Wiley
Journal of Sensors
Volume 2025, Article ID 3138385, 12 pages
https://doi.org/10.1155/js/3138385

WILEY

Research Article

Detecting and Mapping Invasive *Populus alba* Species in Mountainous Ecosystems Using Sentinel-2 Imagery and Random Forest Classification

Morena Mapuru[1,2] **Sifiso Xulu**[1] **Michael Gebreslasie**[1] **and Maleho Sadiki**[3,4]

[1]*School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Westville Campus, Durban 4000, South Africa*
[2]*Scientific Services, Conservation Services Division, South African National Parks, Kimberly 8306, South Africa*
[3]*Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Hatfield, Pretoria 0002, South Africa*
[4]*Digital Earth Africa, Rosebank, Johannesburg 2196, South Africa*

Correspondence should be addressed to Morena Mapuru; johnmapuru97@gmail.com

Received 29 January 2025; Revised 30 July 2025; Accepted 1 August 2025

Academic Editor: Pramita Mishra

Mapping invasive alien plants (IAPs) has become essential for land and biodiversity conservation authorities, as these species can transform the areas they invade. Fortunately, advances in remote sensing using publicly available products such as Sentinel-2 have improved this process, especially in hard-to-access mountainous regions. In South Africa, poplar (*Populus alba*) is among the IAPs of concern and is found in the eastern Free State and elsewhere in the country, but remote sensing has not yet been used to map this species. Using Sentinel-2 imagery and the random forest (RF) algorithm, this study allowed us to: (a) map and distinguish poplar trees from other land covers throughout the year in the eastern Free State's mountainous region, (b) evaluate influential bands and their combinations in classification, and (c) assess the accuracy of the classification for the first and second halves of the year. The results showed that images from the first half of the year (January–June) had higher classification accuracy (overall accuracy [OA] = 91% and kappa = 0.89) than those from the second half (Jul–Dec) (OA = 87% and kappa = 0.84). Poplar and other classes were separable, with poplar mostly found in riparian areas. The study identified variables such as short-wave infrared-1 (SWIR-1), normalized difference vegetation index (NDVI), blue, poplar detection index-1 (PI-1), modified normalized difference water index (MNDWI), near-infrared (NIR), and PI-3 as key parameters for classifying poplar trees in mountainous regions. Overall, our findings demonstrate that Sentinel-2 bands and indices combined with an RF classifier provide an effective method for mapping poplar invasive trees in mountainous ecosystems.

Keywords: invasive alien plant; poplar; random forests; Sentinel-2

1. Introduction

The invasive alien plants (IAPs), referring to species introduced to new habitats, intentionally or unintentionally, by humans or natural influences, such as wind [1, 2], pose a significant threat to natural and human-modified environments worldwide, reportedly affecting biodiversity, ecosystem services, the economy, and society [3–5]. Most of these species are adaptable and naturalize in new ecosystems, outcompeting native species for limited resources [6, 7]. Several IAPs, including poplar species, have been documented to spread and replace important native species in many locations [8–10], particularly in grazing and remote areas, negatively impacting various agricultural and socioeconomic elements [5]. Knowledge of IAPs' geographical distribution and extent of invasion is necessary to reduce these impacts [11], and remote sensing has demonstrated remarkable capabilities in this regard [12]. More recently, Mapuru et al. [13] reviewed remote sensing of poplar studies and found a growth trajectory worldwide, demonstrating varying degrees of success in ground-to-satellite sensors for mapping and monitoring of this species.

Despite these encouraging patterns and the poplar invasion in South Africa, few studies have explored poplar specifically using remote sensing approaches in the country. Publicly available products, such as Landsat and Sentinel-2, have shown satisfactory results in discerning poplar in complex landscapes [14–16]. For this purpose, Sentinel-2 products appear as a preferred choice over Landsat [17], especially in regions affected by frequent cloud cover. This is partly due to their relatively shorter acquisition cycle of three to 5 days (increasing the probability of cloud-free scenes) and 13 spectral bands. These properties have facilitated detailed characterization of poplar, including biophysical structure and biomass from tree to landscape level [18, 19]. Other commonly used products, such as light detection and ranging (LiDAR), are limited by higher costs, particularly in developing countries with limited financial resources [13]. Therefore, the use of Sentinel sensor images has become a practical solution [16, 20], especially in mountainous areas [21]. Studies using a combination of machine learning and Sentinel products to map poplar have achieved optimal classification accuracies [16, 22]. For instance, Ozturk and Colkesen [23] used random forest (RF), support vector machine (SVM), and Adaboost along with Sentinel-2 imagery to map poplar tree growing areas in the Akyazi district of Turkey. The study yielded remarkable results, with RF achieving an accuracy of 99%, outperforming SVM (98%) and Adaboost (97%). Recent studies have improved accuracy by computing various indices suitable for poplar classification in addition to using bands of the images [20]. While there have been successful remote sensing and machine learning techniques for mapping poplar in agriculture and natural forested areas, less has been done to quantify the distribution of poplar in the mountainous landscapes of South Africa. To our knowledge, our study will be the first to spectrally map poplar distribution against other land cover features in South Africa.

Poplar (*Populus alba*) is a fast-growing and highly adaptable species that has been introduced to many regions outside of its natural range, including South Africa [24]. Due to its occurrence and impact in fragile landscapes such as riparian zones, it has been classified as a species of high concern by the South African National Parks (SANParks) [25, 26]. It is listed as Category 2 in the published lists of IAPs under the National Environmental Management: Biodiversity Act (NEMBA; Number 10 of 2004). Poplar species are often found near water bodies and consume more water, leaving native plants with less water [27, 28]. This signifies the need for mapping and monitoring of its distribution and abundance for effective management and conservation of natural resources. Mapuru et al. [10] modeled the potential habitat for poplar invasion in the country, which includes the eastern Free State province as one. In the same region, Steenhuisen et al. [29] manually identify the presence of poplar trees in the Golden Gate Highlands National Park (GGHNP) using traditional survey methods. In this study, we present a first attempt to demonstrate the ability of remote sensing techniques for poplar monitoring in South Africa, which holds much promise to complement other mapping efforts.

The objective of this study is to discern the *P. alba* species over a mountainous landscape in the eastern Free State province of South Africa using Sentinel-2 imagery and the RF algorithm. Specifically, our study allowed us to: (1) distinguish poplar from other types of land cover classes at different times of the year, (2) identify the key spectral bands and indices in the classification process, and (3) evaluate the accuracy of the classification process over the study period.

2. Materials and Methods

2.1. Description of the Study Area. The research study was conducted at Phuthaditjhaba, an area situated in the northeastern Free State (Figure 1). The area's elevation ranges from 1675 to over 2200 m above sea level. Phuthaditjhaba is intersected by streams, such as Metsi-Matso, Mphukojwane, and Namahadi, which converge to form the Elands River. According to the Agricultural Research Council Institute for Soil Climate and Water [30], Phuthaditjhaba has moist and warm summers and dry and cold winters. The region receives an average annual rainfall of about 629 mm, with most of it falling between October and March. The Karoo Supergroup is the primary geological formation in this area [31]. Glenrosa, lithosols, and mispah are the main soil types in Phuthaditjhaba [32]. From a vegetation perspective, the study area is part of the grassland biome, with some forest patches.

2.2. Methodological Approach. The overall approach for executing this study is outlined in the workflow diagram (Figure 2). The workflow includes seven stages: data source, image selection, feature extraction, time series analysis, feature selection, RF classification, and classification accuracy assessment.

2.3. Sentinel-2 Data Acquisition. The study utilized composite imagery from the Sentinel-2 Surface Reflectance (SR) dataset for the first and second half of 2023. This timeframe was chosen to cover all phenological phases. Composite images were preferred over individual ones because they minimize the impact of noisy data caused by elements such as clouds, cloud shadows, and haze. By merging multiple images into one, it becomes easier to compare images over time. The study used a simple median compositing method to produce composites for the first and second halves of the year. The median method is more resilient to outliers and prioritizes observations closest to the central tendency [33]. The data was obtained from the Digital Earth Africa Data Cube [34].

2.4. Image Selection. The image selection process involved filtering the imagery to retain only scenes with less than 15% cloud cover and full (100%) spatial coverage of the study area. Sentinel-2 Level-2A SR products are already atmospherically corrected; no additional atmospheric correction was required. All selected images were subsequently clipped to the boundaries of the study area. This process took place within the Digital Earth Africa Sandbox, a cloud-based computational environment hosted in JupyterLab.

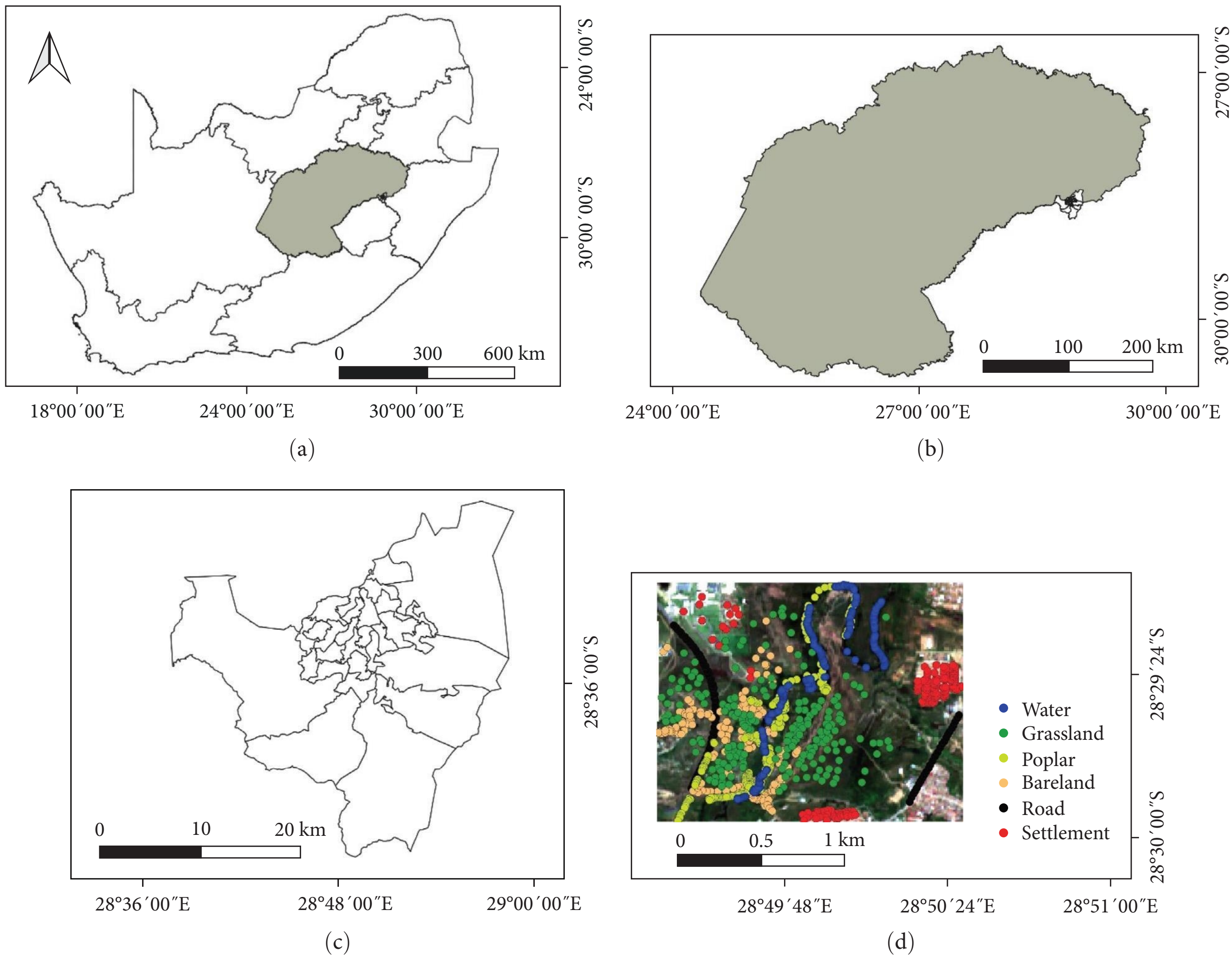


Figure 1: Study area map showing (a) South Africa; (b) Free State; (c) QwaQwa; and (d) Sentinel-2 natural color image of the actual study site, with ground truth data.

2.5. Feature Extraction. For this study, only Sentinel-2 spectral bands relevant to vegetation analysis, specifically bands 2, 3, 4, 5, 6, 7, 8, 11, and 12, were extracted. Bands mainly used for atmospheric studies (bands 1, 9, and 10) were excluded. To ensure consistency in spatial resolution across all inputs, bands originally at 20 m resolution were resampled to 10 m using the Nearest Neighbor method. This standardization enabled accurate analysis and integration of all selected bands. Table 1 shows the spectral bands included in the analysis. Sentinel-2 SR imagery was also used to calculate vegetation indices chosen for their ability to distinguish vegetation from other land cover types while reducing soil background interference. In particular, indices utilizing red-edge and near-infrared (NIR) wavelengths were applied to improve vegetation detection and mapping [35–37]. These indices include the normalized difference vegetation index (NDVI) [38], modified normalized difference water index (MNDWI) [39], leaf area index (LAI) [40], and four variants of the poplar detection index (PI; PI-1, PI-2, PI-3, and PI-4) [20].

2.6. Ground Truth Data (Training and Validation). To produce an accurate land cover map, the first step involved gathering high-quality reference data to train the classification models, following solid standard practices in remote sensing [20, 41]. *Populus alba* presence samples were recorded using a Garmin Global Positioning System (GPS) device set to operate at 2 m positional accuracy. Data collection was carried out under optimal conditions, mainly during clear-sky periods with minimal multipath interference to ensure precise geolocation. To improve accuracy, field measurements were taken after allowing enough time for the GPS receiver to lock onto multiple satellite signals. The field-based dataset was further supplemented with additional reference points taken from high-resolution imagery accessible through Google Earth Pro. Poplar presence samples were collected between January 5 and 15, 2023, a time when *P. alba* shows distinct spectral features that help distinguish it from other land cover types. Sample points were chosen from areas where poplar was the dominant vegetation, especially along riparian zones and wetlands that support its growth. To ensure a balanced and thorough classification, more samples were gathered for other land cover classes, including bare land, grassland, settlements, roads, and water bodies within poplar-invaded areas, using the same GPS method (Table 2). The complete dataset was split into two parts: 70% was used to

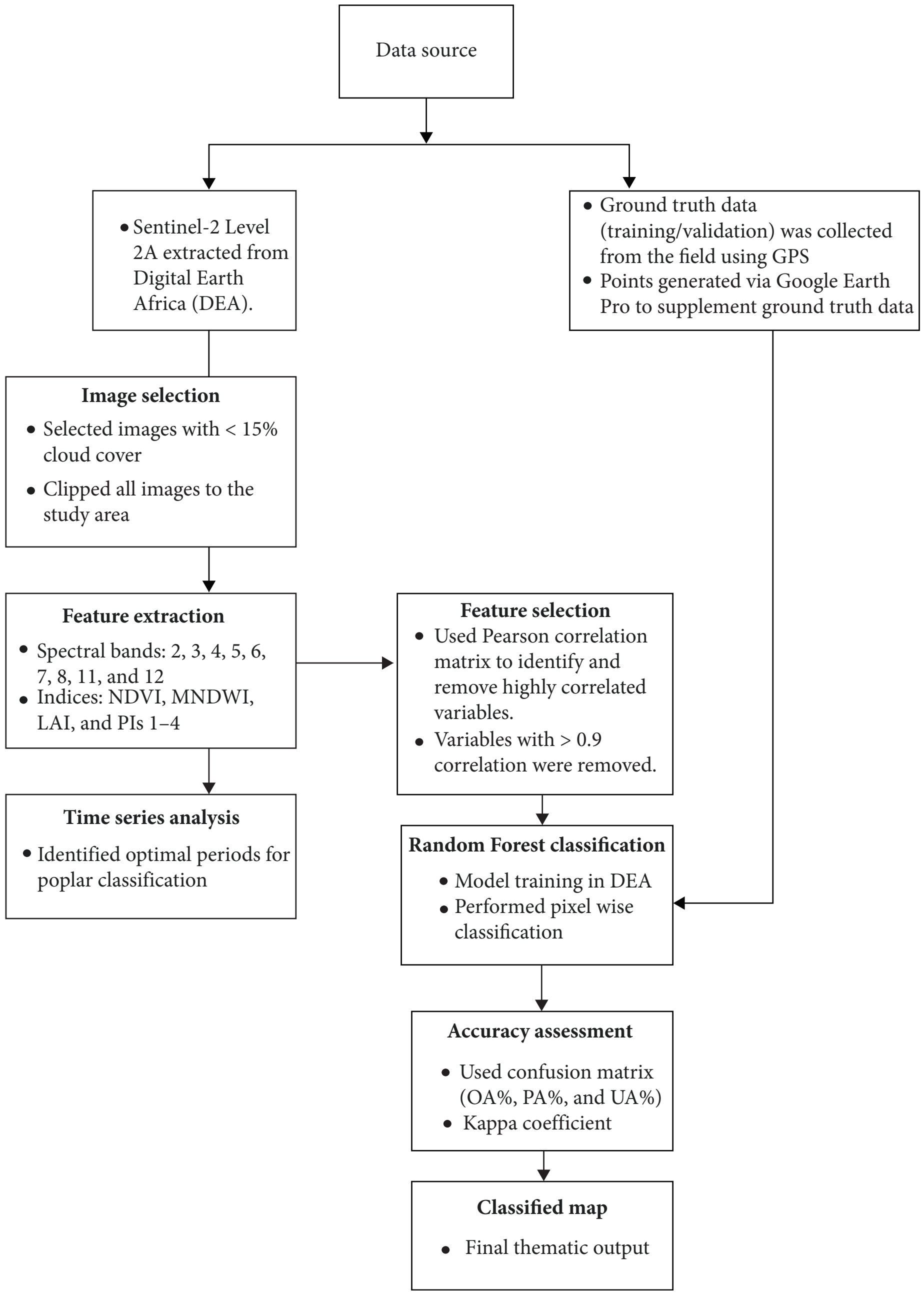


Figure 2: Workflow diagram.

train the classification model to learn the relationship between input features and land cover types, while the remaining 30% was set aside for validation to evaluate the model's accuracy and fine-tune its parameters (Table 2 and Figure 3).

2.7. Time Series Analysis. The time series analysis was conducted to determine the optimal months of the year to include in the machine learning model for mapping poplar species. The primary goal was to identify the periods when spectral characteristics of different land use and land cover (LULC) types, including poplar species, were most distinctive and informative for classification purposes. By understanding these temporal variations, we aimed to enhance the accuracy and reliability of the machine learning models. For each LULC class, median values of the spectral indices were calculated across the time series. This step aimed to identify the months when the spectral characteristics of poplar species were most distinguishable from other LULC types. The time series analysis revealed critical periods during the year when the spectral signatures of different LULC classes, including poplar species, exhibited maximum distinctiveness. These insights guided the selection of specific months for inclusion in the machine learning model, ensuring that the temporal variability captured in the training data would enhance the classifier's performance. Based on the insights gained from the time series plots, it was decided to create semi-annual medians for the periods January–June and July–December.

Table 1: Sentinel-2 imagery bands and vegetation indices used.

Band/index	Description	Central wavelength (nm)	Resolution (m)
B2	Blue	490	10
B3	Green	560	10
B4	Red	665	10
B5	Red edge 1	705	20
B6	Red edge 2	740	20
B7	Red edge 3	783	20
B8	Near-infrared (NIR) wide	842	10
B11	Short-wave infrared (SWIR) 1	1610	20
B12	Short-wave infrared (SWIR) 2	2190	20
NDVI	$\frac{\text{NIR}-\text{RED}}{\text{NIR}+\text{RED}}$	—	—
MNDWI	$\frac{\text{Green}-\text{SWIR}}{\text{Green}+\text{SWIR}}$	—	—
LAI	$3.618 \times \text{EVI} - 0.118$	—	—
PI-1	SWIR1-SWIR2	—	—
PI-2	$\text{Red edge1} - (\text{SWIR1} + \text{SWIR2})$	—	—
PI-3	$\frac{\text{Red edge1}-(\text{SWIR1}+\text{SWIR2})}{\text{Red edge1}+(\text{SWIR1}+\text{SWIR2})}$	—	—
PI-4	$\text{SWIR1} + \text{SWIR2}$	—	—

Table 2: Training and validation data corresponding to land cover classes.

Classes	Training points	Validation points	Total
Bare-land	139	59	198
Grassland	138	59	197
Settlement	106	45	151
Poplar	141	61	202
Road	71	30	101
Water	143	61	204

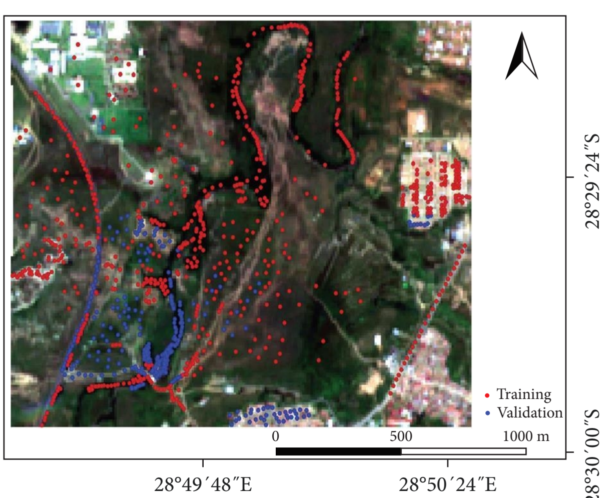


Figure 3: Sentinel-2 true color image with overlaid training and validation data.

2.8. Feature Selection. To improve classification accuracy and make the model easier to interpret, a thorough feature selection process was performed before implementing the RF algorithm. The initial set of predictor variables included 10 Sentinel-2 spectral bands (bands 2, 3, 4, 5, 6, 7, 8, 8A, 11, and 12) along with several derived vegetation and water indices. These indices included the NDVI, NDWI, LAI, and four variants of the PI designed to highlight the spectral features of *P. alba* stands compared to other land covers. A Pearson correlation matrix was created to evaluate multicollinearity among all input variables. Variables with a correlation coefficient above 0.90 were considered highly correlated, and redundant features were removed to prevent overfitting and redundancy. When high correlation existed between variables, the feature with stronger ecological or spectral relevance to poplar detection was kept.

2.9. RF Classification. The RF algorithm was used to classify *P. alba* against other land cover types using selected Sentinel-2 spectral bands and vegetation indices. RF is a strong, nonparametric ensemble learning method that builds multiple decision trees and combines their outputs to enhance classification accuracy and minimize overfitting [42]. Its capacity to handle high-dimensional, collinear, and nonlinear datasets makes it especially suitable for remote sensing land cover mapping [43, 44]. The classification employed a carefully selected set of input features derived through correlation analysis and variable importance ranking. The final variable set included Sentinel-2's short-wave infrared-1 (SWIR-1) band (band 11), blue band (band 2), NIR (band 8), NDVI, MNDWI, and two variants of the PI (PI-1 and PI-3). These features were chosen for their strong ability to distinguish *P. alba* from other land cover types, especially in diverse mountainous landscapes.

To ensure the model's robustness and prevent overfitting, a 10-fold cross-validation was used [45]. The ground truth dataset, comprising samples from six land cover classes, that is, bare land, grassland, settlement, poplar, road, and water, was randomly split into 10 parts. In each cycle, nine parts trained the model and one part validated it. This process was repeated

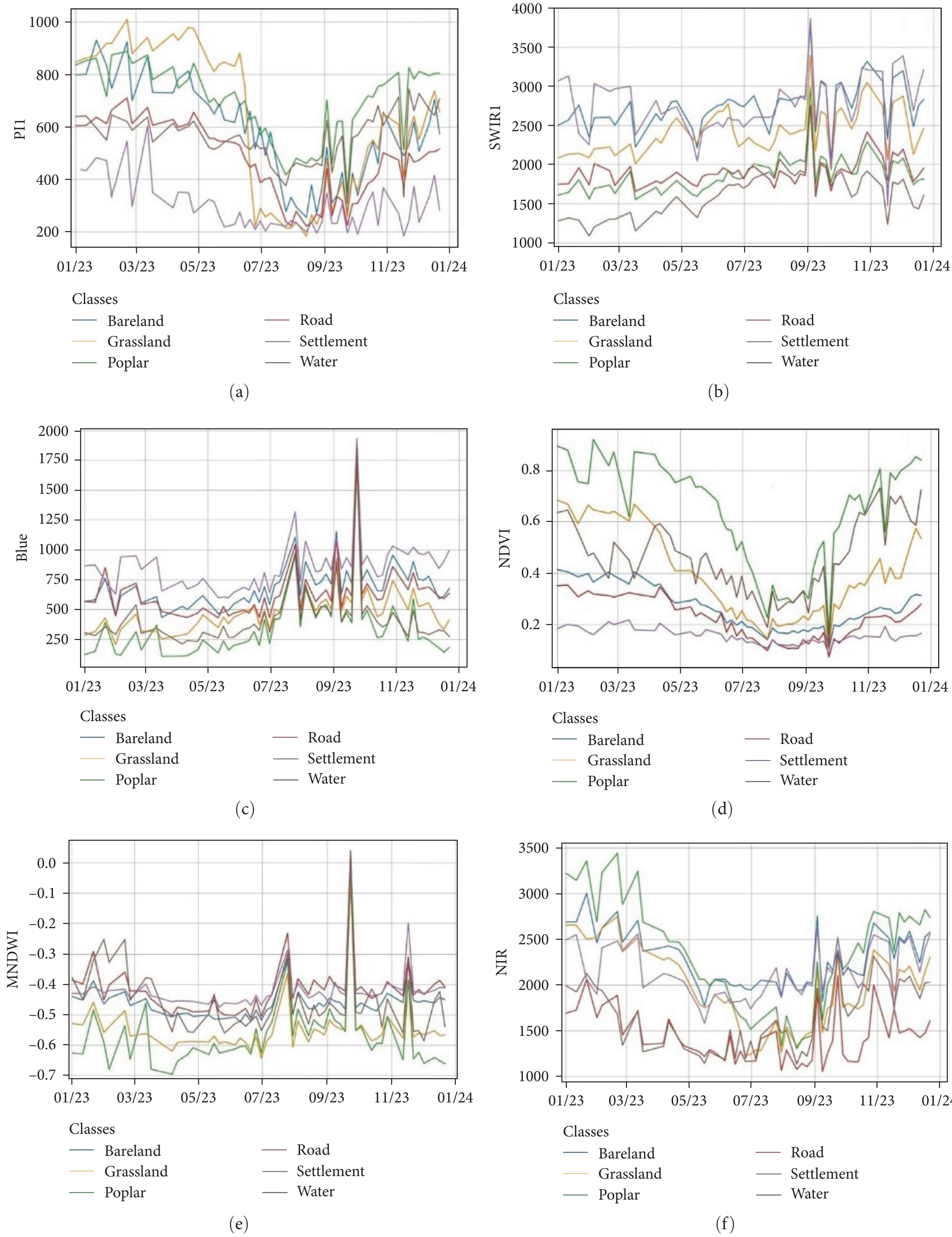
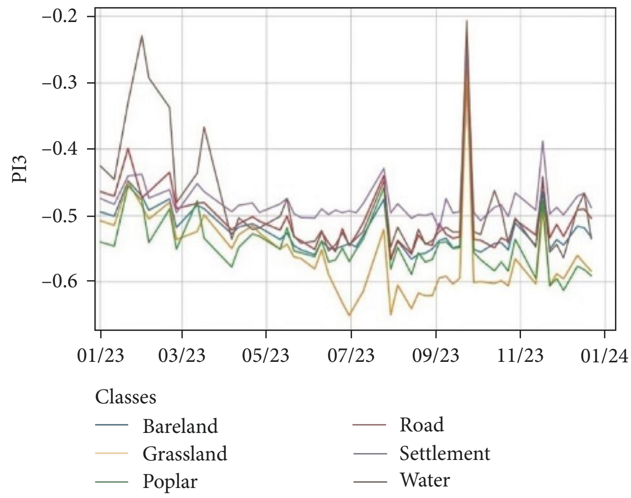


FIGURE 4: Continued.



(g)

FIGURE 4: Spectral signatures for land cover classes ([a] PI-1, [b] SWIR-1, [c] Blue, [d] NDVI, [e] MNDWI, [f] NIR, and [g] PI-3).

10 times, with accuracy metrics averaged across all cycles to provide a reliable performance estimate [45]. The number of trees (ntree) in the RF model was explicitly tested and optimized, while other parameters, such as the number of variables tried at each split (mtry), remained at default settings. To find the best number of trees, the model was run repeatedly with ntree values from 10 to 1000. The configuration achieving the highest overall accuracy (OA) during cross-validation was selected for the final classification. The optimized RF model was then used to classify the entire study area into six land cover classes. The study measured variable importance using the mean decrease Gini (MDG) value, which measures how much a variable reduces the Gini impurity within a specific class. A higher MDG value indicates a purer and more influential variable [46, 47].

2.10. Classification Accuracy Assessment. RFs classification maps were created for images from January to June and from July to December. The confusion matrix, OA, producer accuracy (PA), and user accuracy (UA), as well as, the kappa coefficient, were used to evaluate the accuracy and reliability of the classification process as follows:

$$\text{Overall accuracy} = \frac{\sum_{i=1}^{k} C_{\text{ii}}}{N},$$

where C_{ii} = number of correctly classified pixels for class i (diagonal values of the confusion matrix), N = total number of observations (sum of all elements in the confusion matrix), and k = total number of classes.

$$\text{Producer accuracy}_i = \frac{C_{\text{ii}}}{\sum_{i=1}^{k} C_{\text{ij}}},$$

where C_{ii} = correctly classified samples for class i and $\sum_{i=1}^{k} C_{\text{ij}}$ = total reference samples for class i (column total for class i).

$$\text{User accuracy}_i = \frac{C_{\text{ii}}}{\sum_{i=1}^{k} C_{\text{ji}}},$$

where C_{ii} = correctly classified samples for class i and $\sum_{i=1}^{k} C_{\text{ji}}$ = total predicted samples for class i (row total for class i).

$$\text{Kappa coefficient } (k) = \frac{P_{\text{o}} - P_{\text{e}}}{1 - P_{\text{e}}},$$

where P_{o} = observed accuracy (same as OA) and P_{e} = expected accuracy.

PA indicates the likelihood that a randomly selected point in the field has the same class value as indicated on the map, while UA indicates the likelihood that a randomly selected point on the map matches the class value in the field [48]. OA is calculated as the total number of correctly classified values, and divided by the total number of values [48, 49]. Kappa coefficient measures the agreement between the classified map and reference/ground truth datasets.

3. Results

3.1. Spectral Separability of Classes. We begin this analysis with a detailed consideration of the spectral separability of all classes, with a particular focus on poplars. As already mentioned, we plot temporal signatures of different Sentinel-2 derived bands and their combinations for all classes over the year 2023. The plots of the most influential variables are shown in Figure 4. Our results show a clear pattern of class separability in some variables at certain times of the year while spectral overlap is relatively small but noticeable. In particular, the blue band, NIR, SWIR-1, PI-1, and NDVI show a high degree of separability between classes in the summer and autumn months (Figure 4). Poplar is more distinct than other classes with NDVI, blue band, NIR, and SWIR-1 during this period. In autumn, the blue band showed moderate separability like SWIR-1 in spring. Low separability between winter and spring

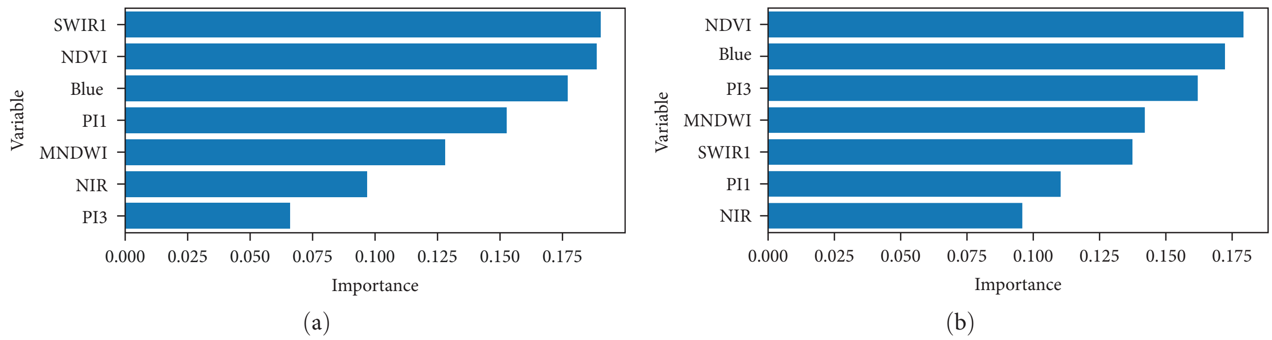


FIGURE 5: Variable importance for (a) January–June and (b) July–December as measured by the MDG indicates the usefulness of the different image bands and indices. The higher the MDG value, the higher the predictor importance.

TABLE 3: Confusion matrix for (a) January–June and (b) July–December.

Classes	Bareland	Grassland	Settlement	Poplar	Roads	Water	UA (%)
			(a)				
Bare land	181	6	2	5	3	1	91
Grassland	4	190	0	1	0	2	96
Settlement	7	1	141	0	2	0	93
Poplar	0	1	0	173	5	22	86
Road	2	0	4	3	92	0	91
Water	0	0	0	19	2	183	90
PA (%)	93	96	96	86	88	88	—
OA (%)				91			
Kappa coefficient				0.89			
(b)							
Bare land	174	8	4	1	3	8	88
Grassland	8	184	0	2	0	3	93
Settlement	14	1	133	0	3	0	88
Poplar	4	2	0	147	9	39	73
Road	2	0	7	3	89	0	88
Water	1	0	0	18	0	185	91
PA (%)	86	94	92	86	86	79	—
OA (%)				87			
Kappa coefficient				0.84			

was observed for the blue band, NIR, SWIR-1, NDVI, PI-1, and PI-3, with notable overlaps of spectral curves. The MNDWI consistently shows low separability throughout the year. Overall, our results suggest that there is high separability between classes in the summer and autumn months, while moderate to low separability is observed in the winter and spring months.

3.2. Influential Variables. In Figure 5, we show the most important spectral bands and indices used to distinguish poplar from other types of land cover in mountainous regions for (a) January–June and (b) July–December 2023. Out of 16 spectral variables, only seven were identified as important for the classification; however, their order of importance slightly varied in the first and second half of 2023. As Figure 5a illustrates, the first half had SWIR-1 as highly ranked, followed by the NDVI, blue band, PI-1, MNDWI, NIR, and PI-3. In the second, the NDVI was the most important in the classification process, followed by the blue band, PI-3, MNDWI, SWIR-1, PI-1, and NIR (Figure 5b).

3.3. Classification Algorithm Performance. We evaluated the classification results based on the combined Sentinel-2 spectral bands and indices, resulting in an overall classification accuracy of over 80% for the January–June and July–December periods (Table 3). The accuracy for January–June classification was higher (UA [86%–96%], PA [86%–96%], OA [91%], and kappa coefficient [0.89]) than for July–December (UA [73%–93%], PA [79%–94%], OA [87%], and kappa coefficient [0.84]). Our results suggest that the RF classifier and images acquired during the January–June period are more optimal for the classification and separability of poplar from other land cover types.

Figure 6 depicts classification maps created from (a) January–June and (b) July–December Sentinel-2 imageries. Both maps exhibit similar spatial distribution patterns for

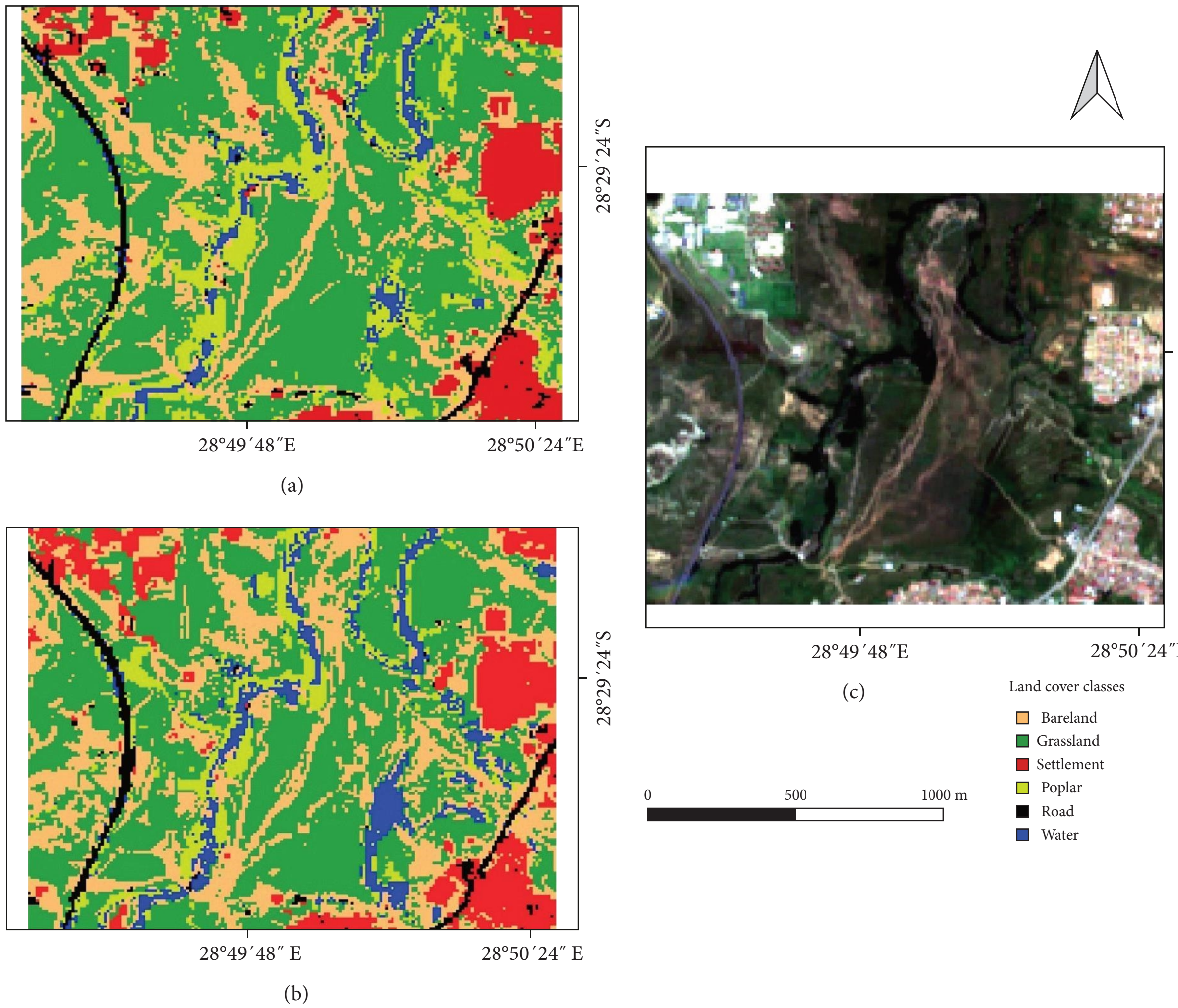


Figure 6: Random forest classification maps for (a) January–June and (b) July–December Sentinel-2 imageries. Sentinel-2 true color image (c) is included for visual inspection against classified maps.

different classes, with a slight overestimation of water, bare land, settlement, and poplar in the July–December images. Specifically, poplar appears to be concentrated along the riparian zones, which aligns with findings from previous studies on poplar distribution trends.

4. Discussion

In this study, we utilized Sentinel-2 satellite imagery in combination with the RF classification algorithm to detect the presence of poplar alien invasive plant species in the mountainous landscapes of Phuthaditjhaba. The classification process involved imagery from the first and second halves of the year, with the goal of assessing the mapping accuracy of both sets of images. Firstly, we evaluated the distinguishability of poplar from other land cover types across the different phases of the year through time series analysis. Secondly, we employed the RF classifier to identify the most important spectral bands and indices in the classification process for both halves of the year. Finally, we computed and produced a classification map depicting the spatial distribution of poplar in relation to other land cover classes.

Our analysis of time series using Sentinel-2 spectral bands and indices showed that there are specific months when poplar and other land cover types can be distinguished from each other, and there are also months when they are not easily separable. We found that during summer and autumn, poplar and other land cover types have distinct spectral characteristics, while during most of the winter and spring, they exhibit somewhat similar spectral characteristics, rendering them difficult to distinguish with high accuracy. Our findings align with Cho et al. [50], who observed that *Acacia* species is distinct from surrounding features during the growing season (November–January) and peak period of vegetation productivity, and less separable during the winter period. Our results provide detailed information on the bands and indices that are associated with high, medium, and low separability at different times of the year. This information is important for selecting the appropriate bands and indices for models that use images from periods where they are particularly distinctive and efficient for classification. We believe that this information will help develop robust models that can produce highly accurate results.

Our study assessed the important variables required for a robust model using Sentinel-2 bands and indices with an RF classifier to classify poplars against other land cover classes. We found that SWIR-1, NDVI, blue, PI-1, MNDWI, NIR, and PI-3 are crucial parameters for this classification. These findings are consistent with a previous study by Kaila et al. [51], which identified the Blue, SWIR, and NIR bands as important parameters for classifying *parthenium* weed. SWIR bands are effective for distinguishing classes [52], while NIR bands are sensitive to vegetation health and density [53]. Additionally, the NIR band, combined with the red band, is used to calculate NDVI, which is utilized to measure vegetation greenness, biomass, and diversity [54, 55]. Incorporating NDVI in a model enables discrimination of vegetation and non-vegetation features based on chlorophyll content sensitivity [56]. PI-1 and PI-3 help highlight the difference between poplar and other landscape features [20], making them important variables for poplar spatial distribution modeling. The inclusion of MNDWI allows for the explicit differentiation of water features from other land cover types [57, 58]. Since poplar is a water-dependent species found in riparian environments, the inclusion of MNDWI parameter is imperative to enhance discrimination efforts. The identification of these parameters has implications for the implementation of the methodology in various landscape settings as well as for the monitoring and management of poplar species, especially in mountainous areas.

Overall, our study produced encouraging results with overall classification accuracy exceeding 80% in both study periods, where the performance from Jan to Jun was higher than from Jul to Dec. As such, poplars were clearly separable from other landscape features, as bare patches and buildings/settlements were highly distinctive, leading to improved classification accuracy. However, during the winter and early spring periods, when vegetation loses its green color and appears brownish like bare land and settlements, the classification accuracy is compromised. The high OA during January–June suggests that the combination of Sentinel-2 bands, derived indices, and RF can effectively map poplar in mountainous regions. These findings align with previous studies [16, 59, 60]. The produced maps could help fill the gap in updating local atlas inventories in South Africa, where deficiencies exist in alien invasive plants land cover maps. While this study was conducted at a local spatial scale, the same approach and dataset (with more additions) could be used to classify other alien invasive plant species in mountainous and various other environmental settings at a much larger spatial scale.

5. Limitations and Future Direction

This study demonstrated the effectiveness of using Sentinel-2 imagery and the RF algorithm to classify *P. alba* in a mountainous region of the north-eastern Free State province, South Africa. While the results were promising, several limitations should be acknowledged, and avenues for future research are proposed. One of the key strengths of this study was the use of seasonal composite imagery derived from both the first and second halves of the year, effectively capturing seasonal variations in vegetation phenology. However, despite accounting for intra-annual variability, the spatial extent of the analysis was restricted to a localized portion of the north-eastern Free State. As a result, the model may not generalize well to other geographic regions with different climatic conditions, vegetation assemblages, or land management practices. Expanding the spatial coverage of training and validation data is therefore necessary to improve the robustness and transferability of the model.

Another limitation is the species-specific focus of the study. The classification was limited to detecting only *P. alba*, excluding other *Populus* species. While this targeted approach was suitable for mapping *P. alba* distribution, it may limit the model's utility in broader invasive species management contexts, where multiple species may coexist or exhibit overlapping spectral characteristics. Future studies should consider incorporating multiple invasive species or functionally similar vegetation types to enhance detection sensitivity and applicability. Although Sentinel-2 provides high-resolution multispectral imagery, its spectral limitation, such as the absence of thermal or hyperspectral bands, may have constrained the ability to fully discriminate between spectrally similar classes. Integrating additional data sources, such as Sentinel-1 Synthetic Aperture Radar (SAR), hyperspectral imagery, or LiDAR, could further improve classification accuracy, especially in densely vegetated or spectrally complex areas.

Moreover, this study employed the RF algorithm as a single classification approach due to its strong performance in remote sensing applications. However, the inclusion and comparison of other advanced machine learning models, such as SVM, gradient boosting machines, or deep learning approaches, such as convolutional neural networks (CNNs), could provide additional insights and possibly enhanced performance.

6. Conclusion

This study aimed to evaluate the effectiveness of Sentinel-2 spectral bands and derived vegetation indices, in combination with an RF classifier, for distinguishing and mapping poplar trees within mountainous landscapes. Specifically, the study sought to identify the optimal seasonal period for accurate poplar detection, as well as to determine the most influential spectral bands and indices to enhance model performance. The results indicated that the January–June period was most suitable for poplar classification, yielding superior performance metrics (OA = 91%; kappa = 0.89) compared to the July–December period (OA = 87%; kappa = 0.84). Key variables that significantly contributed to the classification process included the SWIR, blue, and NIR bands, along with indices such as the NDVI, MNDWI, and two PIs (PI-1 and PI-3). Poplar trees were found to be most distinguishable from other land cover classes when using NDVI, blue, NIR, and SWIR-1. Overall, the findings underscore the substantial potential of integrating Sentinel-2 spectral information with RF classification for operational mapping of poplar trees. Beyond the methodological advantages, the freely available nature of Sentinel-2 data, coupled with its wide swath, high spatial resolution, and frequent revisit time, along with the robustness and efficiency of the RF algorithm, positions this approach as a valuable tool for monitoring invasive alien species in complex mountainous terrains.

Data Availability Statement

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Funding

No funding was received for this manuscript.

References

[1] R. Kannan, C. M. Shackleton, and R. Uma Shaanker, "Reconstructing the History of Introduction and Spread of the Invasive Species, Lantana, at Three Spatial Scales in India," *Biological Invasions* 15, no. 6 (2013): 1287–1302.

[2] T. B. Robinson, K. Peters, and B. Brooker, "Coastal Invasions: the South African Context," in *Biological Invasions in South Africa*, (Springer International, 2020): 229–247.

[3] B. Gallardo, S. Bacher, B. Bradley, et al., "InvasiBES: Understanding and Managing the Impacts of Invasive Alien Species on Biodiversity and Ecosystem Services," *NeoBiota* 50 (2019): 109–122.

[4] K. K. Shah, I. Tiwari, S. Tripathi, S. Subedi, and J. Shrestha, "Invasive Alien Plant Species: A Threat to Biodiversity and Agriculture in Nepal," *Agriways* 8, no. 1 (2020): 62–73.

[5] P. J. Haubrock, R. N. Cuthbert, A. Sundermann, C. Diagne, M. Golivets, and F. Courchamp, "Economic Costs of Invasive Species in Germany," *NeoBiota* 67 (2021): 225–246.

[6] Z. Zhang and M. van Kleunen, "Common Alien Plants are More Competitive Than Rare Natives but not Than Common Natives," *Ecology Letters* 22, no. 9 (2019): 1378–1386.

[7] B. A. Bradley, B. B. Laginhas, R. Whitlock, et al., "Disentangling the Abundance–impact Relationship for Invasive Species," *Proceedings of the National Academy of Sciences* 116, no. 20 (2019): 9919–9924.

[8] I. B. Ngondya and L. K. Munishi, "Impact of Invasive Alien Plants *Gutenbergia cordifolia* and *Tagetes minuta* on Native Taxa in the Ngorongoro Crater, Tanzania," *Scientific African* 13 (2021): e00946.

[9] T. M. Poland, T. Patel-Weynand, D. M. Finch, C. F. Miniat, D. C. Hayes, and V. M. Lopez, *Invasive Species in Forests and Rangelands of the United States: A Comprehensive Science Synthesis for the United States Forest Sector* (Springer Nature, 2021).

[10] M. Mapuru, S. Xulu, and M. Gebreslasie, "Modelling Possible Habitats for Poplar Invasion in South Africa," *KOEDOE - African Protected Area Conservation and Science* 66, no. 1 (2024): 1783.

[11] F. Nininahazwe, J. Théau, G. Marc Antoine, and M. Varin, "Mapping Invasive Alien Plant Species With Very High Spatial Resolution and Multi-Date Satellite Imagery Using Object-Based and Machine Learning Techniques: A Comparative Study," *GIScience & Remote Sensing* 60, no. 1 (2023): 2190203.

[12] P. Zhang, B. Li, J. Wu, and S. Hu, "Invasive Plants Differentially Affect Soil Biota Through Litter and Rhizosphere Pathways: A Meta-Analysis," *Ecology Letters* 22, no. 1 (2019): 200–210.

[13] M. J. Mapuru, S. Xulu, and M. Gebreslasie, "Remote Sensing Applications in Monitoring Poplars: A Review," *Forests* 14, no. 12 (2023): 2301.

[14] L. Zhu, X. Liu, L. Wu, Y. Tang, and Y. Meng, "Long-Term Monitoring of Cropland Change Near Dongting Lake, China, Using the LandTrendr Algorithm With Landsat Imagery," *Remote Sensing* 11, no. 10 (2019): 1234.

[15] P. R. Marbán, J. M. Mullinax, J. P. Resop, and D. J. Prosser, "Assessing Beach and Island Habitat Loss in the Chesapeake Bay and Delmarva Coastal Bay Region, USA, Through Processing of Landsat Imagery: A Case Study," *Remote Sensing Applications: Society and Environment* 16 (2019): 100265.

[16] G. D'Amico, S. Francini, F. Giannetti, et al., "A Deep Learning *Approach* for Automatic Mapping of Poplar Plantations Using Sentinel-2 Imagery," *GIScience & Remote Sensing* 58, no. 8 (2021): 1352–1368.

[17] M. Wang, Y. Zheng, C. Huang, et al., "Assessing Landsat-8 and Sentinel-2 Spectral-Temporal Features for Mapping Tree Species of Northern Plantation Forests in Heilongjiang Province, China," *Forest Ecosystems* 9 (2022): 100032.

[18] R. Cartisano, W. Mattioli, P. Corona, et al., "Assessing and Mapping Biomass Potential Productivity From Poplar-Dominated Riparian Forests: A Case Study," *Biomass and Bioenergy* 54 (2013): 293–302.

[19] M. Liu, C. Cao, W. Chen, and X. Wang, "Mapping Canopy Heights of Poplar Plantations in Plain Areas Using ZY3-02 Stereo and Multispectral Data," *ISPRS International Journal of Geo-Information* 8, no. 3 (2019): 106.

[20] Y. Hamrouni, E. Paillassa, V. Chéret, C. Monteil, and D. Sheeren, "Sentinel-2 Poplar Index for Operational Mapping of Poplar Plantations Over Large Areas," *Remote Sensing* 14, no. 16 (2022): 3975.

[21] J. Hemmerling, D. Pflugmacher, and P. Hostert, "Mapping Temperate Forest Tree Species Using Dense Sentinel-2 Time Series," *Remote Sensing of Environment* 267 (2021): 112743.

[22] H. Tonbul, I. Colkesen, and T. Kavzoglu, "Classification of Poplar Trees With Object-Based Ensemble Learning Algorithms Using Sentinel-2A Imagery," *Journal of Geodetic Science* 10, no. 1 (2020): 14–22.

[23] M. Y. Ozturk and I. Colkesen, "Mapping of Poplar Tree Growing Fields With Machine Learning Algorithms Using Multi-Temporal Sentinel-2A Imagery," in *41th Asian Conference on Remote Sensing (ACRS)*, (Deqing, China, 2020), 9–11.

[24] M. Mbedzi, P. Tshisikhawe, and I. Sinthumule, "The Rate of Recolonization by Native Plant Species After the Eradication of the Invasive Alien Plant Species (*Populus alba* L.) in the Limpopo Province, South Africa," *Ecology Environment and Conservation* 25, no. 4 (2019): 1673–1678.

[25] D. C. Le Maitre, G. G. Forsyth, S. Dzikiti, and M. B. Gush, "Estimates of the Impacts of Invasive Alien Plants on Water Flows in South Africa," *Water SA* 42, no. 4 (2016): 659–672.

[26] SANParks (South African National Parks), *Golden Gate Highlands National Park Management Plan* (SANParks, Pretoria, 2020).

[27] Q. Zhai, L. Xu, T. Zhang, et al., "Transpiration Water Consumption by *Salix matsudana* and *Populus simonii* and Water Use Patterns at Different Developmental Stages on Sandy Land," *Water* 15, no. 24 (2023): 4255.

[28] P. Bakewell-Stone, *Populus nigra (Black Poplar)* (CABI Compendium, 2023).

[29] S. Steenhuisen, V. R. Clark, G. Martin, and N. Cole, "The Current Status of Invasive Alien Plants in Build-up Areas Within Golden Gate Highlands National Park," *Department of Plant Sciences* (2022): 1–18.

[30] Agricultural Research Council Institute for Soil Climate and Water, South Africa, “Meteorological Data,” Nearest Station: QwaQwa: UNIQWA; Station ID: 30426; Latitude: -28.48287; Longitude: 28.82522; Altitude: 1699 m.a.s.l. Date: 01/01/1986-31/08/2016 (2016).

[31] P. Hobbs, C. Colvin, A. Maherry, et al., “Groundwater Reserve Determination for the Middle Vaal Water Management Area,” *WRC* 312 (2013): 13.

[32] Centre for Environmental Management, “State of the Namahadi and Harts rivers (October 2002 to March 2003): a biomonitoring report. Report prepared for Sedibeng Water, CEM, University of the Free State, Bloemfontein, South Africa,” (2003).

[33] M. Jia, D. Mao, Z. Wang, et al., “Tracking Long-Term Floodplain Wetland Changes: A Case Study in the China Side of the Amur River Basin,” *International Journal of Applied Earth Observation and Geoinformation* 92 (2020): 102185.

[34] L. Liu, L. de Simone, C. Fisk, et al., “National Land Cover and Crop Mapping Through the Digital Earth Africa Platform,” in *Space and Geospatial Technologies for the Africa We Want*, Southern Space Studies, (Springer Nature, Switzerland, 2022): 353–362.

[35] R. Zwiggelaar, “A Review of Spectral Properties of Plants and Their Potential use for Crop/Weed Discrimination in Row-Crops,” *Crop Protection* 17, no. 3 (1998): 189–206.

[36] Y. Xie, Z. Sha, and M. Yu, “Remote Sensing Imagery in Vegetation Mapping: A Review,” *Journal of Plant Ecology* 1, no. 1 (2008): 9–23.

[37] Ç. Ö. Genç and A. O. Altunel, “Monitoring the Operational Changes in Surface Reflectances After Logging, Based on Popular Indices Over Sentinel-2, Landsat-8, and ASTER Imageries,” *Environmental Monitoring and Assessment* 197, no. 1 (2025): 120.

[38] J. W. Rouse Jr, R. H. Haas, J. A. Schell, and D. W. Deering, “Third Earth Resources Technology Satellite-1 Symposium: The Proceedings of a Symposium Held by Goddard Space Flight Center at Washington,” 351(1973): 309Paper a 20.

[39] H. Xu, “Modification of Normalised Difference Water Index (NDWI) to Enhance Open Water Features in Remotely Sensed Imagery,” *International Journal of Remote Sensing* 27, no. 14 (2007): 3025–3033.

[40] E. Boegh, H. Soegaard, N. Broge, et al., “Airborne Multispectral Data for Quantifying Leaf Area Index, Nitrogen Concentration, and Photosynthetic Efficiency in Agriculture,” *Remote Sensing of Environment* 81, no. (2-3) (2002): 179–193.

[41] J. Holloway and K. Mengersen, “Statistical Machine Learning Methods and Remote Sensing for Sustainable Development Goals: A Review,” *Remote Sensing* 10, no. 9 (2018): 1365.

[42] P. Duncan, E. Podest, K. J. Esler, S. Geerts, and C. Lyons, “Mapping Invasive Herbaceous Plant Species With Sentinel-2 Satellite Imagery: *Echium plantagineum* in a Mediterranean Shrubland as a Case Study,” *Geomatics* 3, no. 2 (2023): 328–344.

[43] F. M. Martin, J. Müllerová, L. Borgniet, F. Dommanget, V. Breton, and A. Evette, “Using Single- and Multi-Date UAV and Satellite Imagery to Accurately Monitor Invasive Knotweed Species,” *Remote Sensing* 10, no. 10 (2018): 1662.

[44] L. Royimani, O. Mutanga, J. Odindi, K. S. Zolo, M. Sibanda, and T. Dube, “Distribution of Parthenium Hysterophoru L. With Variation in Rainfall Using Multi-Year SPOT Data and Random Forest Classification,” *Remote Sensing Applications: Society and Environment* 13 (2019): 215–223.

[45] L. A. Yates, Z. Aandahl, S. A. Richards, and B. W. Brook, “Cross Validation for Model Selection: A Review With Examples From Ecology,” *Ecological Monographs* 93, no. 1 (2023): 1557.

[46] H. Han, X. Guo, and H. Yu, *Variable Selection Using Mean Decrease Accuracy and Mean Decrease Gini Based on Random Forest* (ICSESS, 2016): 219–224.

[47] M. Belgiu and L. Drăguţ, “Random Forest in Remote Sensing: A Review of Applications and Future Directions,” *ISPRS Journal of Photogrammetry and Remote Sensing* 114 (2016): 24–31.

[48] C. Sammut and G. I. Webb, *Encyclopedia of Machine Learning and Data Mining* (Springer Publishing Company, Incorporated ISBN: 978-1-4899-7685-7, 1335, 2017).

[49] D. A. Çelik and A. O. Altunel, “Is Dynamic World a Contender in Global Land-Cover Making Race? A Swift Field Assessment From Kastamonu, Türkiye,” *The Egyptian Journal of Remote Sensing and Space Sciences* 28, no. 2 (2025): 205–213.

[50] M. A. Cho, A. Ramoelo, and L. Dziba, “Response of Land Surface Phenology to Variation in Tree Cover During Green-Up and Senescence Periods in the Semi-Arid Savanna of Southern Africa,” *Remote Sensing* 9, no. 7 (2017): 689.

[51] Z. Kiala, O. Mutanga, J. Odindi, K. Y. Peerbhay, and R. Slotow, “Automated Classification of a Tropical Landscape Infested by Parthenium Weed (Parthenium Hyterophorus),” *International Journal of Remote Sensing* 41, no. 22 (2020): 8497–8519.

[52] J. T. Atkinson, R. Ismail, and M. Robertson, “Mapping Bugweed (Solanum mauritianum) Infestations in Pinus patula Plantations Using Hyperspectral Imagery and Support Vector Machines,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7, no. 1 (2014): 17–28.

[53] P. Addabbo, M. Focareta, S. Marcuccio, C. Votto, and S. L. Ullo, “Contribution of Sentinel-2 Data for Applications in Vegetation Monitoring,” *Acta IMEKO* 5, no. 2 (2016): 44–54.

[54] A. Michella, F. D. Vescovi, and I. Mustafee, “Sentinel-2 Vegetation Health Index and Early-Onset Drought Detection in Kenya,” *Airbus Defence and Space* (2019): 1–7.

[55] N. Puletti, W. Mattioli, F. Bussotti, and M. Pollastrini, “Monitoring the Effects of Extreme Drought Events on Forest Health by Sentinel-2 Imagery,” *Journal of Applied Remote Sensing* 13, no. 2 (2019): 020501.

[56] R. P. Singh, N. Singh, S. Singh, and S. Mukherjee, “Normalized Difference Vegetation Index (NDVI) Based Classification to Assess the Change in Land Use/Land Cover (LULC) in Lower Assam, India,” *International Journal of Advanced Remote Sensing and GIS* 5, no. 1 (2016): 1963–1970.

[57] S. Szabo, Z. Gácsi, and B. Balazs, “Specific Features of NDVI, NDWI and MNDWI as Reflected in Land Cover Categories,” *Landscape & Environment* 10, no. 3-4 (2019): 194–202.

[58] H. Farhadi, H. Ebadi, A. Kiani, and A. Asgary, “A Novel Flood/Water Extraction Index (FWEI) for Identifying Water and Flooded Areas Using Sentinel-2 Visible and Near-Infrared Spectral Bands,” *Stochastic Environmental Research and Risk Assessment* 38, no. 5 (2024): 1873–1895.

[59] M. Kluczek, B. Zagajewski, and T. Zwijacz-Kozica, “Mountain Tree Species Mapping Using Sentinel-2, PlanetScope, and Airborne HySpex Hyperspectral Imagery,” *Remote Sensing* 15, no. 3 (2023): 844.

[60] M. Wang, M. Li, F. Wang, and X. Ji, “Exploring the Optimal Feature Combination of Tree Species Classification by Fusing Multi-Feature and Multi-Temporal Sentinel-2 Data in Changbai Mountain,” *Forests* 13, no. 7 (2022): 1058.