Makobe *et al.*
Environmental Systems Research (2024) 13:24
https://doi.org/10.1186/s40068-024-00351-w

Environmental Systems Research

RESEARCH

Open Access

Monitoring the invasion of *Campuloclinium macrocephalum* (less) DC plants using a novel MaxEnt and machine learning ensemble in the Cradle Nature Reserve, South Africa

Benjamin Makobe[1], Paidamwoyo Mhangara[1], Eskinder Gidey[1,2*] and Mahlatse Kganyago[3]

Abstract

The proliferation of non-native plant species has caused significant changes in global ecosystems, leading to a surge in international interest in the use of remote sensing technologies for both local and global detection applications. The Greater Cradle Nature Reserve, a UNESCO World Heritage Site, is facing a decline in its global status due to the spread of pompom weeds, affecting its biodiversity. A significant reduction in grazing capacity leads to the displacement of game animals and the replacement of native vegetation. We used Sentinel-2A multispectral images to map the distribution of pompom weeds. At the nature reserve from 2019 to 2024, which allowed us to distinguish it from other land cover types and determine the appropriateness of the habitat. The SVM model provided 44% and 50.7% spatial coverage of pompom weed at the nature reserve in 2019 and 2024, respectively, whereas the RF model yielded 31.1% and 39.3%, respectively. The MaxEnt model identified both soil and rainfall as the most important environmental factors in fostering the aggressive proliferation of pompom weeds at the nature reserves. The MaxEnt predictive model obtained an area under curve score of 0.94, indicating outstanding prediction model performance. Classification of above 75%, indicating that they could distinguish pompom weeds from existing land cover types. For sustainable environmental management, this study suggests using predictive models to effectively eradicate the spatial distribution of invasive weeds in the present and future.

Keywords *Campuloclinium macrocephalum (Less.)* DC, Sentinel-2A, MaxEnt model, Machine learning models, Cradle Nature Reserve, South Africa

*Correspondence:
Eskinder Gidey
eskinder.gidey@mu.edu.et
[1] School of Geography, Archaeology and Environmental Studies, Faculty of Science, University of Witwatersrand, Private Bag: 2050, Johannesburg, Gauteng, South Africa
[2] Department of Land Resources Management and Environmental Protection (LaRMEP), College of Dryland Agriculture and Natural Resources, Mekelle University, P.O Box: 231, Mekelle, Tigray, Ethiopia
[3] Faculty of Science, Geography, Environmental Management & Energy Studies, University of Johannesburg, Johannesburg 2092, South Africa

Introduction

Invasive alien plants (IAPs) are non-native plants transplanted into new ecosystems (Peerbhay et al. 2016). Preston et al. (2018) note that their introduction has economic and ecological effects. IAPs compete with native vegetation for water and space, and they can grow and reproduce across broad areas unaided and negatively influence ecosystem services (Peerbhay et al. 2016). IAPs have negative effects on soil characteristics and water retention, thereby threatening biodiversity, human welfare, and agricultural productivity (Kganyago et al. 2018). At the country level, South Africa reports invasive alien plant status and dispersion, joining the worldwide community. According to Mafanya (2022), South Africa takes steps to reduce biodiversity loss, allocate funds for initiatives, and execute strategic environmental management. *Campuloclinium macrocephalum* (Less) DC, also known as pompom weed, is a hemicryptophytic herb that invades disturbed rangelands. The 1.3-m-tall perennial weed has green to purple stems and light purple to pink flowerheads that bloom in summer (Goodall et al. 2011). In South America and Central America, weeds are native to Argentina and Brazil. Pompom weed is a prominent IAP in South Africa that grows along roads, disturbed grounds, and endangered rangelands.

Pompom weed poses a significant threat in the Gauteng, Limpopo, Northwest and Mpumalanga provinces of South Africa, where it falls under category 1b of invasive species. The National Environmental Management: Biodiversity (NEM: BA) Act of August 2014 suggests categorizing the remaining five provinces of the country under category 1a of the listed invasive species. The presence of pompom weed dominates the central highveld in the Gauteng and Mpumalanga provinces, As well as in the coast of KwaZulu Natal province. Since the 1960s invasion, the intensity has increased (Goodall et al. 2011). The weed is wind-driven, and human mobility has dispersed powdery seeds through mud on shoes and wheels. The frost conditions in the Highveld encouraged the successful emergence of pompom weeds (Goodall et al. 2011). Pompom weeds pose environmental and economic threats to grasslands and wetlands. The Greater Cradle Nature Reserve in Krugersdorp, Gauteng, is a UNESCO World Heritage Site, yet the weed is flourishing there, threatening biodiversity and ecosystem services and reducing its cultural value. The UNESCO General Assembly established the 1972 World Heritage Convention to conserve world cultural and heritage objects of cultural significance. The Convention protects and preserves designated sites with outstanding significance and biodiversity within inscribed sites for all humanity and promotes international cooperation. The ecosystem and sustainability of inscribed World Heritage Sites depend on land cover and biodiversity protection. Protecting heritage sites and monitoring biodiversity can help prevent biodiversity loss and IAP invasions. Understanding how different plant species affect a region's ecology requires vegetation mapping (Saini and Ghosh 2021). South Africa lacks accurate localized spatial data on IAPs, which is critical for planning and eradicating IAPs (Kganyago et al. 2018).

Site surveys, which are a traditional form of terrestrial mapping, are expensive and arduous, and certain locations of interest are inaccessible and difficult to reach (Matongera et al. 2016; Al-dowski et al. 2020). Advanced remote sensing technology has opened new avenues for challenging existing methodologies (Al-Dowski et al. 2020). Remote sensing is useful for gathering IAP spatial data with comprehensive coverage, temporal observations, and affordability (Mafanya et al. 2022). Vegetation mapping studies measure leaf area index (LAI) and biomass using vegetation indicators and spectral characteristics. Satellite remote sensing imagery has low to high resolution. Low-spatial-resolution sensors, while frequently used in short cycles, are not suitable for monitoring. The cost and duration of high-resolution sensors limit periodic monitoring. UAV's are an option to low and high spatial resolution sensors (Katternboorn et al. 2019). UAVs can record remote data with high temporal and geographical resolutions and provide real-time and exact spectrum information (Zhang et al. 2021). Further studies are necessary to understand how environmental factors such as climatic conditions, topography, and biology promote IAP infestations, despite progress in using remote sensing technologies to identify and map IAPs worldwide. These environmental conditions can help or hinder the adaptability of IAPs (Ndlovu and Shoko 2023). Species distribution models (SDMs) have helped quantify how environmental variables play a significant role in IAP adaptability (Miller 2012). SDMs are statistical and mathematical methods used to forecast species proliferation and environmental variables.

Significantly, SDMs aid in the analysis of interactions between the environment and species, as well as predicting the spread of species in different landscapes. This, in turn, facilitates informed decision making regarding the allocation of resources for conservation planning and environmental protection (Mkungo et al. 2023). There are several SDMs techniques that are employed in modelling IAP landscape invasions these includes but not limited to maximum entropy (MaxEnt), generalized linear model (GLM), bioclimatic envelope model (BEM), and Logistic Regression are examples. MaxEnt's computational efficiency, resilience, and ability to analyze incomplete data have contributed to its popularity and extensive publication. Ndlovu and Shoko (2023) considered rainfall and temperature when mapping and forecasting the spread of L. camara in the Inkomati catchment of Mpumalanga. Image classification employs remote sensing techniques to gather data on land-use and land-cover types (Xu 2021). Selecting an image classification algorithm is challenging because of the diverse range of data sources and variations in training data sizes (Saini and Ghosh 2021). A multitude of studies have assessed parametric and non-parametric image classifiers to determine their suitability for land cover and IAP mapping, with the aim of aligning them with scientific and environmental objectives. Distance-based and probabilistic parametric image classifiers can categorize objects or features by determining their similarities to predefined thresholds. Parametric image classifiers include the maximum likelihood classifier (MLC), the minimum-to-distance mean, and Iterative Self-Organizing Data Analysis Technique (ISODATA). The robust support vector machine (SVM), random forest (RF), artificial neural network (ANN), and convolutional neural network (CNN) are all image classifiers that don't use parameters. Instead, they use training samples made by analysts to find the best object boundaries. Remote sensing studies have focused on combining machine learning (ML) and deep learning (DL) to address the issue of data duplication and improve the accuracy of classifying land cover challenges posed by large datasets.

South Africa's Greater Cradle Nature Reserve at the Cradle of Humankind is a private game reserve and heritage site northwest of Gauteng Province. In 1999, Durand et al. The Cradle of humankind is declared a world-historic site owing to its outstanding cultural value and distinct biodiversity. The Cradle of Humankind World Heritage Site (COH-WHS) is a popular tourist destination, as well as a paleo-scientific and archaeological research-intensive site (Durand et al. 2010). The 1972 Convention on Cultural and Natural Heritage has protected all world heritage sites. The World Heritage Convention Act of 1999 (Act No. 49 of 1999), the National Heritage Resources Act of 1999 (Act No. 25 of 1999), the National Environmental Protected Areas Act of 2003 (Act No. 57 of 2003), the National Environmental Management Biodiversity Act (Act No. 10 of 2004), and the Physical Planning Act of 1967 all protect COH-WHS from mining. This law ensures the protection of world-historic sites from mining. Any development must undergo an environmental impact assessment (UNESCO 1972). According to Article 11.4, the World Heritage Committee must report all inscribed world heritage sites threatened by rapid urban development, illegal land invasions, and military, natural, or human-caused environmental degradation (UNEP World Conservation Monitoring Center 2000).

The plant known as *Campuloclinium macrocephalum* (Less.) DC, which is native to South America and Central America, is highly invasive in the Greater Cradle Nature Reserve.

In South Africa, Regulation 15 of the Conservation of Agricultural Resources (Act 43 of 1983) and Section 97(1) of the Alien and Invasive Species Regulations (2014) of the National Environmental Management: Biodiversity Act (Act 10 of 2004) have declared the plant an invasive or exotic weed and placed it in category 1b invasive species. A highly invasive rangeland weed, *Campuloclinium macrocephalum* (Less.) DC is currently causing unprecedented growth in the Greater Cradle Nature Reserve. "Pompom weed" invades grasslands and wetlands, turning green landscapes pink in spring and autumn. Soil erosion accelerates land degradation and diminishes the capacity of wetland water retention, thereby providing an opportunity for poisonous pompom weeds to rapidly surpass the native South African species. This poisonous weed displaces the native South African species and threatens biodiversity. Pompom weed is a rapidly spreading invasive alien plant in South Africa that threatens the highveld grassland and savannah biomes. Pompom weeds degrade UNESCO World Heritage Sites and harm biodiversity. The unregulated spread of pompom weeds reduces game animals' grazing capacity and leads to migration. The nature reserve loses its value and recognition. Thus, Gauteng tourism suffers, resulting in job losses owing to lower revenue. To effectively manage the environment sustainably, it is crucial for community, government, and non-governmental organizations to collaborate (Gebregergs et al. 2021).

The National Environmental Management Biodiversity Act (Act 10 of 2004) prohibits invasive weed planting, propagation, and sale. Therefore, we must monitor and eradicate pompom weeds to stop their spread throughout South Africa and preserve the heritage designation of the Greater Cradle Nature Reserve. The eradication measures that are currently used to control IAPs include the use of registered herbicides (Plenum, Access, and Climax), uprooting and burning of the weeds. Eradication efforts provide only temporary solutions for IAP control. Spring-growing pompom weeds possess water-retaining roots and are fire- and herbicide resistant. Some remote sites may be inaccessible to control. Researchers have successfully applied remote sensing to monitor and map IAPs at various locations. Remote sensing can detect invasive weeds early and accurately track their spatial distribution, thereby aiding decision-making and control. Understanding the proliferation and geographical shifting of IAPs under different environmental conditions can assist in reducing their environmental impact and improving environmental management. This research study uses remote sensing to help the nature reserve manage pompom weeds and improve environmental management practices. The report highlights the following research challenges: Numerous studies have not demonstrated the superiority of any machine learning or deep learning algorithms. Tracking IAPs improves environmental management. However, tracking the spatial distribution of pompom weeds is not sufficient to address its complexity and quick expansion. As a result, we must investigate soil moisture, topography, and climate to understand their impact on weed growth, as well as forecast sensitive landscapes and potential future invasions. The aim of this study was to: i) model the spatial distribution of *Campuloclinium macrocephalum* (Less.)

DC invasive weed at the Greater Cradle Nature Reserve using multispectral Sentinel-2A data and machine learning models; ii) use the MaxEnt species distribution model to strengthen the SVM and RF model findings and recommend effective eradication control measures.

Materials and methods

Study area

The study was conducted in the Greater Cradle Nature Reserve, which is situated in the Cradle of Humankind between Johannesburg and Pretoria, two of the country's largest cities in the Gauteng Province of South Africa. It is located between 27°43′30 ″ E and 27°540 ″ E and 27°57′0 " S to 25°50′0 " S on the Kromdraai Road (Fig. 1). The privately owned Greater Cradle Nature Reserve spans 3000 to 9000 ha of pristine dolomite grassland on Muldersdrift Farm, located near Krugersdorp in the municipality of Mogale. The Blaaubank River Valley runs on the northern side. It has exceptional paleoanthropological value because it includes sites (Swartkrans, Coopers Cave, and Bolts Farm) that document over three and a half million years of landscape, faunal, environmental, and human evolution (Stradford et al. 2016). Owing to the discovery of fossilized remains, The Greater cradle nature reserve is situated within the Cradle of humankind, has served as a significant research hub for archaeologists and anthropologists for over a decade. This site is associated with the origin of the modern human race. Researchers have discovered numerous fossils, and the variety of plant and wildlife species preserved in the nature reserve represents South Africa's diversity. The United Nations Educational and Scientific Council Organization (UNESCO) accorded the Greater Cradle Nature Reserve a World Heritage Site. The pompom weed, *Campuloclinium macrocephalum* (Less.) DC, is a 1.3 m-high perennial, erect herb flowering mostly in the spring–autumn season (McConnachie et al. 2011; Mafanya et al. 2022). Fluffy pink flowerheads with light green leaves adorn the plant, strew along the entire length of the green stem, and cluster at the base to form a rosette (McConnachie et al. 2011). The plant begins flowering in spring and dies in autumn (Goodall et al. 2011). The 1960s saw the introduction of pompom weed for ornamental purposes in South Africa, but between the 1990s and the 2000s, it underwent dramatic expansion and became invasive (McConnachie et al. 2011). Seven of the nine South African provinces currently host pompom

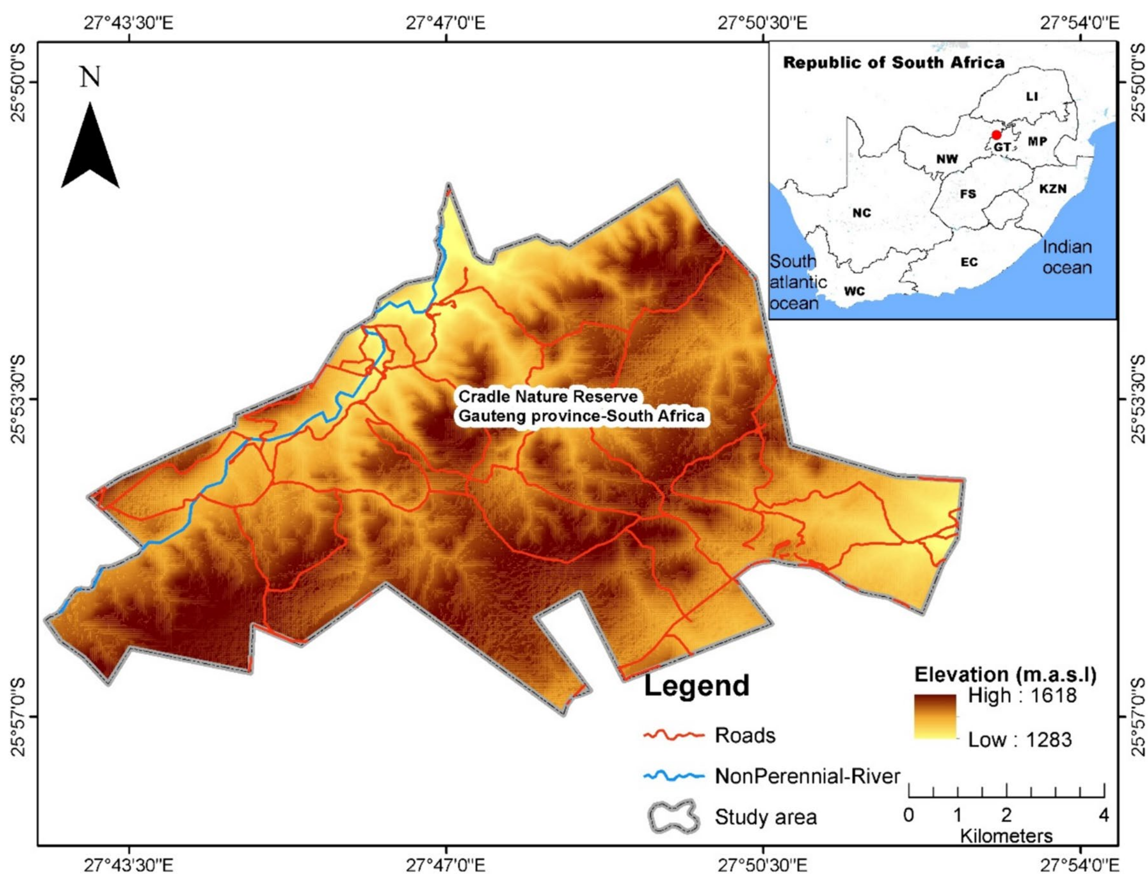

Fig. 1 Location of the study area

weeds. In Gauteng Province, there are frequent infestations of pompom weeds. This invasive weed invades grassland and savannah biomes, where it has a detrimental impact on biodiversity and ecosystem services. It degrades the rangeland and reduces the grazing capacity of large herbivores.

Methods of data acquisition

Field data

Its 26 and 27, 2024, when pompom weeds had aggressively encroached on the Cradle Nature Reserve, field data collection took place. We conducted an initial field survey in the winter of June 2023, when the pompom weed had completely died. The second survey took place in November 2023 during the flowering stage of the pompom weed. However, erratic rainfall in 2023 caused the weed to flower late, resulting in sparse availability in nature reserves. The field survey of 25th and 26th January 2024 was conducted following the full flowering and encroachment of pompom weed, which transformed the nature reserve into a vibrant pink-purple hue. We constructed a 10 m × 10 m plot and randomly collected 469 pompom weed ground control points (GCP) using a handheld Garmin Etrex 10 Global Positioning System. To perform classification, validation, and species modeling using MaxEnt software and machine learning models, such as SVM and RF, we split the ground truth data into a ratio of 70:30. We utilized existing knowledge of the study area and employed pixel image classification to generate additional land cover categories (such as bushland, riparian zones, water bodies, bare land, and cropland) on satellite imagery. We recorded the Ground Control Points (GCPs) using Microsoft Excel spreadsheets, saved them as comma-separated values (CSV), imported them into Google Earth Engine Pro, and superimposed them on the shapefile of the study area.

Earth observation data

The European Space Agency (ESA) (https://scihub.copernicus.eu) operates the Copernicus Open Access Hub, which produces multispectral Sentinel-2A data presented in Table 1. The present study employed Sentinel-2A, with a spatial resolution of 10 m. The obtained satellite imagery corresponded to the date of field data collection, which was January 2024. As a result, we recorded cloud cover below 10%. In addition, for January 2019, we obtained data from the same sensor with a spatial resolution of 10 m and less than 1% cloud cover. Initially intended for land and coastal applications, Sentinel-2A satellite imagery has gained popularity owing to its global coverage and unrestricted access. The advanced satellite Sentinel-2A has a multispectral imager (MSI) with a 290 km swath width, allowing it to monitor land cover. The 13 spectral bands, which encompass the visible, near-infrared (NIR), and shortwave infrared, range in pixel size from 10 to 60 m. The temporal resolution is 10 days when conducted using a solitary satellite; when utilized in conjunction, the satellites provide data with a spatial resolution ranging from 10 to 60 m within a revisit time of 5 days (Miranda et al. 2018). We arranged the sensor's thirteen bands as follows (Table 1).

Table 1 The study presents the Sentinel-2A image characteristics for the years 2019 and 2024

Band	Resolution (m)	Central wavelength (nm)	Description
B1	60	443	Ultra blue
B2	10	490	Blue
B3	10	560	Green
B4	10	665	Red
B5	20	705	Visible and NIR
B6	20	740	Visible and NIR
B7	20	783	Visible and NIR
B8	10	842	Visible and NIR
B8A	20	865	Visible and NIR
B9	60	940	SWIR
B10	60	1375	SWIR
B11	20	1610	SWIR
B12	20	2190	SWIR

For vegetation monitoring research, the Sentinel-2A multispectral sensor is better than Landsat and the Moderate Resolution Imaging Spectroradiometer (MODIS) because it can provide information at a higher resolution (Royimani et al. 2019; Mafanya et al. 2022). Biophysical information about plants, like the amount of chlorophyll in the leaf and the leaf area index (Xie et al. 2019), is easier to get with the Sentinel-2A MSI red-edge bands between 705 and 750 nm. When environmental factors are added to the extra vegetation-sensitive bands of Sentinel-2A MSI, they can help determine the differences between species and make predictions and maps more accurate (Mtengwana et al. 2021).

Image pre-processing and analysis techniques

To correct geometric and radiometric errors and achieve a cloud-free study area, image preprocessing is a prerequisite (Wong and Sarker 2014). Sen2Cor is a software application that processes Sentinel-2A satellite data, generates Level 2A products, and formats outputs. A classification approach in which a user oversees the pixel-classification process is known as supervised classification (Miranda et al. 2018). For land cover mapping, the user selects various pixel values or spectral signatures that represent a specific class (Miranda et al. 2018).

The supervised classification process begins by identifying sample locations for the various types of land cover training sites. Following that, the computer algorithm codes the training site's spectral signature and sorts the whole remote sensing image into groups based on the pixel values or the spectral reflectance of the different types of land cover and land use (Civico 1993; Miranda et al. 2018; Akalu et al. 2019). Ideally, classes should not overlap or overlap slightly with other classes (Miranda et al. 2018). This study utilized ESRI ArcGIS Pro to perform supervised image classification. We employed the SVM and RF classification models to perform supervised image classification.

Support vector machines (SVM)

Support Vector Machine (SVM) are a widely published supervised machine learning algorithm for classification and regression modeling (Kganyago et al. 2018). The support vector machine (SVM) algorithm sorts things into groups by fitting a hyperplane, which is ideally an expression of a two-dimensional plane in three-dimensional space, to mathematical spaces with any number of dimensions. We select the hyperplane with the maximum margin, which indicates the distance between the classifier and training datasets. An optimal hyperplane with the maximum margin reduces the generalization error of the overall classifier (Vapnik 1999). The kernel approach follows the principle of SVM, which transforms data into a higher-dimensional space through nonlinear transformation. The strength of SVM lies in its ability to overcome high dimensionality and perform well with a small number of training samples. Several studies have indicated that SVM produces high accuracy for land cover classification and alien species distribution mapping (Kganyago et al. 2018; Mafanya et al. 2022).

Random Forest (RF)

Breiman (2001) defined random forests (RF) as "a tree-based algorithm that depends on the value of an independent random vector sampled for all trees in the forest." RF employs a variety of tree classifications and classifies a new input vector according to the number of trees within the forest. Then, RF assigns a classification to each tree, symbolizing the tree's "votes" for the class with the most frequent input data. During the classification process, the forest prioritized the class with the most "votes" over forest trees (Adelabu et al. 2015). The tree regression-based model, when trained with sufficient field data plots representative of vegetation variability at the national scale, produces satisfactory results. Random forests use bootstrap samples from other trees combined with ensemble regression and tree classification to build binary classifications. Random forest, similar to any other algorithm, has both advantages and disadvantages. Random forest is efficient in implementing large datasets and has an easily saved structure for the future use of pre-generated trees. The algorithm is not sensitive to noise, avoids overfitting, and has high accuracy. It has spectral bands and feature selection layers like NDVI, soil index, and water index. It also has texture features for classification like entropy, variance, morphology, and line features (Chaturvedi and de Vries 2021). The RF algorithm shows enormous potential for solving environmental problems, such as water resources and natural hazard management (Talukdar et al. 2020). Breiman (2001) integrated a random forest with a decision tree algorithm, utilizing both classification and regression trees.

Accuracy assessment

We conducted an accuracy assessment using SVM and RF models to validate the classified Sentinel-2A imagery. The traditional confusion matrix shows the degree of agreement between the classified image and reference ground data for overall accuracy, user accuracy, producer accuracy, and kappa coefficient. Overall accuracy (OA) measures the proportion of accurately classified LULC classes (Petropoulos et al. 2012). The overall kappa coefficient measures the agreement between training and validation datasets. However, many studies have not used the kappa coefficient due to conceptual flaws. However, this study added a kappa coefficient to the assessment of accuracy. Additionally, precision can be evaluated using various machine learning measures, including the F-score (Gidey and Mhangara 2023). We used the F-score to gain a thorough understanding of the performance of both the SVM and RF models. We achieved this by combining precision and recall into a single metric. Equations 1, 2, 3, 4, 5 demonstrate our consideration of the overall accuracy (OA), consumer accuracy (CA), and producer accuracy (PA).

$$CA = x_{ii} \big/ x_{i+} \times 100 \tag{1}$$

$$PA = x_{ii} \big/ x_{+i} \times 100 \tag{2}$$

$$OA = D \big/ V \times 100 \tag{3}$$

$$\mathrm{F-score} = 2 \times (PA \times CA)/(PA + CA) \tag{4}$$

$$\hat{K} = \frac{N\sum_{i=1}^{k} x_{ii} - \sum_{i=1}^{k}(x_{i+} \times x_{+i})}{N^2 - \sum_{i=1}^{k}(x_{i+} \times x_{+i})} \quad (5)$$

where $\hat{K}$ = **K**-coefficient, $\boldsymbol{x_{ii}}$ refers to the total number of observations in both the row$_i$ and the column$_i$, $\boldsymbol{x_{i+}}$ and $\boldsymbol{x_{+i}}$ refers to the respective marginal totals, **N** refers to the total number of observations, and **D** refers to the total number of correct pixels in the diagonal, which is the same as the total number of pixels in the **V** error matrix. The **F-score** is calculated as the harmonic mean of the producer's accuracy (**PA**) and consumer's accuracy (**CA**) (Gidey and Mhangara 2023).

Moreover, we used both Pearson correlation coefficients and regression models to assess the effectiveness of the SVM and RF models using the STATA software version 14.

MaxEnt-based pompom weed distribution modeling

Ecological studies have used species distribution models (SDMs) to study the spatial distribution of alien plants (Mtengwana et al. 2021; Dai et al. 2022). SDM models, as IAPs, are effective in modeling the spatial distribution of alien plants and forecasting their current and future land distributions (Ndlovu and Shoko 2023). If the environmental conditions of an ecosystem invaded by IAPs are the same as those of their native areas, then the conditions necessitate the survival of IAPs and influence further invasion into other areas. This study uses open-access Maximum Entropy Species Distribution Modeling version 3.4.4 for its computer efficiency and ability to calculate environmental variables such as rainfall, temperature, and topography. The software divides a user-defined landscape into grid cells and integrates presence-only data and a sample of captured species locations (Mtengwana et al. 2022; Mkungo et al. 2023). It produces alien species whose habitat suitability ranges from high to low (Ndlovu and Shoko 2023). We employed the MaxEnt species distribution model to explain the spatial distribution of pompom weed, considering selected environmental variables such as (i) elevation, (ii) land cover, (iii) rainfall, (iv) temperature, and (v) soil. We downloaded the environmental variables (rainfall, temperature, soil, and elevation) from NASA Power Data at https://power.larc.nasa.gov/data-access-viewer/. It is important to note that the climatic data are historical from 2019 to 2023 and aggregated monthly. We used the 2024 SVM-classified imagery as the land cover variable, as it demonstrated the highest and most accurate classification. We processed the environmental variable data using ArcGIS, version 10.8.2. We converted the imported environmental variables in ArcMap from raster to ASCII, the American Standard Code for Information Interchange, to ensure compatibility with MaxEnt. We converted 469 pompom weed ground control points to comma-separated values (CSV) and added them to the MaxEnt sample file. The MaxEnt model divided the pompom weed GCPs (samples) into 70% training and 30% validation groups. We used the environmental layer in MaxEnt to import continuous environmental variables. To run ten-fold cross-validation, we set the model to 1 with 500 iterations, and we set the output to the highly cited and recommended "Clog-log" format (Ndlovu et al. 2018; Ndlovu and Shoko 2023).

MaxEnt model evaluatio

We used the AUC to evaluate MaxEnt SDM's performance. The AUC measures the classifier's ability to correctly predict species presence-only data (sensitivity) versus absence (specificity) by comparing actual and predicted species distributions (Mkungo et al. 2023). The AUC rates the model with values between 0.5 and 0.6 as poor, 0.7 and 0.8 as decent, and 0.9 and 1 as excellent model performance prediction. Therefore, we used the AUC to evaluate the performance of the MaxEnt predictive model. MaxEnt uses a jackknife test to assess the efficacy of predictor variables in predicting landscape vulnerability to invasion by IAPs, as well as to produce distinct information on species distribution.

Results

Spatial distribution of *Campuloclinium macrocephalus* (Less) DC in the greater cradle nature reserve for 2019 to 2024

Figures 2 and 3 show The spatio-temporal trends of pompom weed distribution from 2019 to 2024. Figure 2 presents the changes in area coverage by pompom weeds in the Cradle Nature Reserve between 2019 and 2024. The Cradle Nature Reserve total area coverage is 92.38 km2, and in January 2019, the pompom weed area coverage was 31.1 km2 and 29.7.42 km2 using SVM and RF, respectively. The SVM reported 46.84 km2 and RF of 12.0 km2 area coverage for January 2024. The comparison of the SVM and RF models for 2019 and 2024 yielded quite different results. However, the SVM model's analysis of how pompom weed spreads over time across nature reserves matches what we saw when we looked at the weed on satellite images and in the field. The nature reserve experienced a swift surge in pompom weeds between 2019 and 2024, thereby establishing the SVM model as a potent supervised machine learning method. Figure 3 shows the SVM and RF models' performance in mapping the spatial distribution of pompom weed against the co-existing land cover types at the Cradle Nature Reserve between January 2019 and January 2024. In both years, we detected the presence of invasive pompom weeds throughout the nature reserve. We found

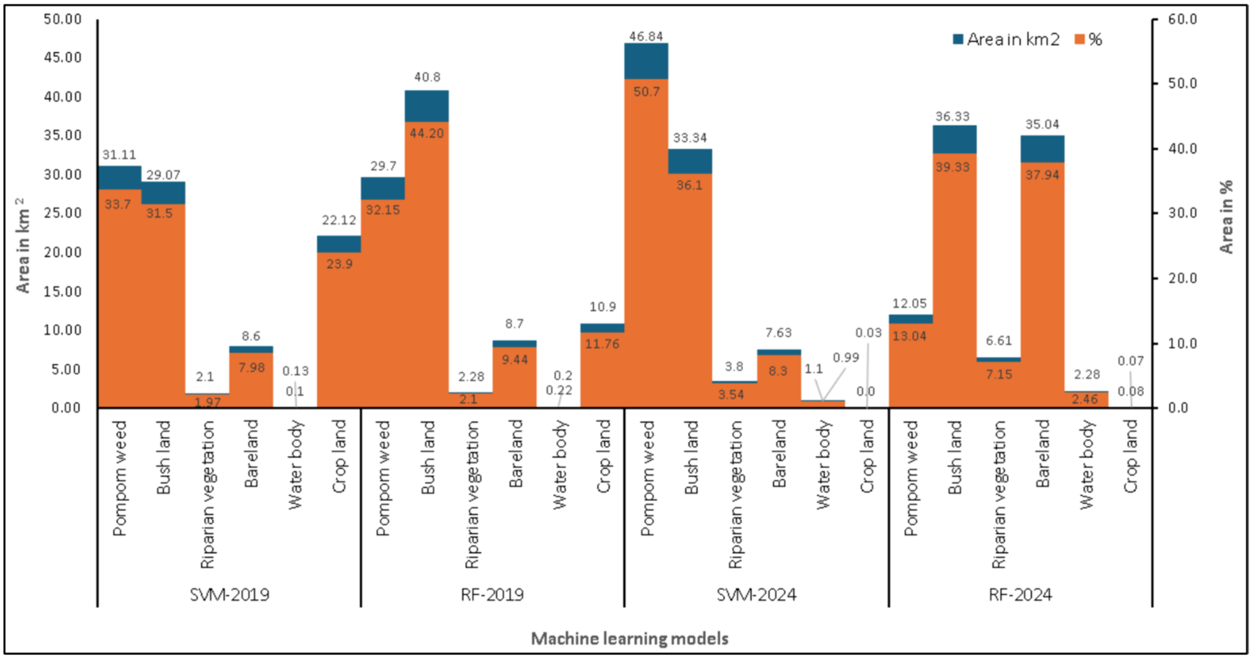

Fig. 2 Pompom weed coverage in the Greater Nature Reserve compared to other land cover types in the study area

that the spatial distribution of pompom weeds increased from 44% in 2019 to 50.7% in 2024 using the SVM model Fig. 4.

In contrast, the RF model indicated that the spatial distribution of pompom weed at the nature reserve was 31.14% in 2019 and increased further to 39.3% in 2024. Pompom weeds exhibit a patchy and heterogeneous presence, invading various land-cover types and vegetation species, with a notable concentration in bushland and riparian zones. We found that the SVM and RF models effectively discriminated and displayed the spatial extent of pompom weeds in the Cradle Nature Reserve. Current observations reveal that uncontrolled exotic pompom weeds heavily invade 50% of the nature reserve, masking it in pink and purple during the spring and summer seasons. The map outputs for both years revealed that previously uninvaded regions, such as the northwestern side of the nature reserve, experienced an expansion of pompom weed, while the former experienced a marginal decline. The encroachment of pompom weeds causes environmental degradation at the nature reserve, leading to a high-risk migration of game animals, reducing biodiversity, and lowering the nature reserve's world-assigned status. This leads to the rapid degradation of potential land (Gidey et al. 2023).

We applied SVM and RF models to illustrate the dynamics of the spatial distribution of pompom weeds (Fig. 5). The SVM model estimated that from 2019 to 2024, there has been an increase in pompom weed presence in the nature reserve; the dominant land cover is pompom weed. In contrast, the RF model estimated that pompom weeds declined. This explains the discrepancies between the two models and highlights each model's weaknesses. However, the SVM model results were consistent with the current visual assessments in the field, indicating a rapid encroachment of pompom weed in the study area. In relation to other land cover types, the RF model indicated a high decline in bushland over the years, and the SVM also indicated a gradual decline in bushland. Land disturbances and pompom weed invasions could explain the decline in bushlands. Both models showed a decrease in cropland during the same periods, with the SVM model indicating a significant decline in cropland. In the RF model, bare land dominates the land cover, whereas in the SVM model, the opposite is observed.

Analysis of pompom weed distributions along various land cover types

The spatial distribution of pompom weeds across other land cover types (Fig. 6) indicates that pompom weeds strongly invaded the bushland and riparian zones (58% and 30%, respectively). In such cases, bushlands are susceptible to pompom weed invasion due to numerous factors, such as bush thickening and land disturbances in terms of road construction. The excessive encroachment of invasive plants causes the infectious bush to thicken,

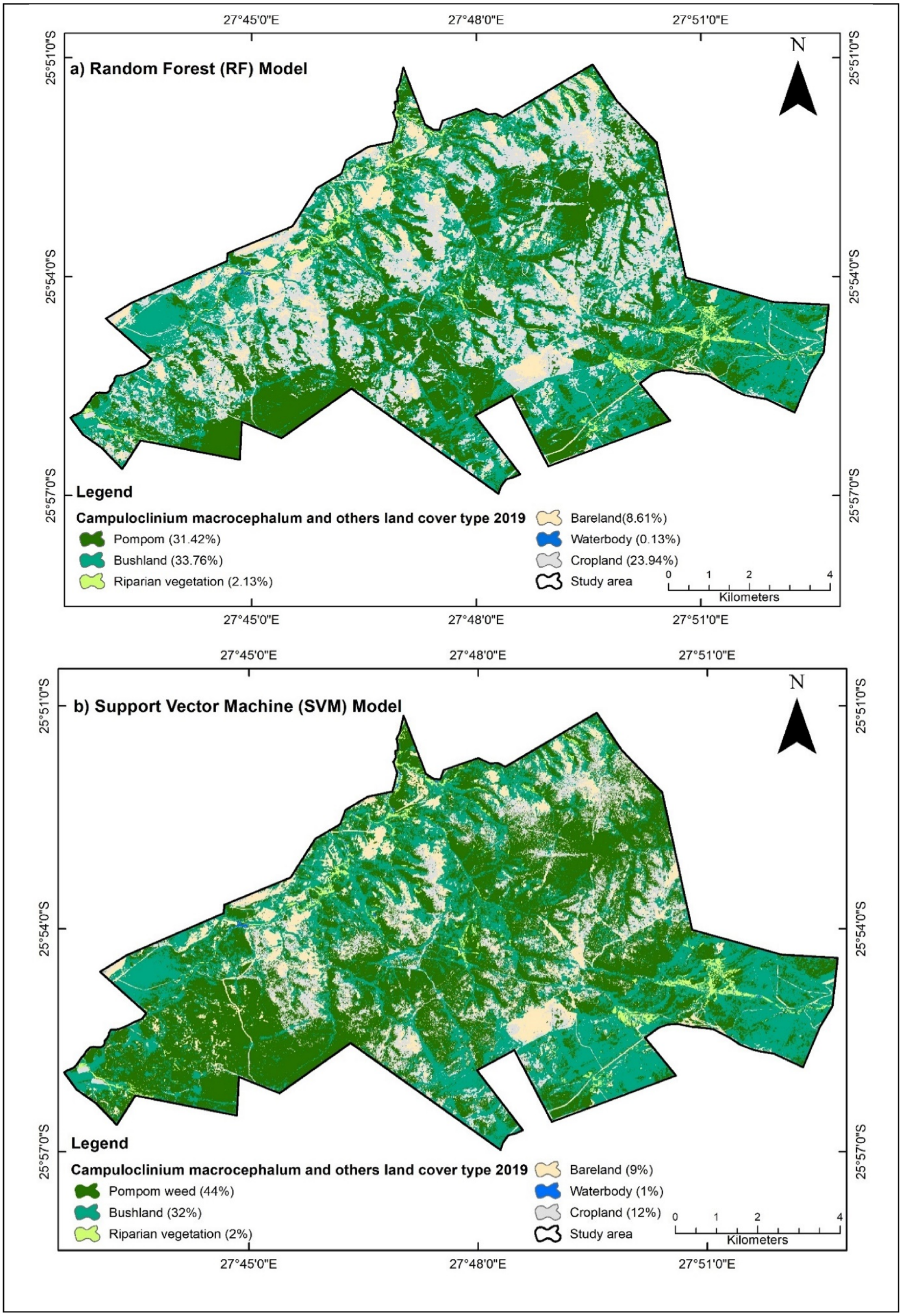

Fig. 3 RF (**a**) and SVM (**b**) model based spatial distribution of pompom weed at Cradle nature reserve for 2019

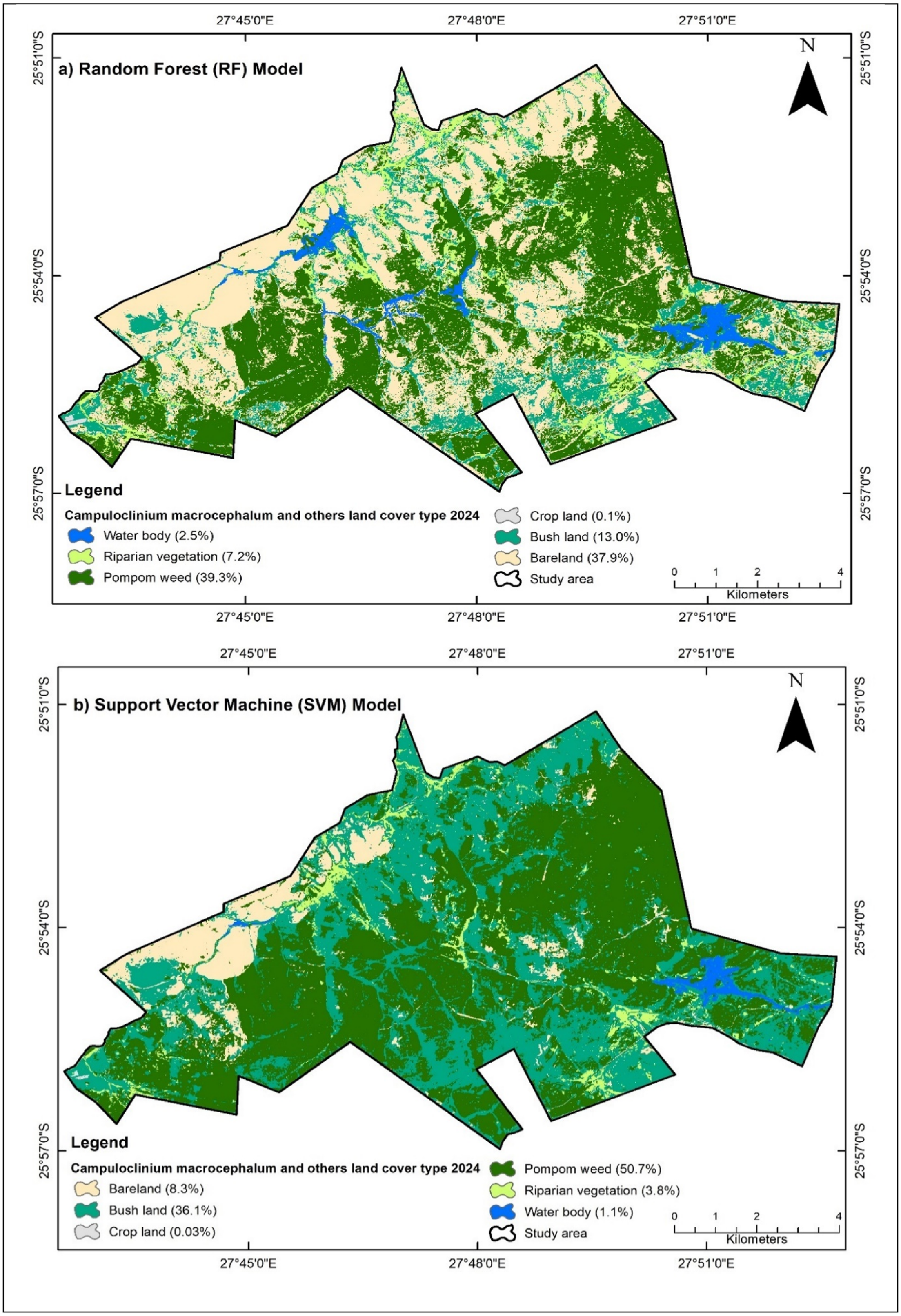

Fig. 4 RF (**a**) and SVM (**b**) model based spatial distribution of pompom weed at Cradle nature reserve for 2019

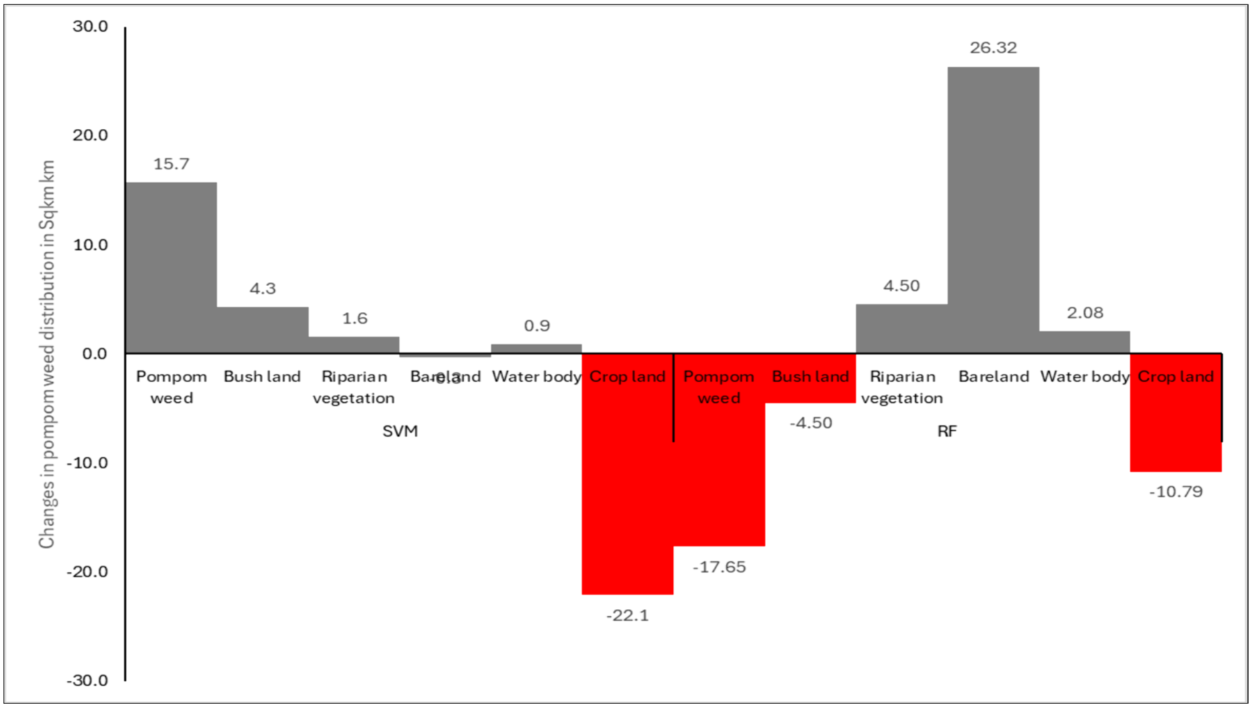

Fig. 5 Showing temporal changes in pompom weed distribution from 2019 to 2024

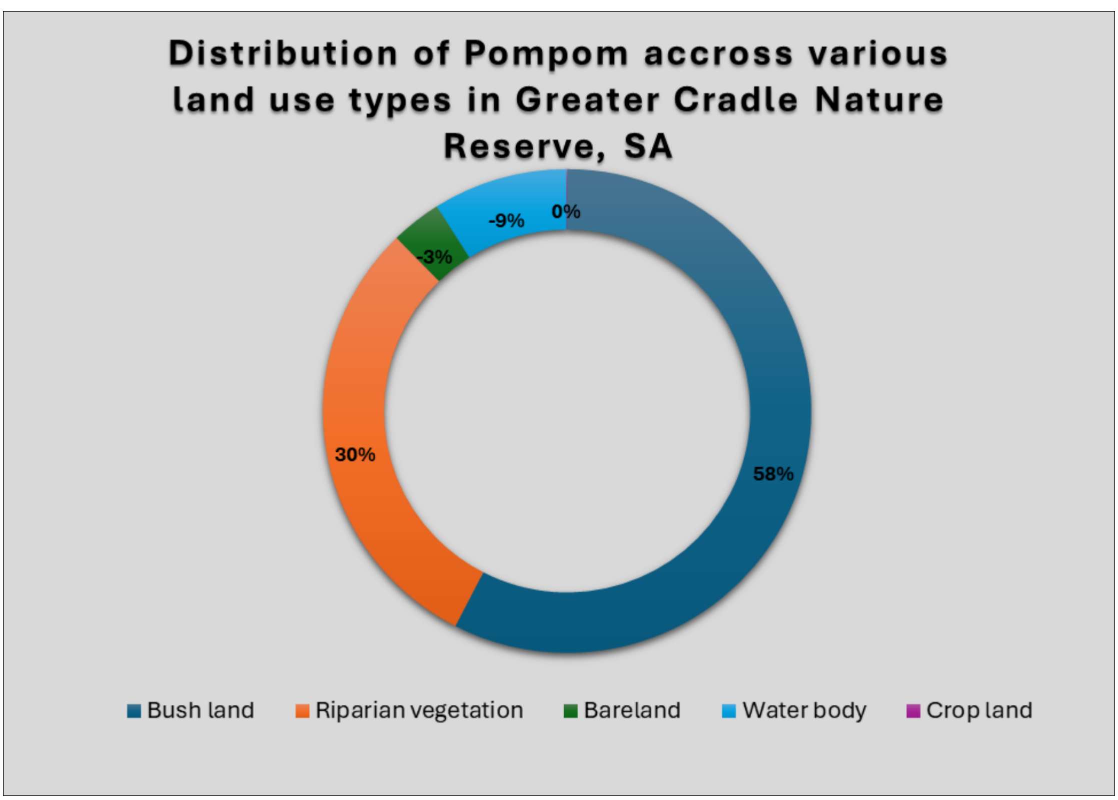

Fig. 6 Distribution of pompom weed across various land cover types in the nature reserve

making it a serious threat to savanna rangelands. Riparian vegetation grows along river streams with its natural pompom weed and invades wetlands (Gooddall et al. 2011); therefore, riparian vegetation is also susceptible to pompom weed invasion. Pompom weed did not infiltrate 0% of the cropland, whereas it only affected 3% of the bare land.

SVM and RF models classification accuracy and confusion matrix assessment

We performed accuracy assessments for both 2019 and 2024 Sentinel-2A classified imagery using the popular confusion matrix; Tables 2, 3, 4, 5, present the results. We manually calculated the overall accuracy by taking the sum of correctly classified values and dividing it by the total number of values. Despite the criticisms of redundancy and misleadingness in remote sensing applications

Table 2 Confusion matrix for SVM 2024

Land use type	Pompom weed	Bush land	Riparian vegetation	Bare land	Water body	Crop land	Total	U-accuracy	F-score	Kappa
Pompom weed	30	0	0	0	0	0	30	0.97	0.92	–
Bushland	1	29	0	0	0	0	30	0.96	0.96	–
Riparian	1	0	29	0	0	0	30	0.97	0.94	–
Bareland	0	1	1	28	0	0	30	0.93	0.96	–
Waterbody	0	0	1	0	29	0	30	0.97	0.98	–
Cropland	2	0	1	0	0	27	30	0.90	0.95	–
Total	34	30	32	28	29	27	180	–	–	–
P-Accuracy	0.88	0.97	0.91	1	1	1	0	0.96	–	–
Kappa	–	–	–	–	–	–	–	–	–	0.94

Table 3 Confusion matrix for RF 2024

Land use type	Pompom weed	Bush land	Riparian vegetation	Bare land	Water body	Crop land	Total	U-Accuracy	F-score	Kappa
Pompom weed	28	1	0	0	0	0	30	0.96	0.85	–
Bushland	1	29	0	0	0	0	30	0.95	0.94	–
Riparian vegetation	0	1	27	0	2	0	30	1	1.00	–
Bareland	8	2	0	20	0	0	30	0.65	0.79	–
Waterbody	0	0	5	0	25	0	30	1.00	1.00	–
Cropland	0	0	0	0	0	30	30	1.00	1.00	–
Total	39	31	30	20	30	30	180	–	–	–
P-Accuracy	0.76	0.93	1	1	1	1	0	0.94	–	–
Kappa	–	–	–	–	–	–	–	–	–	0.92

Table 4 Confusion matrix for SVM model 2019

Land use type	Pompom weed	Bush land	Riparian vegetation	Bare land	Water body	Crop land	Total	U-Accuracy	F-Score	Kappa
Pompom weed	26	2	1	0	0	1	30	0.94	0.83	–
Bushland	2	28	0	0	0	0	30	1.00	0.82	–
Riparian	0	1	24	0	5	0	30	1.00	1.00	–
Bareland	0	2	1	26	0	1	30	1.00	1.00	–
Waterbody	0	2	5	0	23	0	30	1.00	1.00	–
Cropland	0	7	0	5	0	18	30	0.58	0.73	–
Total	28	42	31	31	28	20	180	–	–	–
P-Accuracy	0.8	0.7	1	1	1	1	0	0.90	–	–
Kappa	–	–	–	–	–	–	–	–	–	0.91

Table 5 Confusion matrix for RF model 2019

Land use type	Pompom weed	Bush land	Riparian vegetation	Bare land	Water body	Crop land	Total	U-accuracy	F-score	Kappa
Pompom weed	26	2	1	0	0	1	30	0.94	0.83	–
Bushland	2	28	0	0	0	0	30	1.00	0.81	–
Riparian	0	1	24	0	5	0	30	1.00	1.00	–
Bareland	0	2	1	26	0	1	30	1.00	1.00	–
Waterbody	0	2	5	0	23	0	30	1.00	1.00	–
Cropland	0	7	0	5	0	18	30	0.58	0.73	–
Total	28	42	31	31	28	20	180	–	–	–
P-Accuracy	0.75	0.68	1	1	1	1	0	0.92	–	–
Kappa	–	–	–	–	–	–			–	0.9

(Pontius and Millones 2011), we include both the user's and producer's accuracy, as well as the kappa coefficient. We rank the Kappa coefficient as follows: we consider any value from 0 to 0.4 to be moderate, any value from 0.4 to 0.8 to be substantial agreement, and any value above 0.8 to be excellent agreement. In addition, the F-score values for the same period and model indicate that they exceed the minimum limit thresholds, i.e., 0.5 or 50%. All our findings in this case exceed the minimum standards. The F-score values for the pompom and waterbody range from 0.92 to 0.98, respectively.

The SVM model achieved a kappa coefficient of 0.94 for the classification accuracy of 2024 classified imagery, indicating excellent agreement with Table 2. The SVM model successfully classified pompom weed, achieving user and producer accuracy of 0.96 and 0.88, respectively, compared to other land cover types. The bushland and riparian vegetation obtained high accuracies above 70%, and the overall classification accuracy for detecting pompom weed using the SVM model was 95%. The RF model achieved an overall classification accuracy of 93% when detecting pompom weeds in 2024 classified imagery. The Kappa coefficient was 0.92, user accuracy was 0.94, and producer accuracy was 0.76, as shown in Table 3. Using RF, we accurately classified the water body with no spectral confusion in 2024 imagery, compared to other land cover types. The classification accuracy for the year 2024 proved successful in obtaining high accuracy; additionally, the SVM outperformed RF. The F-score obtained using the RF model ranged from 0.79 to 1.00. The riparian vegetation, water bodies, and crop land cover types exhibited the highest F-scores, reaching 1.00. Nevertheless, in bare land, pompom weed, and bush land, the values were 0.79, 0.85, and 0.94, respectively.

The overall classification accuracy for 2019 using the SVM model was 94%, with a kappa coefficient of 0.91 (Table 4). The user's accuracy and the producer's accuracy were 0.92 and 0.75, respectively. A 2019 study using the RF model found that the overall classification accuracy was 80%, with kappa coefficients of 0.90 for producer accuracy and 0.92 for user accuracy (Table 5). In summary, the SVM model performed better than the RF model in accurately detecting pompom weeds against co-existing land cover types in 2019 and 2024. The F-score values acquired through SVM varied between 0.73 and 1.00 for each land cover category. The vegetation types that exhibited the highest F-score values (1.00) were riparian, barren land, and water bodies. In contrast, the F-score values for cropland, bushland, and pompom weed were 0.73, 0.82, and 0.83, respectively (Table 4). Conversely, the F-scores produced by the RF model exhibited similar outcomes. This varies between 0.73 and 1.00. The F-score values for riparian vegetation, bareland, and water bodies were the highest (i.e., 1.00), whereas the values for crop land, bush land, and pompom weed were comparatively lower (0.73, 0.82, and 0.83, respectively) than those of the remaining land cover categories (Table 5).

SVM and RF model testing

We assessed the effectiveness of the SVM and RF models using a dataset of 120,659 samples. These samples included pompom weeds and other types of land cover within a 2-km radius of the center of nature reserves. Our evaluation considers the computational and processing capabilities of a desktop computer. We use samples to evaluate the identification proficiency of each model. Pearson's correlation coefficients and regression models were used to assess the efficacy of the SVM and RF models. The Pearson correlation coefficients indicated that both the SVM and RF models accurately identified pompom weeds, with a correlation coefficient of 0.672. The correlation coefficient was statistically significant at a p-value of 0.00. Furthermore, a highly significant p-value of 0.000 supported a robust positive linear correlation between RF and SVM in the regression model. Both the

Source	SS	df	MS			
				Number of obs	=	120,659
				F(1, 120657)	=	99495.39
Model	17362.9622	1	17362.9622	Prob > F	=	0.0000
Residual	21055.8799	120,657	.174510222	R-squared	=	0.4519
				Adj R-squared	=	0.4519
Total	38418.8421	120,658	.318411063	Root MSE	=	.41774

rf	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
svm	.5232784	.0016589	315.43	0.000	.5200269	.5265299
_cons	.0508302	.0012891	39.43	0.000	.0483036	.0533568

Fig. 7 Model validation results outputs using STATA v.14 software

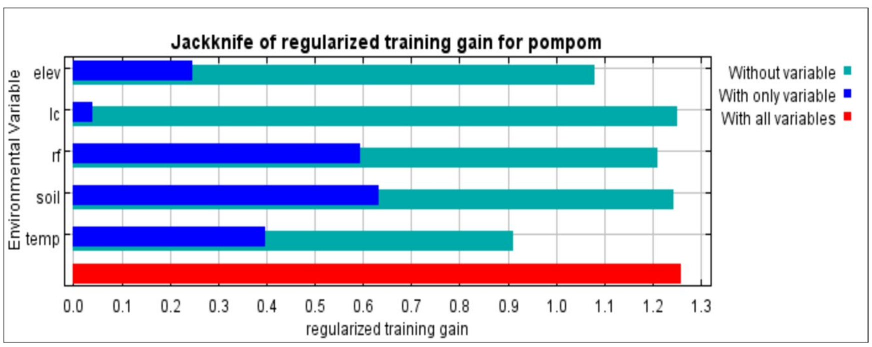

Fig. 8 Jackknife results in training gain for pompom weed. *Note: Elevation (ele); Land cover (lc); Rainfall (rf); soil and Temperature (temp)

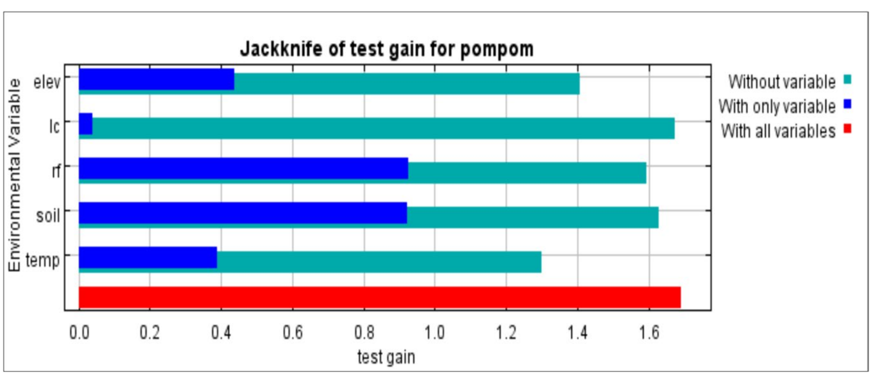

Fig. 9 Jackknife results from test gain for pompom weed

constant term (_cons) and intercept exhibit statistical significance, indicating that the intercept is not equal to zero (Fig. 7).

MaxEnt model-based analysis of pompom along various environmental variables

Figures 8, 9, 10 display the results of the jackknife test for variable significance. When running in isolation, the soil is the environmental variable with the highest training gain (Fig. 8). This means that it provides useful

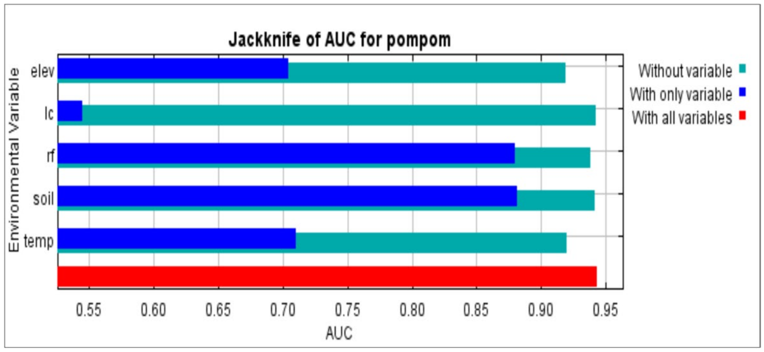

Fig. 10 Jackknife test using AUC on test data

Table 6 Estimates relative contributions of environmental variables

Environmental variables	Percent contribution	Permutation importance
Soil	44.8	39.3
Temp	29.8	30.3
Rf	14.3	21.2
Elev	10.7	8.7
Lc	0.4	0.5

information for explaining the study area's vulnerability to pompom weed invasion. After soil, rainfall and temperature were the environmental variables with the highest gain. However, omitting the temperature reduces the gain, indicating that it contains more information than the other environmental variables. Elevation and land cover were the environmental variables with the lowest training gains.

The jackknife test results in Fig. 8 show that soil and rainfall are still the environmental variables with the highest test gain, which holds the most valuable information for determining the area's susceptibility to pompom weed invasion. In contrast to the results from the Jackknife regularized training gain, the Jackknife test gain identified elevation as the third contributing environmental variable. Land cover remained the least contributing environmental variable Fig. 9.

Finally, Fig. 10 shows the jackknife test results using the pompom AUC. We selected soil, rainfall, and temperature as environmental variables that yielded the greatest gains. The results of our analysis indicate that soil, rainfall, and temperature are the most influential environmental variables affecting the spatial distribution of pompom weeds in nature reserves. Soil contains the most valuable information for explaining the pompom weed invasion. The nature reserve's soil nutrient composition creates a favorable environment for the pompom weed to thrive. Elevation and land cover contributed less to the pompom weed infestation.

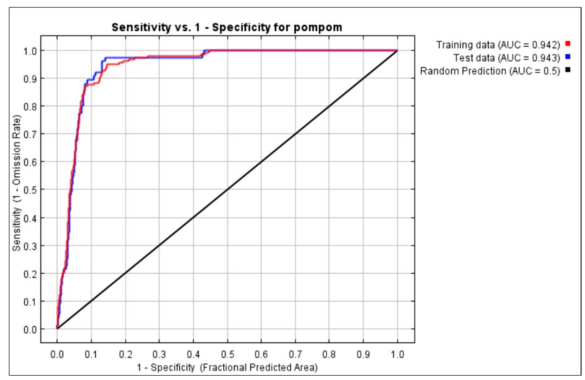

Fig. 11 The AUC curve results in predicting pompom weed habitat suitability

Analysis of each environmental variable contributions to pompom

Table 6 presents the estimated relative contributions of the chosen environmental variables trained using the MaxEnt model. These estimates show how each environmental variable affects the suitability of the habitat for pompoms. Soil was the leading environmental variable (44.8%), which explains why the soil and nutrient composition of the study area are favorable for pompom weed growth. Temperature and rainfall also showed high percent contributions of 29.8 and 14.3, respectively. Elevation and land cover had fewer contributions; therefore, they were less significant.

MaxEnt model performance evaluation

We assessed the MaxEnt model's performance by calculating its AUC. We consider AUC model rates between 0.5 and 0.6 as poor, 0.7 to 0.8 as decent, and 0.9 to 1 as excellent. Figure 11 shows the results of the area under the receiver operating characteristic curve (AUC). We used selected environmental variables and presence-only data (pompom weed samples) to predict the distribution of pompom weeds and habitat suitability. We achieved an AUC score of 0.94 with training and test data to accurately predict the spatial distribution of pompom weed in the study area for the year 2024. This study regarded 0.94 as an excellent model performance prediction.

Spatial distribution of pompom weed based on a maximum predictive model

The MaxEnt model predicts the growth locations of pompom weeds based on the selected environmental variables, as shown in Fig. 12. The presence of pompom weeds dominates the central part of the nature reserve; the lower southern part of the nature reserve is highly characterized by pompom weed presence, and this is where the locations of pompom weeds were collected during the field survey. This explains why the MaxEnt model predicted that the central, southern, and eastern tips of the nature reserve would have high habitat suitability for pompom weeds. The western region showed low habitat suitability. However, parts that exhibit low suitability and are adjacent to the central part of the nature reserve are at risk of future invasion. The MaxEnt predictive model's results suggest prioritizing parts of the nature reserve with high pompom weed for effective weed control and environmental management.

Responsive curves

In Fig. 13, the response curve graphs indicate how the individual environmental variables affect the maximum prediction. The curves illustrate the changes in the estimated probability of the adjusted environmental variables, while maintaining the average sample value for all other environmental variables.

In contrast to the response curve graphs in Fig. 14, each of the following curves indicates a MaxEnt model created using only individual environmental variables. These plots reflect the predicted suitability's dependence on the selected variable, as well as the dependencies induced by correlations between the selected variable and other variables.

Discussions

Application of SVM and RF models in mapping the spatial distribution of pompom weed

Invasion by alien plants is considered the second-largest threat to biodiversity (Newete et al. 2023). They outcompete native vegetation in terms of space, nutrients, and water retention (Mafanya et al. 2022). To implement control measures and minimize their impact on the environment, precise data on the spatial distribution of IAPs are still lacking (Kganyago et al. 2018). We adopted remote sensing technology to accurately monitor and map the spatial distribution of invasive weeds. The temporal resolution of remote sensing technology gives users the advantage of obtaining historical information. We can obtain historical data to monitor IAPs, forecast their future distribution, and develop effective eradication methods. The current study used the Sentinel-2A MSI product to successfully show how invasive pompom weeds spread across the Cradle Nature

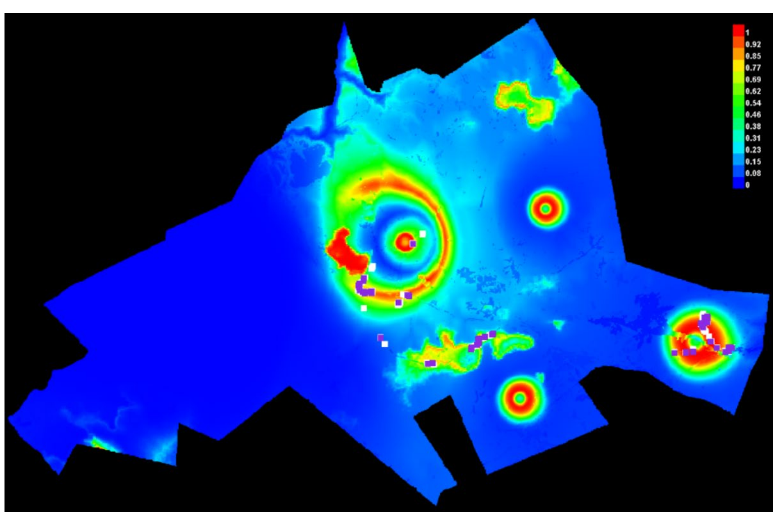

Fig. 12 Spatial distribution of pompom weed using Maxent predictive model

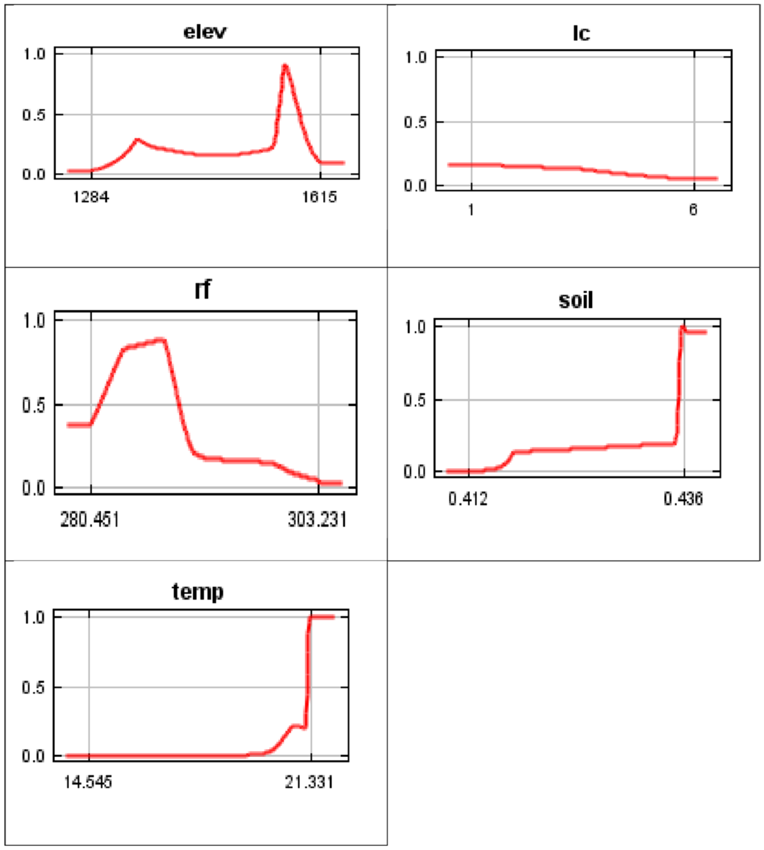

Fig. 13 Responsive curve of various environmental variables

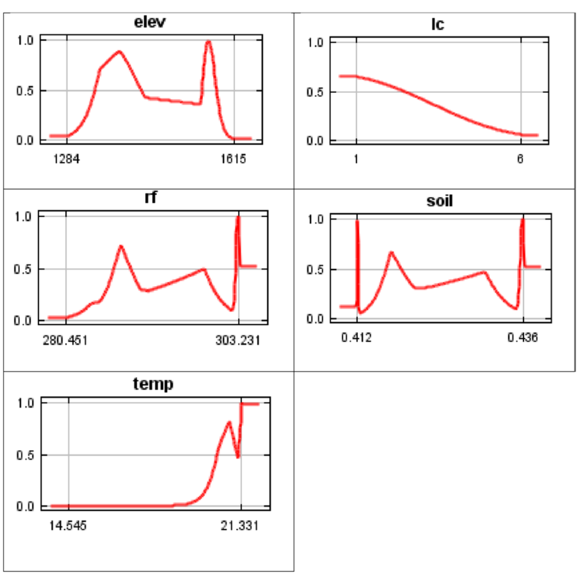

Fig. 14 MaxEnt model responsive curve of each environmental variable

Reserve from 2019 to 2024. This is possible with Band 4: Red (665 nm), Band 3: Green (560 nm), and B2: Blue (490 nm), which come with the product. Sentinel-2A's red band effectively maps vegetation and provides detailed 10 m spatial resolution. Sentinel-2A's red edge band played a key role in pompom-weed detection. The adoption of SVM and RF models was significant in discriminating pompom weeds from existing land cover types. There were slight differences between the two models; however, the overall accuracy was statistically significant. The results of this study align with those of previous studies that evaluated the effectiveness of machine learning models in mapping invasive alien plants (Mafanya et al. 2022; Kganyago et al. 2018; Ndlovu and Shoko 2023), despite SVM's overall superior accuracy over RF in identifying the locations of pompom weeds.

Findings of this study indicate that invasive pompom weeds heavily invaded the nature reserve. Pompom weed encroachment into nature reserves has been increasing for the periods chosen for this study. The 2019 imagery showed that pompom weed was present in the nature reserve; the presence of pompom weed was 44% using the SVM model, and 31.1% using the RF model. For the current year (2024), the SVM indicates 50.7% encroachment of pompom weed in the nature reserve, which is an increase of 44% in 2019. The RF model predicted this to be 39.3% in 2024. The presence of pompom weeds in nature reserves has increased, and this trend is likely to continue into the future. Therefore, we should implement effective seasonal monitoring and eradication management strategies. Given the current land cover types in nature reserves, pompom weed has significantly infiltrated bushland. Approximately 58% of bushlands are susceptible to pompom weed infestation. This poses environmental and economic challenges, reducing the grazing capacity of game animals, causing their migration, and jeopardizing tourism. The construction of roads in nature reserves for game vehicles is a form of land disturbance that results in bush thickening. The densification of alien plants causes a serious environmental problem known as infectious bush thickening (Kellner 2020). In nature reserves, the encroachment of bush thickening replaces the native vegetation; in turn, pompom weeds take advantage of this and grow rapidly with less competition. Bush thickening thus led to a significant invasion of bushland by pompom weed. Pompom weeds heavily affected other land cover types, such as riparian zones (30%). Other land cover types, such as bare land, water bodies, and cropland, were less affected by pompom weeds. Pompom weeds may not invade crops due to the use of herbicides or chemicals in agriculture.

In this study, the use of ESA Sentinel-2A imagery was successful in distinguishing the presence of pompom weeds from the existing land cover types. Overall, the study yielded an accuracy of > 75% for both years. Previous studies that used Sentinel-2A to map the spatial distributions of alien plants can supplement this study's findings. Ndlovu and Shoko (2023) used Sentinel-2A imagery to map the spatial distribution of L. camara and differentiate it from other LULC types. They obtained an overall classification accuracy of 90.27% using the RF model on Sentinel-2A data. Newete et al. conducted a different study in 2023, using Sentinel-2A images and RF and SVM algorithms to determine the locations of invasive genotypes and their relationship to diverse types of land cover in the Leeu, Swart, and Olifant River valleys of the Western Cape Province. The utilization of Sentinel-2A imagery proved to be successful in mapping invasive species, resulting in an impressive overall classification accuracy of 85%.

Modeling potential distributions of pompom weed using environmental variables in the Maxent species distribution model

This study found that the MaxEnt predictive model worked effectively with a small sample of data, providing robust and accurate estimations. Previous studies (Mtengwana et al. 2021; Dai et al. 2022; Ndlovu and Shoko 2023; Mkungo et al. 2023), which used the Maxent model to predict species distribution, agree with the results of this study. We chose historical environmental variables (like temperature, rainfall, soil type, and elevation) and land cover from the 2024 classified Sentinel-2A imagery to look into how they affected the spread of pompom weed in Cradle Nature Reserve. Research has consistently shown that environmental and climatic variables significantly influence the distribution of invasive species (Ncube et al. 2020; Ndlovu et al. 2018; Ndlovu and Shoko 2023). This study used the environmental variable in the MaxEnt species distribution model to predict areas in the nature reserve that are suitable for pompom weeds. The AUC for this study's predictions was 0.94, which is high compared to other studies in Southern Africa that used the MaxEnt species distribution model (Mtengwana et al. 2021; Mkungo et al. 2023; Ndlovu and Shoko 2023), and the same model. The model predicted a concentration of pompom weed in the central and southern regions, as well as the eastern tip of the nature reserve. We consider these parts of the nature reserve to be high-quality habitats for the pompom weeds. According to Jackknife's MaxEnt model results, soil, rainfall, and temperature all have an impact on pompom weed growth. These three environmental factors provided significant insights into weed initiation and rapid spread within the study area. We identified soil as the environmental variable that had the greatest influence on the establishment of pompom weeds. Nature reserves' soil chemistry encourages weeds to thrive. Therefore, we recommend researching the soil composition of nature reserves. Rainfall also contributes to infestation by pompom weeds. Rainfall is a significant climatic variable that promotes the growth and spatial distribution of pompom weeds (Mtengwana et al. 2021).

The highveld moist conditions are suitable for pompom weed, and the Cradle Nature Reserve, which is in the highveld region, receives rainfall during the spring and summer seasons. The wettest seasons, when the nature reserve receives rainfall, provide moisture to the pompom weed, which is important for its germination and growth. During the rainfall season, the availability of water or moisture in the soil promotes the germination and establishment of pompom weeds. Gooddall et al. (2011) have observed the growth of pompom weeds along roadsides. Land disturbance during road construction and carbon feeding from vehicles can explain the infestations of pompom weeds along roadsides. We recommend conducting studies on the effects of carbon on pompom weeds.

Environmental impacts of pompom weed at the cradle nature reserve

UNESCO has granted cradle nature reserve world status, making it a protected and sensitive nature reserve. The introduction of pompom weed into nature reserves has resulted in adverse impacts on the environment, society, and economy of Gauteng Province. The proliferation of pompom weeds has an impact on nature reserves' biodiversity. The invasion of pompom weeds decreases grazing capacity and water content. The invasive pompom weed grows unaided and replaces the natural vegetation. The predictive model results indicate that pompom weed invasions in nature reserves will continue to grow in the future. Implementation of preventive measures will determine the outcomes. The implementation of integrated environmental management approaches is necessary to curb the rapid spread of pompom weeds in nature reserves. In nature reserves, the current methods of eradication and weed control are mechanical, including uprooting and chemical spraying. To prevent further spread, the predictive model recommends prioritizing the currently used eradication methods in areas highly suitable for pompom weeds. Community engagement and raising awareness about invasive species could inform members of their impact.

Conclusions

This study investigated the ability of Sentinel-2A to map the spatial distribution of pompom weeds in Cradle Nature Reserve. SVM and RF were effective in accurately detecting pompom weeds against existing land cover. The findings revealed that the spatial distribution of invasive pompom weeds will increase between 2019 and 2024. Current observations and model estimations indicate that the presence of pompom weeds is worse than that in previous years. Under current environmental conditions, the number of pompom weeds may increase in the future. We investigated this using the MaxEnt species distribution model to assess which environmental variables significantly support the germination and flowering of pompom weeds in the study area. The model indicated that the soil was the most significant variable influencing the distribution of pompom weeds. The findings of this study revealed the robustness and capability of MaxEnt SDM in predicting habitat suitability. Environmental practitioners and conservationists can use the findings of this study to implement effective eradication methods, monitor areas that are susceptible to pompom weeds, and formulate the best environmental practices. The findings of this study recommend further research to determine whether the presence of carbon from automobiles causes pompom weed infestations along roadsides. We recommend investigating the efficacy of eradication measures at nature the reserves, starting with a five-year change detection period.

Acknowledgements
The University of the Witwatersrand, School of Geography, Archaeology, and Environmental Studies, Johannesburg, South Africa, supported this research.

Author contributions
P.M, B.M, E.G; Conceptualisation and methodology: B.M. wrote the draft manuscript text supported by E.G, P.M. E.G., P.M, and M.K. critically reviewed and improved the manuscript to the standard of publication.

Funding
We are grateful for the fieldwork funding provided by Genus DSI-NRF Centre of Excellence in Palaeosciences, University of the Witwatersrand, South Africa.

Availability of data and materials
Data will be available upon request to the corresponding author.

Declarations

Ethics approval and consent to participate
Not applicable.

Competing interests
The authors declare no competing interest.

Received: 6 April 2024 Accepted: 7 June 2024
Published online: 10 July 2024

References

Adelabu S, Mutanga O, Adam E (2015) Testing the reliability and stability of the internal accuracy assessment of random forest for classifying tree defoliation levels using different validation methods. Geocarto Int 30(7):810–821
Akalu F, Raude JM, Sintayehu EG, Kiptala J (2019) Evaluation of land use and land cover change (1986–2019) using remote sensing and GIS in Dabus Sub-Catchment, Southwestern Ethiopia. J Sustain Res Eng 5(2):91–100
Al-dowski J, Mansor SB, San H, Khuzaimah Z (2020) Land cover mapping using remote sensing data. Am J Geograph Inform Syst 9(1):33–45
Breiman L (2001) Random forests. Machine Learn 45(1):5–32
Chaturvedi V, de Vries WT (2021) Machine learning algorithms for urban land use planning: a review. Urban Science 5(3):68
Civco DL (1993) Artificial neural networks for land-cover classification and mapping. Int J Geogr Inf Sci 7(2):173–186
Dai X, Wu W, Ji L, Tian S, Yang B, Guan B, Wu D (2022) MaxEnt model-based prediction of potential distributions of Parnassiawightiana (Celastraceae) in China. Biodiversity Data J 10
Durand JF, Meeuvis J, Fourie M (2010) The threat of mine effluent to the UNESCO status of the Cradle of Humankind World Heritage Site TD. JTransdiscipl Res Southern Afr 6(1):73–92
Gebregergs T, Teka K, Taye G, Gidey E, Dikinya O (2021) Status and challenges of integrated watershed management practices after-project phased-out in Eastern Tigray, Ethiopia. Modeling Earth Systems and Environment 1–7.

Gidey E, Mhangara P (2023) An application of machine-learning model for analyzing the impact of land-use change on surface water resources in Gauteng Province. South Africa Remote Sensing 15(16):4092
Gidey E, Dikinya O, Sebego R, Segosebe E, Zenebe A, Mussa S, Birhane E (2023) Land use and land cover change determinants in Raya Valley, Tigray Northern Ethiopian Highlands. Agriculture 13(2):507
Goodall J, Witkowski ET, Morris CD, Henderson L (2011) Areenvironmental factors important facilitators of pompom weed (*Campuloclinium macrocephalum*) invasion in South African rangelands? Biol Inv 13:2217–2231
Katternboorn T, Lopatin J, Förster M, Braun AC, Fassnacht FE (2019) UAV data as alternative to field sampling to map woody invasive species based on combined Sentinel-1 and Sentinel-2 data. Remote Sens Environ 227:61–73
Kganyago M, Odindi J, Adjorlolo C, Mhangara P (2018) Evaluating the capability of Landsat 8 OLI and SPOT 6 for discriminating invasive alien species in the African Savanna landscape. Int J Appl Earth Obs Geoinf 67:10–19
Mafanya M, Tsele P, Zengeya T, Ramoelo A (2022) An assessment of image classifiers for generating machine-learning training samples for mapping the invasive *Campuloclinium macrocephalum* (Less.) DC (pompom weed) using DESIS hyperspectral imagery. ISPRS J Photogramm Remote Sens 185:188–200
McConnachie AJ, Retief E, Henderson L, Mc Kay F (2011) The initiation of a biological control programme against pompom weed, *Campuloclinium macrocephalum* (Less) DC (*Asteraceae*) South Africa. Afr Entomol 19(1):258–268
Miranda, E., Mutiara, A.B. and Wibowo, W.C. (2018). Classification of land cover from Sentinel-2 imagery using supervised classification technique (preliminary study). In *2018 International Conference on Information Management and Technology (ICIMTech)* pp 69–74
Miller JA (2012) Species distribution models: Spatial autocorrelation and nonstationarity. Progress in Phys Geog 36(5):681–692
Mkungo L, Odindi J, Mutanga O, Matongera TN (2023) Modelling landscape vulnerability to the Bracken fern (Pteridium aquilinum) invasion in a remnant urban Sandstone Sourveldt grassland ecosystem. Scientific African 22:e01947
Mtengwana B, Dube T, Mudereri BT, Shoko C (2021) Modeling the geographic spread and proliferation of invasive alien plants (IAPs) into new ecosystems using multi-source data and multiple predictive models in the Heuningnes catchment, South Africa. GISci & Remote Sens 58(4):483–500
Ncube B, Shekede MD, Gwitira I, Dube T (2020) Spatial modelling the effects of climate change on the distribution of Lantana camara in Southern Zimbabwe. Appl Geog 117:102172
Ndlovu H, Shoko C (2023) Assessing Lantana camara spatial distribution in response to rainfall and temperature variations in Inkomati catchment in Mpumalanga. Phys Chemis Earth Parts A/B/C 132:103470
Ndlovu P, Mutanga O, Sibanda M, Odindi J, Rushworth I (2018) Modelling potential distribution of bramble (rubus cuneifolius) using topographic, bioclimatic and remotely sensed data in the KwaZulu-Natal Drakensberg, South Africa. Appl Geog 99:54–62
Newete SW, Mayonde S, Kekana T, Adam E (2023) A rapid and accurate method of mapping invasive Tamarix genotypes using Sentinel-2 images. PeerJ 11:e15027
Peerbhay K, Mutanga O, Lottering R, Bangamwabo V, Ismail R (2016) Detecting bugweed (Solanum mauritianum) abundance in plantation forestry using multisource remote sensing. ISPRS J Photogrammetry Remote Sens 121:167–176
Pontius Jr RG, Millones M (2011) Death to Kappa: birth of quantitydisagreement and allocation disagreement for accuracy assessment. Int J Remote Sens 32(15):4407–4429
Preston IR, Le Maitre DC, Blignaut JN, Louw L, Palmer CG (2018) The impacts of invasive alien plants on water provision in selected catchments. Water. https://doi.org/10.4314/wsa.v44i4.20
Royimani L, Mutanga O, Odindi J, Dube T, Matongera TN (2019) Advancements in satellite remote sensing for mapping and monitoring of alien invasive plant species (AIPs). Phys Chem Earth Parts A/B/C 112:237–245
Saini R, Ghosh SK (2021) Crop classification in a heterogeneous agricultural environment using ensemble classifiers and single-date Sentinel-2A imagery. Geocarto Int 36(19):2141–2159
Stratford D, Merlo S, Brown S (2016) The development of a new geospatial framework for the palaeoanthropological site of the Sterkfontein Caves, Cradle of Humankind, Gauteng, South Africa. J Field Archaeol 41(2):211–221
Talukdar S, Singha P, Mahato S, Pal S, Liou YA, Rahman A (2020) Land-use land-cover classification by machine learning classifiers for satellite observations—a review. Remote Sensing 12(7):1135
Wong SN, Sarker MLR (2014) Land use/land cover mapping using multi-scale texture processing of high-resolution data in earth and environmental science. IOP Publishing, Bristol
Xie Q, Dash J, Huete A., Jiang A., Yin G, Ding Y, Huang W (2019) Retrieval of crop biophysical parameters from Sentinel-2 remote sens imagery. Int App Earth Obs Geoinf 80:187–195
Xu B (2021) Improved convolutional neural network in remote sensing image classification. Neural Comput Appl 33(14):8169–8180
Zhang H, Wang L, Tian T, Yin J. 2021. A review of unmanned aerial vehicle low-altitude remote sens (UAV-LARS) use in agricultural monitoring in China. Remote Sens 13(6):1221

Publisher's Note