Fatima, H., Imdadullah, Zakariya, S., Abbas, M., Alqahtani, M.S., Usman, M. 2023. Influence of optimal hyperparameters on the performance of machine learning algorithms for predicting heart disease. *Processes*, *11*(3), 734. https://doi.org/10.3390/pr11030734

Al Hindawi, N.A., Shahin, I., Nassif, A.B. 2021. Speaker identification for disguised voices based on modified SVM classifier. *18th International Multi-Conference on Systems, Signals & Devices (SSD),* Monastir, Tunisia, 2021, pp. 687-691, https://doi.org/10.1109/SSD52085.2021.9429403

Aryal, J., Sitaula, C., Frery, A.C. 2023. Land use and land cover (LULC) performance modeling using machine learning algorithms: A case study of the city of Melbourne, Australia. *Scientific Reports*, *13*(1), 13510. https://doi.org/10.1038/s41598-023-40564-0

Ballanti, L., Blesius, L., Hines, E., Kruse, B. 2016. Tree species classification using hyperspectral imagery: A comparison of two classifiers. *Remote Sensing*, *8*(6), 445. https://doi.org/10.3390/rs8060445

Bayas, S., Sawant, S., Dhondge, I., Kankal, P., Joshi, A. 2022. Land use land cover classification using different ml algorithms on sentinel-2 imagery. In *Advanced machine intelligence and signal processing* (pp. 761–777). Springer. https://doi.org/10.1007/978-981-19-0840-8_59

Belete, D.M., Huchaiah, M.D. 2021. Grid search in hyperparameter optimization of machine learning models for prediction of HIV/AIDS test results. *International Journal of Computers and Applications, 44*(9), 875-886. https://doi.org/10.1080/1206212X.2021.1974663

Boukyoud, Z., Ibourki, M., Gharby, S., Sakar, E.H., Bijla, L., Atifi, H., Matthäus, B., Laknifli, A., Charrouf, Z. 2021. Can the water quality influence the chemical composition, sensory properties, and oxidative stability of traditionally extracted argan oil? *Mediterranean Journal of Nutrition and Metabolism*, *14*(4), 383–399. https://doi.org/10.3233/MNM-210005

Czarnecki, W.M., Podlewska, S., Bojarski, A.J. 2015. Robust optimization of SVM hyperparameters in the classification of bioactive compounds. *Journal of Cheminformatics*, *7*(1), 1–15. https://doi.org/10.1186/s13321-015-0088-0

El Ghazali, H., Harrouni, C., Daoud, S., Tlemçani, N.B. 2021. *Impact of Climate Change on the Argan Biosphere Reserve (ABR) in Morocco*. https://doi.org/10.20944/preprints202105.0536.v1

El Moussaoui, E., Moumni, A., Lahrouni, A. 2024. Assessing the influence of different Synthetic Aperture Radar parameters and Digital Elevation Model layers combined with optical data on the identification of argan forest in Essaouira region, Morocco. *iForest-Biogeosciences and Forestry*, *17*(2), 100. https://doi.org/10.3832/ifor4183-016

Gharby, S., Charrouf, Z. 2022. Argan oil: Chemical composition, extraction process, and quality control. *Frontiers in Nutrition*, *8*, 804587. https://doi.org/10.3389/fnut.2021.804587

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R. 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. https://doi.org/10.1016/J.RSE.2017.06.031

Guido, R., Groccia, M.C., Conforti, D. 2023. A hyperparameter tuning approach for cost-sensitive support vector machine classifiers. *Soft Computing*, *27*(18), 12863–12881. https://doi.org/10.1007/s00500-022-06768-8

Hssaisoune, M., Boutaleb, S., Benssaou, M., Bouaakkaz, B., Bouchaou, L. 2017. Physical geography, geology, and water resource availability of the Souss-Massa River Basin. *The Souss-Massa River Basin, Morocco*, 27–56. https://doi.org/10.1007/698_2016_68

Huang, X., Zhang, L., Wang, L. 2009. Evaluation of morphological texture features for mangrove forest mapping and species discrimination using multispectral IKONOS imagery. *IEEE Geoscience and Remote Sensing Letters*, *6*(3), 393–397. https://doi.org/10.1109/LGRS.2009.2014398

Joshi, N., Baumann, M., Ehammer, A., Fensholt, R., Grogan, K., Hostert, P., Jepsen, M.R., Kuemmerle, T., Meyfroidt, P., Mitchard, E.T.A., Reiche, J., Ryan, C.M., Waske, B. 2016. A review of the application of optical and radar remote sensing data fusion to land use mapping and monitoring. *Remote Sensing*, *8*(1), 1–23. https://doi.org/10.3390/rs8010070

Kurani, A., Doshi, P., Vakharia, A., Shah, M. 2023. A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting. *Annals of Data Science*, *10*(1), 183–208. https://doi.org/10.1007/s40745-021-00344-x

Laarıbya, S., Alaouı, A., Ayan, S., Benabou, A. 2021. Spatial Analysis of the Degraded Forest Areas in Idmine Forest-Morocco Using Geoscience Capabilities. *Kastamonu University Journal of Forestry Faculty*, *21*(1), 1–11. https://doi.org/10.17475/kastorman.908568

Main-Knorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U., Gascon, F. 2017. Sen2Cor for sentinel-2. *Image and Signal Processing for Remote Sensing XXIII*, *10427*, 37–48. https://doi.org/10.1117/12.2278218

Mantovani, R.G., Rossi, A.L.D., Vanschoren, J., Bischl, B., Carvalho, A.C.P.L.F. De. 2015. Effectiveness of Random Search in SVM hyper-parameter tuning. *Proceedings of the International Joint Conference on Neural Networks*, *2015-Septe*. https://doi.org/10.1109/IJCNN.2015.7280664

Maxwell, A.E., Warner, T.A., Fang, F. 2018. Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing*, *39*(9), 2784–2817. https://doi.org/10.1080/01431161.2018.1433343

Moussaid, R., El Jaouhari, S. 2021. L'impact de l'entrepreneuriat collectif sur le développement territorial: L'arganeraie du Sud-Ouest du Maroc comme modèle. *Revue Internationale Des Sciences de Gestion*, *4*(1).

Mushtaq, F., Mahmood, K., Hamid, M.C., Tufail, R. 2021. A Comparative Study of Support Vector Machine and Maximum Likelihood Classification to Extract Land Cover of Lahore District, Punjab, Pakistan. *Pakistan Journal of Scientific & Industrial Research Series A: Physical Sciences*, *64*(3), 265–274. https://doi.org/10.52763/PJSIR.PHYS.SCI.64.3.2021.265.274

Oo, T.K., Arunrat, N., Sereenonchai, S., Ussawarujikulchai, A., Chareonwong, U., Nutmagul, W. 2022. Comparing four machine learning algorithms for land cover classification in gold mining: A case study of Kyaukpahto Gold Mine, Northern Myanmar. *Sustainability*, *14*(17), 10754. https://doi.org/10.3390/su141710754

Rana, V.K., Suryanarayana, T.M.V. 2020. Performance evaluation of MLE, RF and SVM classification algorithms for watershed scale land use/land cover mapping using sentinel 2 bands. *Remote Sensing Applications: Society and Environment*, *19*, 100351. https://doi.org/10.1016/j.rsase.2020.100351

Sebbar, B., Moumni, A., Lahrouni, A., Chehbouni, A., Belghazi, T., Maksoudi, B. 2022. Remotely sensed phenology monitoring and Land-cover classification for the localization of the endemic argan tree in the southern-west of Morocco. *Journal of Sustainable Forestry*, *41*(10), 1014–1028. https://doi.org/10.1080/10549811.2021.1897018

Shen, X., Cao, L. 2017. Tree-species classification in subtropical forests using airborne hyperspectral and LiDAR data. *Remote Sensing*, *9*(11), 1180. https://doi.org/10.3390/rs9111180

Sinsin, T.E., Mounir, F., El Aboudi, A. 2020. Conservation, valuation and sustainable development issues of the Argan Tree Biosphere Reserve in Morocco. *Environmental & Socio-Economic Studies*, *8*(1), 28–35. https://doi.org/10.2478/environ-2020-0004

Tian, Y. 2020. Artificial intelligence image recognition method based on convolutional neural network algorithm. *Ieee Access*, *8*, 125731–125744. https://doi.org/10.1109/ACCESS.2020.3006097

Victoria, A.H., Maragatham, G. 2021. Automatic tuning of hyperparameters using Bayesian optimization. *Evolving Systems*, *12*(1), 217–223. https://doi.org/10.1007/s12530-020-09345-2

Wong, T.-T., Yeh, P.-Y. 2019. Reliable accuracy estimates from k-fold cross validation. *IEEE Transactions on Knowledge and Data Engineering*, *32*(8), 1586–1594. https://doi.org/10.1109/TKDE.2019.2912815

Yao, L., Fang, Z., Xiao, Y., Hou, J., Fu, Z. 2021. An intelligent fault diagnosis method for lithium battery systems based on grid search support vector machine. *Energy*, *214*, 118866. https://doi.org/10.1016/j.energy.2020.118866

Yimer, S.M., Bouanani, A., Kumar, N., Tischbein, B., Borgemeister, C. 2024. Comparison of different machine-learning algorithms for land use land cover mapping in a heterogenous landscape over the Eastern Nile river basin, Ethiopia. *Advances in Space Research*, *74*(5), 2180–2199. https://doi.org/10.1016/j.asr.2024.06.010

Zhang, T., Su, J., Xu, Z., Luo, Y., Li, J. 2021. Sentinel-2 satellite imagery for urban land cover classification by optimized random forest classifier. *Applied Sciences*, *11*(2), 543. https://doi.org/10.3390/app11020543