Bulletin of Geography. Physical Geography Series, No. 25 (2023): 63–78
http://doi.org/10.12775/bgeo-2023-0010

Evaluation of machine learning algorithms for forest species mapping based on Sentinel 2 data: a case study of Ait Bouzid forest (Central High Atlas, Morocco)

Ahmed Barakat* [1,a], **Mariem Brhaiberh**[1], **Abderrahim Ettaqy**[2], **Aafaf El Jazouli**[1], **Widad Ennaji**[1]

[1]Sultan My Slimane University, Geomatics, Georesources and Environment Laboratory, Faculty of Sciences and Techniques, Béni-Mellal, Morocco
[2]Regional Directorate of Waters and Forests, Béni-Mellal, Morocco

*E-mail: a.barakat@usms.ma
[a]https://orcid.org/0000-0003-3500-6157

Abstract. In arid and semi-arid environments, producing accurate maps of forest tree cover using optical remote-sensing data is essential to understand their spatial distributions and dynamics. In this respect, the current study aimed to explore the effectiveness of support vector machine (SVM), K nearest neighbors (KNN) and random forest (RF) machine learning (ML) algorithms to map the forest tree species of Ait Bouzid region (Central High Atlas, Morocco) based on Sentinel-2A data.
The results from all models showed that about 19–28%, 21–27%, 16–24%, 15–18% and 0.3–0.32% of the area was covered by euphorbia, red juniper, cedar, holm oak, bare ground, and water body, respectively. According to the overall accuracy (OA) and kappa coefficient, the SVM classifier showed the highest OA (73%) and kappa (0.66) values, followed by KNN (OA=70%, kappa=0.62) and RF (OA=67%, kappa=0.59). Regarding LC classes, water, bare soil and holm oak could be identified with the producer's accuracy attaining 100%. At the same time, red juniper and cedar were the most challenging classes to determine for all ML classifiers, with the producer's accuracy of 40–50% and 40–67%. This study revealed the potential of ML approaches coupled with multispectral Sentinel-2A data for forest species cartography with high accuracy in arid areas. Furthermore, it provides crucial information about forest tree species distribution for developing forest management plans.

Key words:
Ait Bouzid forest,
forest trees,
sentinel 2 image,
ML classifiers,
mapping
Morocco

Introduction

Forest sustains the well-being of many critical terrestrial ecosystems. It provides essential environmental services because it contains more than half of global biodiversity and of the world's carbon stock in soils and vegetation. However, the non-optimized production of forest ecosystem goods and services threatens the planet's forests, potentially driving deforestation and forest degradation (FAO 2023). Therefore, ever-increasing human activities, industrialization and urbanization combined with global climate changes have caused detrimental effects on the forest reserves and threatened their sustainability.

Forests in Morocco covers about 13.5% of the national territory and represents a sector of great economic and social importance (Ministry of Energy Transition and Sustainable Development, Morocco 2023). It represents a most important natural resource for the people who live here and a source of economic activity. The Moroccan forest contributes about 2% of the Gross domestic product (GDP) and 0.4% of the national GDP (Ministry of Energy Transition and Sustainable Development, Morocco 2023).

However, this natural heritage is threatened with disappearance due to climatic hazards (desertification) and human intervention (uncontrolled exploitation of wood, evolution of agriculture, overgrazing and fires) (Barakat et al. 2018; Bouzekraoui et al. 2016; Ettaqy et al. 2020). Population growth in a mountainous region with limited agricultural land has reduced forest density. Subsistence farming, overgrazing and livestock practices, driven by a lack of alternatives, contribute to deforestation (Boubekraoui et al. 2023; Ziadi et al. 2023). Thus, providing the long-term health of Morocco's forest stands under changing climate while providing social, economic and environmental benefits to local communities needs efficient and sustainable planning and management of forest resources using modern techniques. Therefore, the National Agency for Water and Forests (ANEF), Morocco, has drawn up a national strategy and action plan to safeguard the forest and stop the process of its degradation without depriving residents of their livelihoods (ANEF 2023).

In recent decades, digital modeling and mapping of forest canopy have emerged as promising tools that help planners propose strategies for achieving sustainable forest management. Remote Sensing (RS), Geographical Information Systems (GIS) and statistical techniques are the common tools applied in forest stand mapping (Barakat et al. 2019; Soleimannejad et al. 2019a; Nasiri et al. 2023) as they are cost-efficient and timesaving. Remote sensing provides valuable information for forest mapping and inventories on temporal and spatial scales. It makes it possible to precisely map forest stands, their types and their distribution and changes (Decuyper et al. 2022; Grabska et al. 2020; Nasiri et al. 2023). GIS have also become crucial in forest studies for effectively managing and analyzing extensive geospatial data collected from remote sensing and field observations (Mickelson et al. 1998; Appeaning Addo 2010; Fassnacht et al. 2016). Moreover, the geostatistical probabilistic models incorporated with GIS have been used for processing the complex geospatial dataset to assess the status of the vulnerability of forests around the world (Wilson et al. 2005; Gülci 2014; Tuček et al. 2014). Recently, machine learning (ML) has been successfully applied to process the high-dimensional and complex geospatial dataset collected from remote-sensing imagery and field measurements for predicting a hazard and risk in environmental sciences (Sajedi-Hosseini et al. 2018; Vega Isuhuaylas et al. 2018; Mishra et al. 2021; Barakat et al. 2022). These approaches provided multivariate, nonparametric and nonlinear data classifications (Lary et al. 2016; Maxwell et al. 2018). For instance, the most popular ML algorithms used in forest studies are Random Forest (RF) (Cheng and Wang 2019; Soleimannejad et al. 2019b; Grabska et al. 2020; Sothe et al. 2020) and Support Vector Machines (SVM) (Grabska et al. 2020; Radhakrishnan et al. 2020; Sothe et al. 2020; Zagajewski et al. 2021). Thus, all these ML models have proven to identify and map contiguous forest formations in most further studies. However, simple classifiers such as the K-Nearest Neighbor (KNN), which performs better than the SVM or the RF (Ge et al. 2018), receive less attention. Therefore, a sensitivity comparison of the prediction accuracies among the model (SVM, KNN and RF) results is missing and needs to be well analyzed. In this respect, this study attempted to compare SVM, KNN and RF models to indicate the type and density of forest trees.

Given the above, the present study aimed to evaluate the potential of multispectral Sentinel-2 MSI data with ML models (SVM, KNN and RF) for mapping forest stand species in the Ait Bouzid forest, Atlas of Afourer, Morocco. A confusion matrix was applied to the classified images to evaluate the efficiency of the ML models and validate the maps generated from each model. The provided information is precious and will serve as a foundation to direct and enhance the effectiveness of conservation initiatives, restoration of natural ecosystems and land-use planning. Also, the employed research methodology can serve as a model for studying other similar areas.

Materials and methods

Study area

The forest of Ait Bouzid is located in the High Atlas of Afourer, in the center of Morocco. Bounded south by Oued El Abid, a main river supplying the lake of the Bine El Ouidane dam, and north by the Tadla plain, the study forest covers a surface area of 16,477 ha. It lies between 32°07’ and 32°14’ N latitudes and 6°19’ and 6°40’ W longitudes (Fig. 1). It is characterized by a dominance of medium and high mountain landscapes with elevation varying from 695 to 2412 m. Average annual precipitation varies between 200 mm and 600 mm, with minimal rainfall occurring in August, and a yearly temperature ranges between 3°C and 40°C, with the minimum temperature in January. The forest comprises tree-dominant species, including holm oak (*Quercus rotundifolia*), barberry cedar (*Tetraclinis articulata*), and red juniper (*Juniperus turbinata*), often in forest or pre-forest formations (Taïbi et al. 2015). Other plant species are common but often in association with holm oak or alone in particular reliefs (high mountains, very steep slopes, escarpments or valley bottoms) (Bouzekraoui et al. 2016; Barakat et al. 2018).

Data

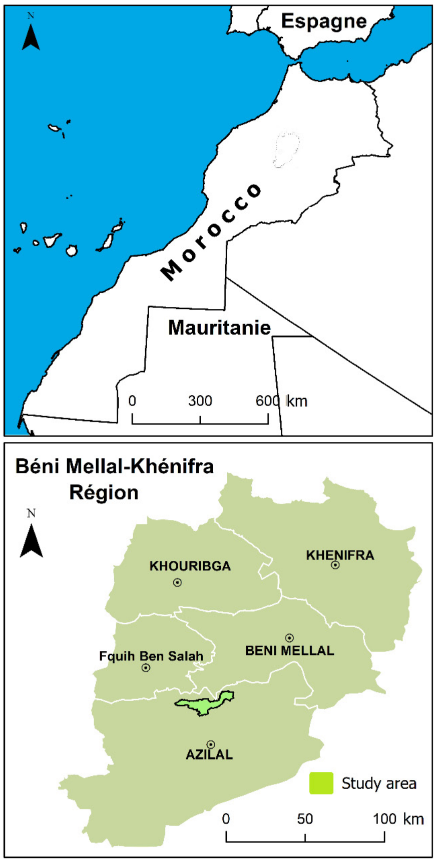
In the present research, multi-temporal remotely sensed and filed data have been used to map the forest canopy species in the Ait Bouzid forest (High Atlas of Afourer, Morocco).

LC classification

Defining the land cover classes is necessary to generate a reliable forest map. Based on common categories defined by the Regional Directorate of Waters and Forests Béni Mellal-Khénifra (Morocco) and field observations, a set of land cover categories of interest were considered for the study area, namely holm oak, cedar, red juniper and euphorbia, water body, and bare ground. These thematic classes are as homogeneous as possible to avoid any mixing between classes during classification (Ettaqy et al. 2020).

Holm oak is the first forest species in terms of its surface area and its production of firewood. This species is found in its pure state in a few places in the forest, at an altitude range between 100 and 1819 m (Barakat et al. 2018). It is generally abundant on the northern slopes corresponding to the bioclimatic stages of the humid, sub-humid, locally semi-arid, temperate and cool and cold types. In addition, this plant species generally colonized all types of geological substrates due to its ecological plasticity and resistance.

The barberry cedar extends over tiny areas, constituting islets experiencing a strong regression under the effect of anthropic action. Its populations are remarkably linked to the warm and temperate



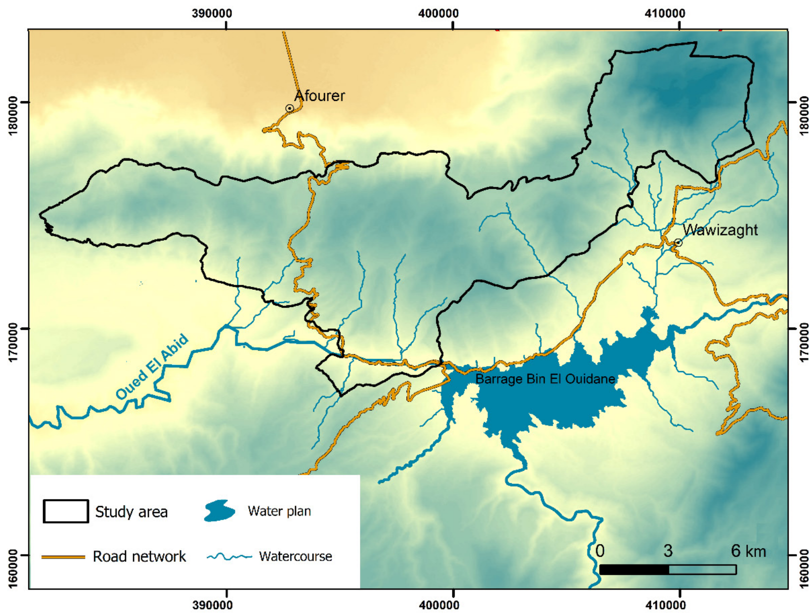


Fig. 1. Location of the study area

variants of the semi-arid thermo-Mediterranean and locally lower sub-humid (Benabid 2000) and have a high plasticity and ecological resistance that allow them to colonize all types of geological substrates.

The red juniper grows in the forest in altitudes between 1000 m and 2200 m and strongly occupies a large part of the two northern and southern slopes of the study area. This softwood is characterized by high plasticity and resistance, allowing it to colonize different substrate types (Benabid 2000).

Euphorbia develops in the study forest at an altitudinal range from 600–1200 m marked by arid to sub-humid bioclimates (Ettaqy et al. 2020). It colonizes cracked calcareous substrates, dolomite and rocky soils. This species has a high socio-economic and environmental interest because it protects the soil against water erosion and allows many cooperatives to produce high quantities of high-quality honey.

The vegetal cover comprised seasonal crops, vegetables, fruit, lower wild plants and shrubs. The bare land corresponds to exposed soils, urban and rural areas and roads (Barakat et al. 2016).

Satellite data

For our case study, the Sentinel-2 MSI product is used for mapping the forest vegetation types because it constitutes a cost-effective technique for spatial data collection and analysis of forest characteristics. Previous research utilizing Sentinel-2 data has conclusively demonstrated its remarkable capability to produce precise vegetation maps, even at the species level. (Kollert et al. 2021; Mahmud et al. 2022; Verhegghen et al. 2022; Potić et al. 2023). The Sentinel-2 MIS also covers a large area with 13 spectral bands (visible, near-infrared and short-wave), as summarized in Table 1.

The Sentinel-2 MIS image used in this study was downloaded from the European Space Agency (https://www.scihub.copernicus.eu) on July 28, 2019, when seasonal crops (especially cereals) become dry and clouds are absent (Jun–Sep).

A field survey was conducted to gather reference data supporting satellite image processing and to assess the precision of the generated land cover map. Seventy-nine samples were taken in the field, Google Earth images, and local forest maps, all representing similar areas of each LC class. (Table 1). All image processing, classification and GIS analyses were performed using QGIS 2.8.6 and ArcGIS 10.2.2 software.

Methods

Data processing and image classification

The present study followed a procedure based on three ML algorithms for discriminating tree species in the Ait Bouzid forest (Central High Atlas) and assessing the efficient classification algorithm by using remote-sensing imagery and geoinformatics tools. To indicate the type and density of the forest trees, the common SVM, KNN and RF classifiers were selected to classify the Sentinel-2 bands resampled to 10. The methodology implemented in this study is illustrated in Figure 2. The work steps consisted of: (I) pre-processing of Sentinel-2 MSI imagery, (II) land cover classification, (III) accuracy

Table 1. Number of training and validation points

LC classes	Number of training points	Number of Validation points
Holm oak	5	9
Thuja	15	4
Red juniper	9	5
Euphorbia	9	6
Water body	2	1
Bare ground	4	8

assessment, and (IV) ML algorithms performance comparison.

Pre-processing of the satellite image

The Sentinel-2 MIS image used in this study was projected to the UTM WGS84 World Geodetic System. Of all the 13 bands (visible, near-infrared and short-wave) that compose the sentinel-2 image, only the bands with a high spatial resolution (10 m and 20 m) can be used for vegetation analysis and mapping according to previous works (Wessel et al. 2018; Kollert et al. 2021). In this study, all bands with a high spatial resolution (B2, B3, B4, B8, B5, B6, B7, 8A, B11, and B12) have been used. The bands at 20 m are resampled to 10 m using the nearest neighbor interpolation and the SEN2COR plugin in SNAP software and GIS environment.

The spectral reflectance remote-sensing data comprehensively identify various ground objects, including vegetation, water and soil (Baumgardner et al. 1986; Khellouk et al. 2020). Therefore, the separability of the selected LC categories in the study forest was checked by their S2A spectral trends. Spectral signatures were generated using QGIS and Excel software from 44 control points (training samples) collected from local forest maps, Google Earth images and field observations (Table 1).

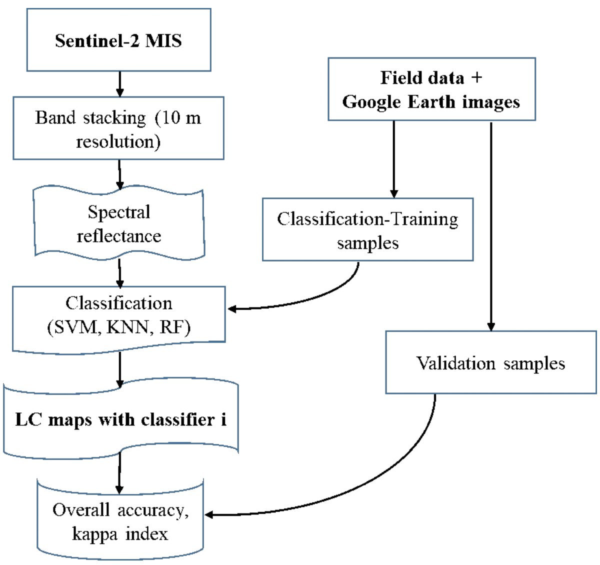


Fig. 2. Schematic overview of steps to create land use maps

ML classifiers selected for forest mapping

To classify the Sentinel-2 bands for mapping forest tree species, the common classifiers SVM, KNN and RF were employed in this study.

- SVM classifier

The SVM is a supervised learning technique employed for both classification and regression tasks, effectively reducing errors for classification and regression, successfully used to minimize the error in data grouping or fit function. It was successfully applied to many classification problems in environmental sciences (George et al. 2014; Lee et al. 2018; Dinh et al. 2021; Han et al. 2021). The input data in the SVM algorithm are divided into testing and training samples. The SVM model splits the training data in feature space based on a user-defined kernel function by a maximal margin hyperplane. The ultimate classification is determined by the position of the unlabeled samples concerning the hyperplane (Evgeniou and Pontil 1999). Compared with other algorithms, the SVM algorithm showed great superiority in solving nonlinear problems, especially small sample data, and often produced higher classification accuracy. Hence, the SVM model was selected for the present study and applied with the predefined parameters ($C=100$ and $\gamma=0.25$).

- KNN classifier

The KNN classifier, developed by Fix and Hodges (1989), is one of the simplest supervised ML algorithms used in various studies for regression and classification (MohanRajan et al. 2020; Nguyen et al. 2020; Ramezan et al. 2021; Barakat et al. 2022). It became a useful nonparametric statistical tool in remote sensing for image classification (Tuominen et al. 2018; Mohammadi et al. 2020; Nidamanuri 2020). KNN computes the similarity between data items considering features. It ranks the sample neighbors among the training examples in the feature space and uses the k most similar neighbor class labels to predict the new class based on maximum votes. In our study, the number of neighboring profiles (k) considered is 5.

- RF classifier

The RF classifier (Breiman 2001) is a popular ML algorithm employed for various natural hazard assessments (Achour and Pourghasemi 2020; Barakat et al. 2022; Sahin 2022), hydrology (Ardabili et al. 2020; Zounemat-Kermani et al. 2021; Mosaffa et al. 2022) and LULC classification (Preidl et al. 2020; Adugna et al. 2022; Barakat et al. 2022). This non-parametric and nonlinear approach combines decision trees constructed from randomly selected predictors and samples in the training dataset. The final classification/prediction decision is generated according to the voting results (Belgiu and Drăguţ 2016). The effectiveness of using the RF algorithm in predicting environmental issues has been discussed in the literature and surpasses the performance of other robust machine learning models like SVM (Ghosh and Joshi 2014; Naghibi et al. 2017; Amiri et al. 2019), in terms of fast computing (Rodriguez-Galiano et al. 2012; Naghibi et al. 2017) and of separating the most important variables from those less important (Catani et al. 2013; Belgiu and Drăguţ 2016; Liaw and Wiener 2021). We employed the RF package with its default settings (classification-RF, fitting 500 trees to all records). The current study utilized the RF model with default settings (100 trees).

Sample data and accuracy assessment

To check the comparability of the SVM, RF and KNN classifiers employed in the present study, the same training samples were used to perform each supervised classification of the sentinel-2 image. The field survey of trees in the studied forest was conducted in June 2022, and 33 sites within the studied forest were surveyed to define the land-use classes and to verify the tree species locations provided by the regional director of Water, Forests and Desertification Control, and validate the generated forest species maps (Table 1). Visual interpretation of Google Earth images was also used to assess accurately the final forest cover maps. For validation, 44 sites of all LU classes were chosen similarly to the training samples (Table 1, Fig. 3).

The accuracy of each model's forest stands species classification was assessed by confusion matrices calculated by the training reference samples (Fig. 3a). The confusion matrices were then used to generate overall classification accuracy, Kappa index (kappa) and producer's accuracy (precision) that are commonly used to evaluate the performance of the applied ML classifiers (Huang et al. 2002; Ghosh et al. 2014). OA and kappa were used to evaluate the proportion of correctly classified imagery pixels and the statistical difference between classifiers. The producer's accuracy enabled assessment of the model's performance based on individual class accuracy.

Results

Analysis of LC maps

The classified maps generated from the three ML algorithms of the Ait Bouzid forest are presented in Figures 4, 5 and 6. The maps showed the spatial distribution pattern of six LC classes, i.e., holm oak, cedar, red juniper, euphorbia, water body and bare ground. Table 2 displays the percentage distribution of each LC class compared to the entire study area for each classifier.

Results from all models established that areas covered by different classes are 0.3–0.32% under water body, 13–21% under bare ground and 78.68–86.7% under forest canopy (Table 2). Similarly, the LC maps showed that 19–28%, 21–27%, 16–24% and 15–18% of the area was covered by euphorbia (32–46 km^2), red juniper (35–44 km^2), cedar (27–39 km^2) and holm oak (24–29 km^2), respectively. Table 2 shows that forest species emerge as the predominant LC class, consistently classified by all classifiers. They cover a total area of about 79 % by KNN, 80% by SVM and 88% by RF.

The comparison of the results of the classification elaborated by SVM and KNN shows a less significant difference in classification that varies between 1% and 2% for the classes of bare soil, holm oak and euphorbia and a significant difference of 8% for the red juniper and cedar classes. Furthermore, it is noticed that the bare ground extends about 13% and 20% of the total studied area using the RF and SVM, respectively. Likewise, the euphorbia regions exhibited a variance of 12 km^2 or 8% of the total area. As for the other classes, the disparities in the occupied area ranged from 1% to 5% % of the total studied area.

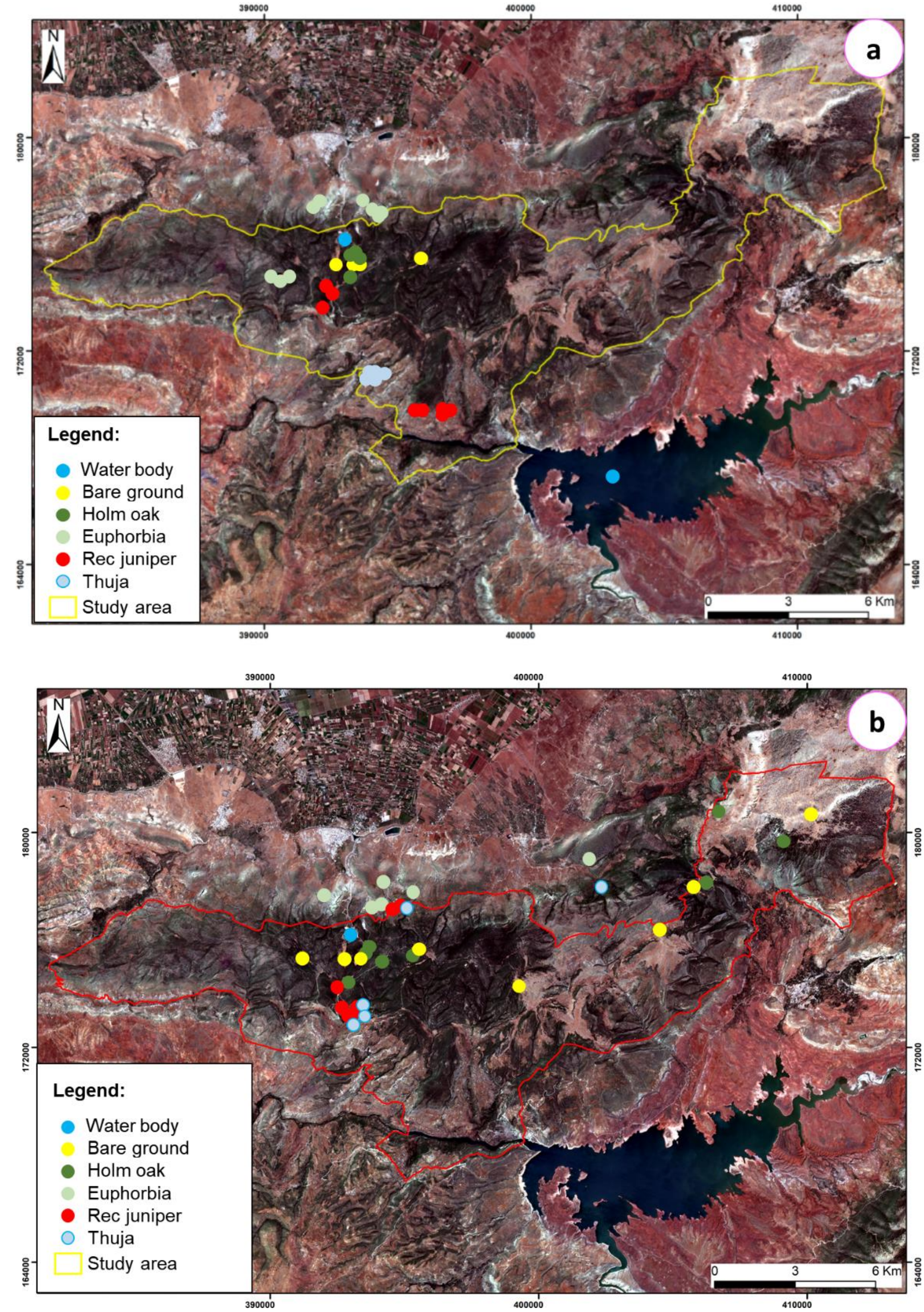


Fig. 3. (a) Location of training points; (b) Location of validation points

Classification accuracy

To identify the optimal ML classifier for the forest canopy mapping at Ait Bouzid forest, a confusion matrix providing average overall accuracy (OA), kappa index and producer's accuracy (Congalton and Green 2019) was used to describe the concordance between the generated classes and the validation samples (Ma et al. 2017; Congalton and Green

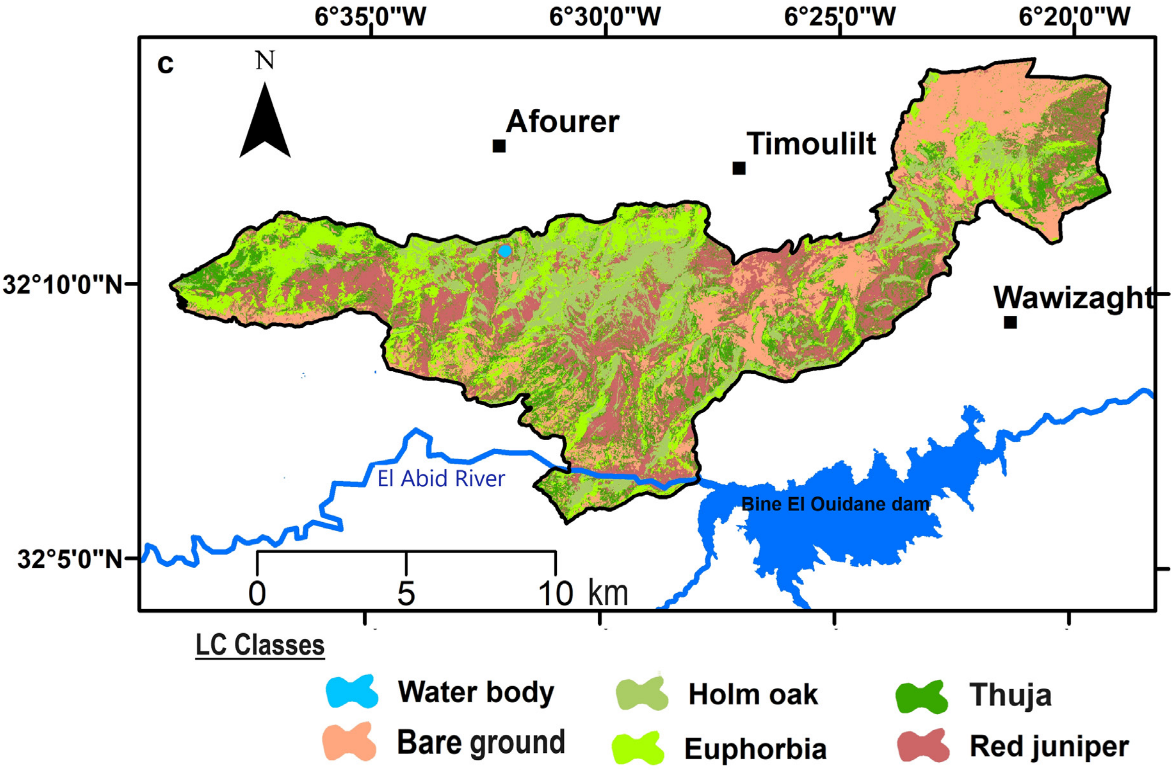


Fig. 4. LC map generated using the SVM classifier

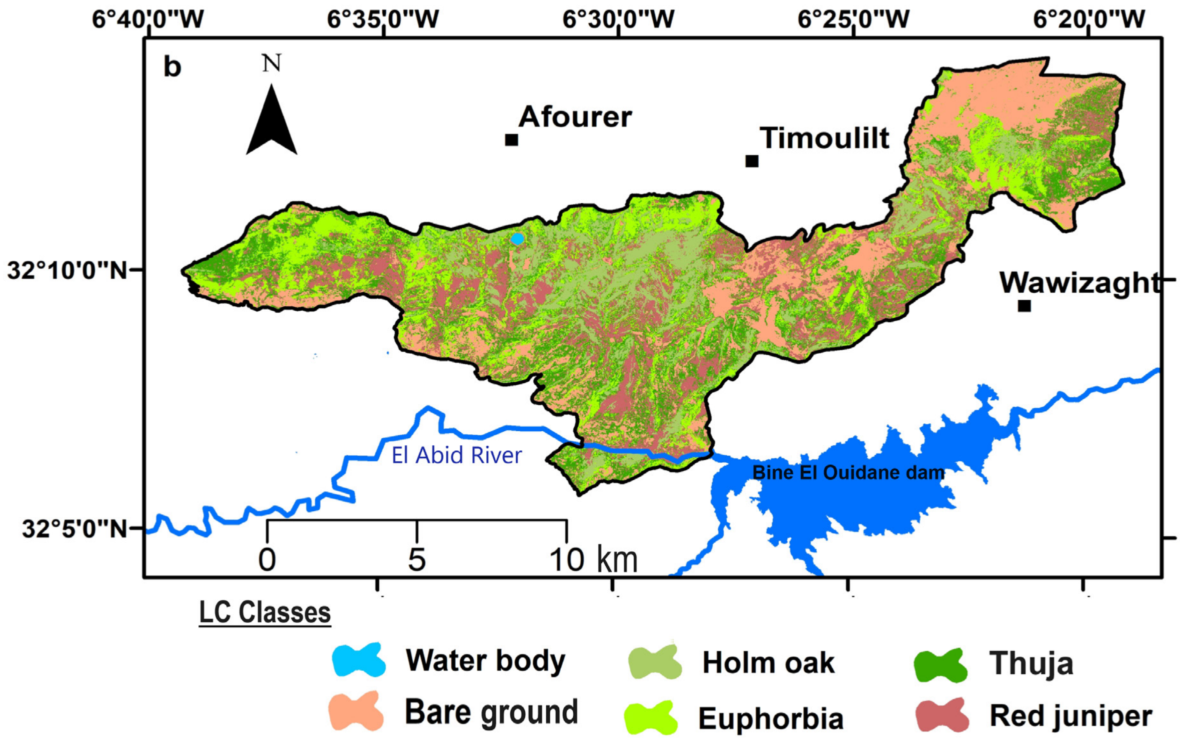


Fig. 5. LC map generated using the KNN classifier

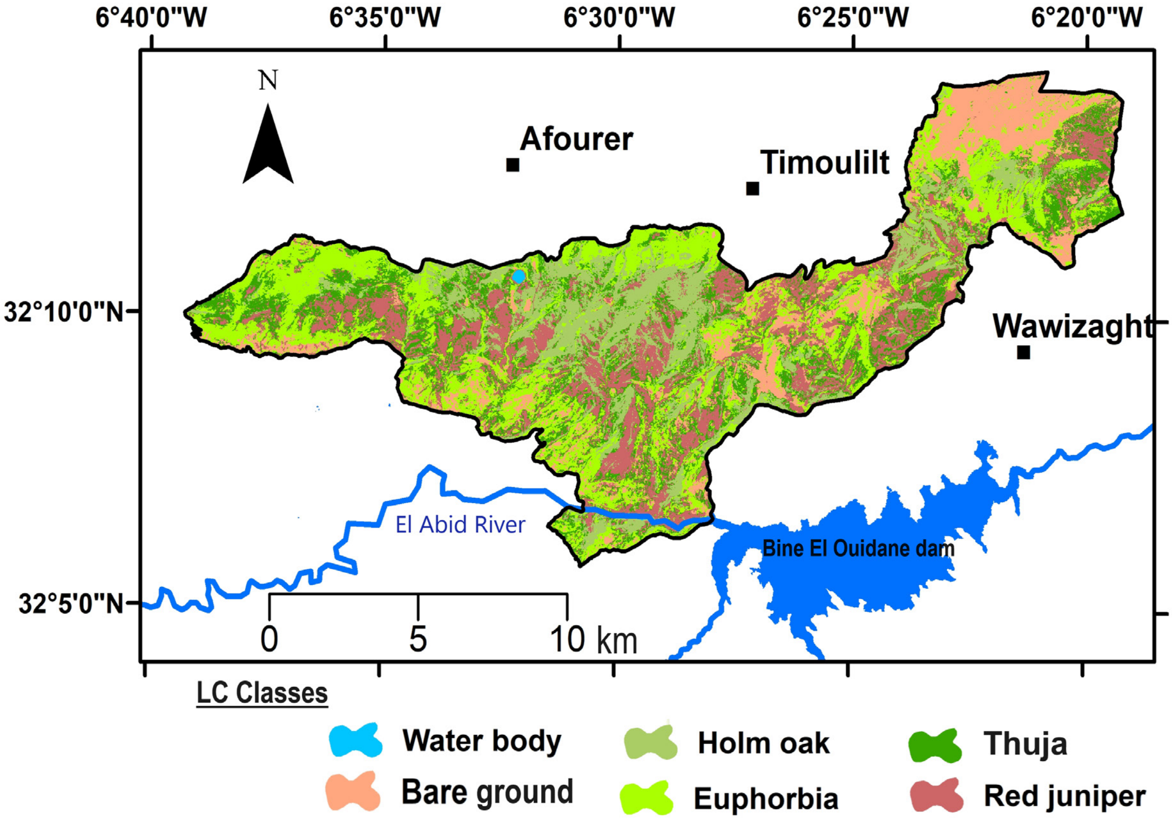


Fig. 6. LC map generated using the RF classifier

2019). Table 3 summarizes the accuracy assessment results for the classification ML algorithms.

The results revealed that the OA reached 73%, 70% and 67% for SVM KNN and RF algorithms, respectively. This indicated that the SVM and KNN models are more effective than the RF model. According to the kappa values that are still moderate (McHugh 2012), the SVM (0.66) produced slightly superior results than the KNN (0.62), followed by the RF (0.59) for LC classifying in the study area.

By analyzing the individual class accuracy, different performances for each of the six LC classes were revealed independently of the employed ML algorithms. The highest (100%) and lowest (40%) UA and PA values were obtained for non-forest classes and mixed forests, respectively.

Regardless of the ML classifiers, the most straightforward classes to identify are water, bare soil and holm oak, with the producer's accuracy attaining 100%. This could be explained by the fact that the classes are relatively pure with distinct spectral variation, i.e., when they have fewer mixed pixels. On average, all three classification methods showed contamination in one or two

Table 2. Comparison of classification results performed by SVM, KNN and RF

LC classes	**SVM**		**KNN**		**RF**	
	Area		Area		Area	
	(km^2)	(%)	(km^2)	(%)	(km^2)	(%)
Holm oak	27	17	24	15	29	18
Thuja	27	16	39	24	32	20
Red juniper	44	27	35	21	35	22
Euphorbia	34	20	32	19	46	28
Water body	0.5	0.3	0.5	0.3	0.52	0.32
Bare ground	32	20	34	21	22	13

other classes. Moreover, the moderate producer's accuracies, ranging between 40% and 67%, were achieved for all other LC classes (Table 3, Fig. 7). Euphorbia showed accuracies of 50%, 51% and 63% generated by SVM, RF and KNN classifiers, likely due to mixed pixels containing spectral signatures of euphorbia, bare soil and red juniper. The results also indicated that red juniper and cedar were the most challenging classes to identify for all ML classifiers, with producer's accuracies of 40–50% and 40–67%, respectively. Intuitively, this could be explained by a significant similarity in the spectral signature of the two types. Even in the field, it is not easy to distinguish between red juniper and cedar due to the strong resemblance of the leaves of these two species belonging to the same family (Cupressaceae).

Previous results revealed notable disparities in OA, kappa index and producer accuracy among the SVM, KNN and RF classifiers. SVM exhibited the highest accuracies, followed by KNN, while RF achieved the lowest accuracy. Therefore, the SVM is more suitable for categorizing tree species classes with near-similar spectral characteristics. Previous works on forests have illustrated that the SVM produced more accurate results in distinguishing forest tree species types (Nasiri et al. 2021). The relatively smaller training dataset may be why SVM outperformed KNN and RF. Numerous studies have already indicated that the precise prediction of ML models employed in our study is highly sensitive to the training sample size. According to Thanh Noi and Kappas (2018), SVM achieved the highest OA with sensitivity to the training sample sizes, followed consecutively by RF and KNN. In contrast, when the training sample size was large enough, all three classifiers showed similar results. Sabat-Tomala et al. (2020) also found that SVM produces a higher accuracy with fewer training samples than RF. This confirms that the SVM is more efficient in mapping forest tree species in the case of small numbers of training samples.

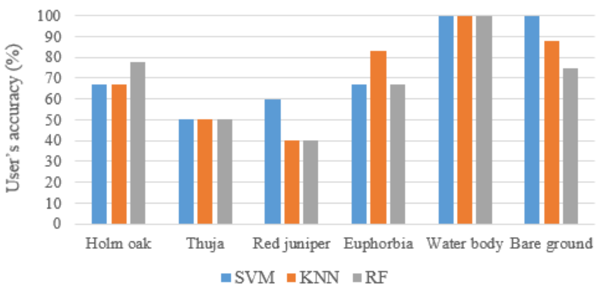


Fig. 7. Comparative graph of classification results

Discussion

The present study investigated the performance of classification algorithms, SVM, KNN and RF to map the Ait Bouzid forest formations (Atlas of Afourer, Morocco) using Sentinel-2 imagery. The mapping of formations, in general, is not a perfect representation of the reality in the field because errors may be present. Therefore, it is important to know the accuracy of these maps generated to exploit them for various applications. For this reason, the accuracy of these three classifiers used in our study was evaluated using the confusion matrix and the Kappa index. Regardless of the results, the SVM significantly outperformed KNN and RF regarding overall accuracies and kappa.

Table 3. Confusion matrix results for the ML algorithms for the LC maps

LC classes	User's accuracy (%)			Producer's accuracy (%)		
	SVM	KNN	RF	SVM	KNN	RF
Holm oak	67	67	78	100	100	100
Thuja	50	50	50	67	40	40
Red juniper	60	40	40	50	40	33
Euphorbia	67	83	67	50	63	51
Water body	100	100	100	100	100	100
Bare ground	100	88	75	89	88	86
OA (%)	73	70	67			
kappa coefficient	0.66	0.62	0.59			

The classification results in this study were affected by mixed signals caused by the small area classes. Moreover, the disadvantage of the directed classification methods used is associated with this confusion of close spectral responses for some forest formations. Some previous studies have reported similar difficulties (N'Dia et al. 2008; Tankoano et al. 2015). Ahmadi et al. (2020) and Hemmerling et al. (2021) demonstrated that multispectral data cannot accurately map diverse plant communities. In addition, the confusion in our case could be related to the definition of homogeneous plots when selecting training sites based on the visual interpretation of Google Earth images. The points representing the sites visited and used to generate the validation points could also be subject to errors related to the complexity of moving through the forest (Fig. 3b). Although the field visit aimed to be as cost-effective as possible regarding the homogeneity of the visited sites, some points are, therefore, close to each other and are sometimes located in small forest formations.

This study confirmed the relevance of the Sentinel-2A sensor in predicting canopy cover in semi-arid regions based on the spectral information provided by the optical remote-sensing data. This ascertainment agrees with previous studies that used Sentinel-2 data in mapping forest canopies (Barakat et al. 2018; Godinho et al. 2018; Suleymanov et al. 2023). Šiljeg et al. (2022) and Liang et al. (2023) found that SVM learning schemes perform well in Object-Based Image Analysis. Immitzer et al. (2016) carried out a study evaluating the potential of Sentinel-2 data for mapping tree species while adopting an object-oriented and pixel-based classification with a supervised classifier (RF); the results obtained are very close: 64% for the first exercise and 66% for the second.

Conclusion

This study employed the common ML algorithms SVM, KNN and RF to classify the Sentinel-2 bands for mapping forest tree species in the Ait Bouzid forest (Central High Atlas, Morocco). The accuracy of these three classifiers was evaluated using the confusion matrix and Kappa index.

Results from all models established that areas covered by different classes are 0.3–0.32% under the water body, 13–21% under the bare ground, and 78.68–86.7% under the forest canopy. Similarly, the LC maps showed that 19–28%, 21–27%, 16–24% and 15–18% of the area was covered by euphorbia (32–46 km^2), red juniper (35–44 km^2), cedar (27–39 km^2) and holm oak (24–29 km^2), respectively. The forest species represent the most dominant LC class in the studied area, classified by all classifiers. They cover of the total area, about 79 % by KNN, 80% by SVM and 88% by RF. An OA of 73%, 67% and 70%, with a kappa coefficient of 0.62, 0.59 and 0.62, were achieved, respectively, for SVM, RF and KNN algorithms for LC classifying in the study area. The classification successfully identified forest trees with water body and bare ground; however, it was affected by mixed signals caused by the small area classes and by confusion related to the definition of homogeneous training sites. Subsequent studies may explore Sentinel time series data across various seasons to assess the precision of forest stand characteristic predictions, determining the optimal season for obtaining more reliable and accurate results.

In a general sense, this study showed that the proposed ML algorithm-based classification for Sentinel-2 images, as outlined in this paper, holds potential for applications in the survey and analysis of forest canopy in other similar areas of Morocco. The SVM classifier has proven its accuracy in mapping forest canopy and is a reliable and fast option. Therefore, further investigations of the SVM algorithm across other variables and datasets are required to understand its limitations and strengths.

Finally, the current study findings could assist decision-makers involved in land-use planning and forest protection in the study region. Also, since we exclusively utilized images from the Sentinel-2 sensor, our approach is well suited for cost-effective classification models to monitor areas similar to those employed in this study.

Disclosure statement

No potential conflict of interest was reported by the authors.

Author contributions

Study design AB, MB, AE; ML; data collection AB, MB, AE; ML; statistical analysis AB, MB, AE, AEl, WE; result interpretation AB, MB, AE; ML; manuscript preparation AB; literature review: AB, MB, AE; ML.

References

ACHOUR Y and POURGHASEMI HR, 2020, How do machine learning techniques help in increasing accuracy of landslide susceptibility maps? *Geoscience Frontiers* 11: 871-883.

ADUGNA T, XU W and FAN J, 2022, Comparison of Random Forest and Support Vector Machine Classifiers for Regional Land Cover Mapping Using Coarse Resolution FY-3C Images. *Remote Sensing* 14: 574.

AHMADI K, KALANTAR B, SAEIDI V, HARANDI EK, JANIZADEH S and UEDA N, 2020, Comparison of machine learning methods for mapping the stand characteristics of temperate forests using multi-spectral sentinel-2 data. *Remote Sensing* 12: 3019.

AMIRI M, POURGHASEMI HR, GHANBARIAN GA and AFZALI SF, 2019, Assessment of the importance of gully erosion effective factors using Boruta algorithm and its spatial modeling and mapping using three machine learning algorithms. *Geoderma* 340: 55-69.

APPEANING ADDO K, 2010, Urban and Peri-Urban Agriculture in Developing Countries Studied using Remote Sensing and In Situ Methods. *Remote Sensing* 2: 497-513.

ARDABILI S, MOSAVI A, DEHGHANI M and VÁRKONYI-KÓCZY AR, 2020, Deep learning and machine learning in hydrological processes climate change and earth systems a systematic review. In *Engineering for Sustainable Future: Selected papers of the 18th International Conference on Global Research and Education Inter-Academia–2019* 18, 52-62, Springer International Publishing.

BARAKAT A, KHELLOUK R, EL JAZOULI A, TOUHAMI F and NADEM S, 2018, Monitoring of forest cover dynamics in eastern area of Béni-Mellal Province using ASTER and Sentinel-2A multispectral data. *Geology, Ecology, and Landscapes* 2: 203-215. DOI: 10.1080/24749508.2018.1452478.

BARAKAT A, OUARGAF Z, KHELLOUK R, EL JAZOULI A and TOUHAMI F, 2019, Land use/land cover change and environmental impact assessment in béni-mellal district (morocco) using remote sensing and gis. *Earth Systems and Environment* 3(1): 113-125.

BARAKAT A, OUARGAF Z and TOUHAMI F, 2016, Identification of potential areas hosting aggregate resources using GIS method: a case study of Tadla-Azilal Region, Morocco. *Environmental Earth Sciences* 75: 1-16.

BARAKAT A, RAFAI M, MOSAID H, ISLAM MS and SAEED S, 2022, Mapping of Water-Induced Soil Erosion Using Machine Learning Models: A Case Study of Oum Er Rbia Basin (Morocco). *Earth Systems and Environment* 7(1): 151-170. DOI: 10.1007/s41748-022-00317-x.

BAUMGARDNER MF, SILVA LF, BIEHL LL and STONER ER, 1986, Reflectance properties of soils. *Advances in agronomy* 38: 1-44.

BELGIU M and DRĂGUŢ L, 2016, Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing* 114: 24-31. DOI: https://doi.org/10.1016/j.isprsjprs.2016.01.011.

BENABID A, 2000, Flore et écosystèmes du Maroc: Evaluation et préservation de la biodiversité.

BOUBEKRAOUI H, MAOUNI Y, GHALLAB A, DRAOUI M and MAOUNI A, 2023, Spatio-temporal analysis and identification of deforestation hotspots in the Moroccan western Rif. *Trees, Forests and People* 12: 100388. DOI: https://doi.org/10.1016/j.tfp.2023.100388.

BOUZEKRAOUI H, EL KHALKI Y, MOUADDINE A, LHISSOU R, EL YOUSSI M and BARAKAT A, 2016, Characterization and dynamics of agroforestry landscape using geospatial techniques and field survey: a case study in central High-Atlas (Morocco). *Agroforestry Systems* 90: 965-978. DOI: 10.1007/s10457-015-9877-8.

BREIMAN L, 2001, Random Forests. *Machine Learning* 45: 5-32. DOI: 10.1023/A:1010933404324.

CATANI F, LAGOMARSINO D, SEGONI S and TOFANI V, 2013, Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. *Natural Hazards and Earth System Sciences* 13: 2815-2831.

CHENG K and WANG JJF, 2019, Forest type classification based on integrated spectral-spatial-temporal features and random forest algorithm—A case study in the qinling mountains. *Forests* 10(7): 559.

CONGALTON RG and GREEN K, 2019, Assessing the accuracy of remotely sensed data: principles and practices. CRC press.

DECUYPER M, CHÁVEZ RO, LOHBECK M, LASTRA JA, TSENDBAZAR N, HACKLÄNDER J, HEROLD M and VÅGEN T-G (2022) Continuous monitoring of forest change dynamics with satellite time series. *Remote Sensing of Environment* 269: 112829. DOI: https://doi.org/10.1016/j.rse.2021.112829.

DINH TV, NGUYEN H, TRAN X-L and HOANG N-D, 2021, Predicting Rainfall-Induced Soil Erosion Based on a Hybridization of Adaptive Differential Evolution and Support Vector Machine Classification. *Mathematical Problems in Engineering* 2021: 6647829. DOI: 10.1155/2021/6647829.

ETTAQY A, TAHA A, ELGHIOUANE A, ELKHOU A, BOULLI A, ABBAS Y, 2020, New data on the ecological distribution of Euphorbia resinifera O. Berg in the Beni Mellal-Khenifra region. E3S Web of Conferences. EDP Sciences.

EVGENIOU T and PONTIL M, 1999, Support vector machines: Theory and applications. In: *Advanced Course on Artificial Intelligence*, 249-257. Berlin, Heidelberg: Springer Berlin Heidelberg.

FAO (Food and Agriculture Organization), 2023, Les societies tirent des avantages colossaux des services écosystémiques forestiers, qui représentent plus d'un cinquième des richesses contenues dans les actifs fonciers. Available at: https://www.fao.org/3/cb9360fr/online/src/html/forests-ecosystem-services-wealth.html (Accessed: 25 July 2023).

FASSNACHT FE, LATIFI H, STEREŃCZAK K, MODZELEWSKA A, LEFSKY M, WASER LT, STRAUB C and GHOSH A, 2016, Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment* 186: 64-87.

FIX E and HODGES JL, 1989, Discriminatory analysis. Nonparametric discrimination: Consistency properties. *International Statistical Review/Revue Internationale de Statistique* 57: 238-247.

GE W, CHENG Q, TANG Y, JING L and GAO C, 2018, Lithological classification using Sentinel-2A data in the Shibanjing ophiolite complex in Inner Mongolia, China. *Remote Sensing* 10: 638.

GEORGE R, PADALIA H and KUSHWAHA SPS, 2014, Forest tree species discrimination in western Himalaya using EO-1 Hyperion. *International Journal of Applied Earth Observation and Geoinformation* 28: 140-149. DOI: https://doi.org/10.1016/j.jag.2013.11.011.

GHOSH A, FASSNACHT FE, JOSHI PK and KOCH B, 2014, A framework for mapping tree species combining hyperspectral and LiDAR data: Role of selected classifiers and sensor across three spatial scales. *International Journal of Applied Earth Observation and Geoinformation* 26: 49-63. DOI: https://doi.org/10.1016/j.jag.2013.05.017.

GHOSH A and JOSHI PK, 2014, A comparison of selected classification algorithms for mapping bamboo patches in lower Gangetic plains using very high resolution WorldView 2 imagery. *International Journal of Applied Earth Observation and Geoinformation* 26: 298-311.

GODINHO S, GUIOMAR N and GIL A, 2018, Estimating tree canopy cover percentage in a mediterranean silvopastoral systems using Sentinel-2A imagery and the stochastic gradient boosting algorithm. *International Journal of Remote Sensing* 39: 4640-4662. DOI: 10.1080/01431161.2017.1399480.

GRABSKA E, FRANTZ D and OSTAPOWICZ K, 2020, Evaluation of machine learning algorithms for forest stand species mapping using Sentinel-2 imagery and environmental data in the Polish Carpathians. *Remote Sensing of Environment* 251: 112103. DOI: https://doi.org/10.1016/j.rse.2020.112103.

GÜLCI N, 2014, Researches on precision forestry in forest planning. Phd thesis, Kahramanmaraş Sütçü İmam University, Faculty of Forestry.

HAN H, SHI B and ZHANG L, 2021, Prediction of landslide sharp increase displacement by SVM with considering hysteresis of groundwater change. *Engineering Geology* 280: 105876. DOI: https://doi.org/10.1016/j.enggeo.2020.105876.

HEMMERLING J, PFLUGMACHER D and HOSTERT P, 2021, Mapping temperate forest tree species using dense Sentinel-2 time series. *Remote Sensing of Environment* 267: 112743.

HUANG C, DAVIS LS and TOWNSHEND JRG, 2002, An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing* 23: 725-749. DOI: 10.1080/01431160110040323.

IMMITZER M, VUOLO F and ATZBERGER C, 2016, First experience with Sentinel-2 data for crop and tree species classifications in central Europe. *Remote sensing* 8: 166.

KHELLOUK R, BARAKAT A, BOUDHAR A, HADRIA R, LIONBOUI H, EL JAZOULI A, RAIS J, EL BAGHDADI M and BENABDELOUAHAB T, 2020, Spatiotemporal monitoring of surface soil moisture using optical remote sensing data: a case study in a semi-arid area. *Journal of Spatial Science* 65: 481-499.

KOLLERT A, BREMER M, LÖW M and RUTZINGER M, 2021, Exploring the potential of land surface phenology and seasonal cloud free composites of one year of Sentinel-2 imagery for tree species mapping in a mountainous region. *International Journal of Applied Earth Observation and Geoinformation* 94: 102208.

LARY DJ, ALAVI AH, GANDOMI AH and WALKER AL, 2016, Machine learning in geosciences and remote sensing. *Geoscience Frontiers* 7: 3-10.

LEE S, HONG S-M and JUNG H-S, 2018, GIS-based groundwater potential mapping using artificial neural network and support vector machine models: the case of Boryeong city in Korea. *Geocarto International* 33: 847-861. DOI: 10.1080/10106049.2017.1303091.

LIANG L, WANG J, DENG F and KONG D, 2023, Mapping Pu'er tea plantations from GF-1 images using Object-Oriented Image Analysis (OOIA) and Support Vector Machine (SVM). *PLoS One* 18: e0263969.

LIAW A and WIENER M, 2021, Classification and Regression by randomForest. *R news* 2(3): 18–22.

MA L, LI M, MA X, CHENG L, DU P and LIU Y, 2017, A review of supervised object-based land-cover image classification. *ISPRS Journal of Photogrammetry and Remote Sensing* 130: 277-293. DOI: https://doi.org/10.1016/j.isprsjprs.2017.06.001.

MAHMUD S, REDOWAN M, AHMED R, KHAN AA and RAHMAN MM, 2022, Phenology-based classification of Sentinel-2 data to detect coastal mangroves. *Geocarto International* 37: 14335-14354.

MAXWELL AE, WARNER TA and FANG F, 2018, Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing* 39: 2784-2817.

MCHUGH ML, 2012, Interrater reliability: the kappa statistic. *Biochemia medica* 22: 276-282.

MICKELSON JG, CIVCO DL and SILANDER J, 1998, Delineating forest canopy species in the northeastern United States using multi-temporal TM imagery. *Photogrammetric Engineering and Remote Sensing* 64: 891-904.

Ministry of Energy Transition and Sustainable Development, Morocco, 2023, National biodiversity strategy and action plan. Available at: https://www.cbd.int/doc/world/ma/ma-nbsap-v2-p15-fr.pdf (Accessed: 19 December 2023).

MISHRA AP, RAI ID, PANGTEY D and PADALIA H, 2021, Vegetation characterization at community level using sentinel-2 satellite data and random forest classifier in western Himalayan Foothills, Uttarakhand. *Journal of the Indian Society of Remote Sensing* 49: 759-771.

MOHAMMADI J, SHATAEE S and NÆSSET E, 2020, Modeling tree species diversity by combining ALS data and digital aerial photogrammetry. *Science of Remote Sensing* 2: 100011.

MOHANRAJAN SN, LOGANATHAN A and MANOHARAN P, 2020, Survey on Land Use/Land Cover (LU/LC) change analysis in remote sensing and GIS environment: Techniques and Challenges. *Environmental Science and Pollution Research* 27: 29900-29926.

MOSAFFA H, SADEGHI M, MALLAKPOUR I, JAHROMI MN and POURGHASEMI HR, 2022, Application of machine learning algorithms in hydrology. *Computers in Earth and Environmental Sciences*: 585-591

NAGHIBI SA, AHMADI K and DANESHI A, 2017, Application of support vector machine, random forest, and genetic algorithm optimized random forest models in groundwater potential mapping. *Water Resources Management* 31: 2761-2775.

NASIRI V, BELOIU M, ASGHAR DARVISHSEFAT A, GRIESS VC, MAFTEI C and WASER LT, 2023, Mapping tree species composition in a Caspian temperate mixed forest based on spectral-temporal metrics and machine learning. *International Journal of Applied Earth Observation and Geoinformation* 116: 103154. DOI: https://doi.org/10.1016/j.jag.2022.103154.

NASIRI V, DARVISHSEFAT AA, AREFI H, PIERROT-DESEILLIGNY M, NAMIRANIAN M and LE BRIS A, 2021, Unmanned aerial vehicles (UAV)-based canopy height modeling under leaf-on and leaf-off conditions for determining tree height and crown diameter (case study: Hyrcanian mixed forest). *Canadian Journal of Forest Research* 51: 962-971. DOI: 10.1139/cjfr-2020-0125.

National Agency for Water and Forests (ANEF), Morocco, 2023, Forestry Development. Available at: http://www.eauxetforets.gov.ma/DeveloppementForestier/Amenagement/Pages/Amenagement.aspx.

NGUYEN HTT, DOAN TM, TOMPPO E and MCROBERTS RE, 2020, Land Use/land cover mapping using multitemporal Sentinel-2 imagery and four classification methods—A case study from Dak Nong, Vietnam. *Remote Sensing* 12: 1367.

NIDAMANURI RR, 2020, Hyperspectral discrimination of tea plant varieties using machine learning, and spectral matching methods. *Remote Sensing Applications: Society and Environment* 19: 100350.

POTIĆ I, SRDIĆ Z, VAKANJAC B, BAKRAČ S, ĐORĐEVIĆ D, BANKOVIĆ R and JOVANOVIĆ JM, 2023, Improving Forest Detection Using Machine Learning and Remote Sensing: A Case Study in Southeastern Serbia. *Applied Sciences* 13: 8289.

PREIDL S, LANGE M and DOKTOR D, 2020, Introducing APiC for regionalised land cover mapping on the national scale using Sentinel-2A imagery. *Remote Sensing of Environment* 240: 111673. DOI: https://doi.org/10.1016/j.rse.2020.111673.

RADHAKRISHNAN S, LAKSHMINARAYANAN AS, CHATTERJEE JM and HEMANTH DJ, 2020, Forest data visualization and land mapping using support vector machines and decision trees. *Earth Science Informatics* 13: 1119-1137.

RAMEZAN CA, WARNER TA, MAXWELL AE and PRICE BS, 2021, Effects of training set size on supervised machine-learning land-cover classification of large-area high-resolution remotely sensed data. *Remote Sensing* 13: 368.

RODRIGUEZ-GALIANO VF, GHIMIRE B, ROGAN J, CHICA-OLMO M and RIGOL-SANCHEZ JP, 2012, An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS journal of photogrammetry and remote sensing* 67: 93-104.

SABAT-TOMALA A, RACZKO E and ZAGAJEWSKI B, 2020, Comparison of Support Vector Machine and Random Forest Algorithms for Invasive and Expansive Species Classification Using Airborne Hyperspectral Data. *Remote Sensing* 12: 516.

SAHIN EK, 2022, Implementation of free and open-source semi-automatic feature engineering tool in landslide susceptibility mapping using the machine-learning algorithms RF, SVM, and XGBoost. *Stochastic Environmental Research and Risk Assessment* 37(3): 1067-1092.

SAJEDI-HOSSEINI F, MALEKIAN A, CHOUBIN B, RAHMATI O, CIPULLO S, COULON F and PRADHAN B, 2018, A novel machine learning-based approach for the risk assessment of nitrate groundwater contamination. *Science of The Total Environment* 644: 954-962. DOI: https://doi.org/10.1016/j.scitotenv.2018.07.054.

ŠILJEG A, PANĐA L, DOMAZETOVIĆ F, MARIĆ I, GAŠPAROVIĆ M, BORISOV M and MILOŠEVIĆ R, 2022, Comparative assessment of pixel and object-based approaches for mapping of olive tree crowns based on UAV multispectral imagery. *Remote Sensing* 14: 757.

SOLEIMANNEJAD L, ULLAH S, ABEDI R, DEES M and KOCH B, 2019a, Evaluating the potential of sentinel-2, landsat-8, and irs satellite images in tree species classification of hyrcanian forest of iran using random forest. *Journal of Sustainable Forestry* 38: 615-628. DOI: 10.1080/10549811.2019.1598443.

SOLEIMANNEJAD L, ULLAH S, ABEDI R, DEES M and KOCH B, 2019, Evaluating the potential of sentinel-2, landsat-8, and Irs satellite images in tree species classification of hyrcanian forest of Iran using random forest. *Journal of Sustainable Forestry* 38(7): 615-628.

SOTHE C, DE ALMEIDA C, SCHIMALSKI M, LA ROSA L, CASTRO J, FEITOSA R, DALPONTE M, LIMA C, LIESENBERG V, MIYOSHI GJG and SENSING R, 2020, Comparative performance of convolutional neural network, weighted and conventional support vector machine and random forest for classifying tree species using hyperspectral and photogrammetric data. *GIScience & Remote Sensing* 57(3): 369-394.

SULEYMANOV A, GABBASOVA I, KOMISSAROV M, SULEYMANOV R, GARIPOV T, TUKTAROVA I and BELAN L, 2023, Random Forest Modeling of Soil Properties in Saline Semi-Arid Areas. *Agriculture* 13: 976.

TAÏBI AN, EL KHALKI Y and EL HANNANI M (2015) *Atlas régional Région du Tadla Azilal Maroc*. Université d'Angers

THANH NOI P and KAPPAS M, 2018, Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. *Sensors* 18: 18.

TUČEK P, CAHA J, JANOŠKA Z, VONDRÁKOVÁ A, SAMEC P, BOJKO J and VOŽENÍLEK V, 2014, Forest vulnerability zones in the Czech Republic. *Journal of Maps* 10: 179-182. DOI: 10.1080/17445647.2013.866911.

TUOMINEN S, NÄSI R, HONKAVAARA E, BALAZS A, HAKALA T, VILJANEN N, PÖLÖNEN I, SAARI H and OJANEN H, 2018, Assessment of classifiers and remote sensing features of hyperspectral imagery and stereo-photogrammetric point clouds for recognition of tree species in a forest area of high species diversity. *Remote Sensing* 10: 714.

VEGA ISUHUAYLAS LA, HIRATA Y, VENTURA SANTOS LC and SERRUDO TOROBEO NJRS, 2018, Natural forest mapping in the Andes (Peru): A comparison of the performance of machine-learning algorithms. *Remote Sensing* 10(5): 782.

VERHEGGHEN A, KUZELOVA K, SYRRIS V, EVA H and ACHARD F, 2022, Mapping canopy cover in african dry forests from the combined use of sentinel-1 and sentinel-2 data: Application to tanzania for the year 2018. *Remote Sensing* 14: 1522.

WESSEL M, BRANDMEIER M and TIEDE D, 2018, Evaluation of different machine learning algorithms for scalable classification of tree types and tree species based on Sentinel-2 data. *Remote Sensing* 10: 1419.

WILSON K, NEWTON A, ECHEVERRÍA C, WESTON C and BURGMAN M, 2005, A vulnerability analysis of the temperate forests of south central Chile. *Biological Conservation* 122: 9-21. DOI: https://doi.org/10.1016/j.biocon.2004.06.015.

ZAGAJEWSKI B, KLUCZEK M, RACZKO E, NJEGOVEC A, DABIJA A and KYCKO MJRS, 2021, Comparison of random forest, support vector machines, and neural networks for post-disaster forest species mapping of the Krkonoše/Karkonosze transboundary biosphere reserve. *Remote Sensing* 13(13): 2581.

ZIADI K, BARAKAT A, EL ALOUI A, OUAYAH M and NAMOUS M, 2023, Modelling and mapping of soil erosion risk based on GIS and PAP/RAC guidelines in the watershed of Tassaoute (Central High-Atlas, Morocco). *Bulletin of Geography Physical Geography Series* 24: 65-83. DOI: 10.12775/bgeo-2023-0005.

ZOUNEMAT-KERMANI M, BATELAAN O, FADAEE M and HINKELMANN R, 2021, Ensemble machine learning paradigms in hydrology: A review. *Journal of Hydrology* 598: 126266.

Received 30 July 2023
Accepted 23 November 2023