Leveraging High-Resolution Satellite Imagery and Gradient Boosting for Invasive Weed Mapping

Yuri Shendryk, *Member, IEEE*, Natalie A. Rossiter-Rachor, Samantha A. Setterfield, and Shaun R. Levick

***Abstract*—An introduced pasture grass (*Andropogon gayanus* - gamba grass) is spreading through the tropical savannas of northern Australia, with detrimental ecosystem consequences including increased fire intensity. In order to monitor and manage the spread of gamba grass, a scalable solution for mapping its distribution over large areas is required. Recent developments in machine learning have proven useful for distinguishing vegetation types in satellite imagery in an automated manner. In this study, we collected field data for supervised learning of very high-resolution (0.3 m) WorldView-3 satellite imagery and tuned the hyperparameters of an extreme gradient boosting classifier to produce a viable solution for detecting the probability of gamba grass presence. To evaluate the performance of WorldView-3 imagery in discriminating gamba grass, we tested the utility of predictors derived from: 1) spectral bands; 2) textural features; 3) spectral indices; and 4) all predictors combined. Our results suggest that gamba grass presence can be mapped from space with an accuracy of up to 91% under optimal environmental conditions.**

***Index Terms*—Extreme gradient boosting (XGBoost), gamba grass, high resolution, machine learning, remote sensing, weed, WorldView-3.**

I. Introduction

Andropogon gayanus (gamba grass) is an introduced perennial grass that has invaded northern Australia's tropical savannas, including large areas in Queensland, the Northern Territory, and small infestations in Western Australia [1]. Gamba grass produces more biomass each year than the native savanna grasses, resulting in increased fuel load and fire intensity, and consequently, reduced biodiversity, tree cover, and carbon stores [2], [3]. Fire frequency is high across the region, with an average of one fire every two years [4]. Gamba grass is fire resistant [5], regenerating quickly from the established tussocks [6], [7] and resulting in a substantially increased fire risk [3], [8]. Due to the detrimental environmental and socioeconomic impacts following invasion, gamba grass has been recognized as one of Australia's 32 weeds of national significance (WoNS), and a key threatening process under Australian federal legislation [1]. Critical to the effective management of this weed is knowledge on its distribution and spread. However, to date, comprehensive mapping of the gamba grass has been limited by the vast spatial extent of the current and potential invasion, and limited on-ground access to much of the region.

In 2015, gamba grass cover was successfully mapped over a 30-km^2 area in Australia's Northern Territory using a point cloud obtained from airborne light detection and ranging (LiDAR) [9]. The authors defined a method for extracting a two-dimensional mask from the point cloud with a voxel-based approach that utilizes height and connectivity information. Airborne LiDAR is a powerful landscape-scale mapping tool but is relatively expensive compared to alternatives such as satellite imagery and is less scalable across large regions, particularly because high point densities dictate a low-speed flight plan.

Airborne and satellite-derived multispectral and hyperspectral imagery offer alternative avenues for mapping of invasive weeds [10]. Airborne hyperspectral imagery has been used to map multiple invasive weeds with accuracies of 57%–97% [11]–[13]. However, as the acquisition of airborne imagery is usually prohibitive for large-scale studies, in recent years, there has been an increase in research focus on the use of very high-resolution (VHR) multispectral satellite imagery [14]–[18]. The application of VHR satellite imagery [ground sample distance (GSD) of $\leq$0.5 m] has confirmed its ability to map weed species with accuracies of 68–93%. The large spread in reported accuracies could be attributed to the presence of multiple classes of weeds and nonweeds in the classification task.

The comparable accuracies in mapping invasive weeds using airborne hyperspectral and VHR multispectral satellite imagery suggest that the latter is a promising tool for mapping gamba grass and may be comparatively cost-effective for systematic mapping of large areas. Indeed, the majority of airborne-based studies have considered a mapping area of less than 30 km^2, and indicated that high accuracies of weed detection has much to do with the higher spatial and spectral resolution of imagery [11]. While both airborne hyperspectral and VHR satellite multispectral sensors can generate imagery with similarly high spatial resolutions (i.e., GSD of $\leq$0.5 m), the latter are limited by spectral resolution with current sensors usually providing

Manuscript received February 28, 2020; revised May 20, 2020 and June 21, 2020; accepted July 16, 2020. Date of publication August 3, 2020; date of current version August 17, 2020. This work was supported by the Australian Government's National Environmental Science Program (NESP) through the Northern Australia Environmental Resources Hub. *(Corresponding author: Yuri Shendryk.)*

Yuri Shendryk is with Agriculture and Food, Commonwealth Scientific and Industrial Research Organisation, Brisbane, ACT 4067, Australia (e-mail: yuri.shendryk@csiro.au).

Natalie A. Rossiter-Rachor is with the Research Institute for the Environment and Livelihoods, Charles Darwin University, Darwin 0909, NT, Australia (e-mail: natalie.rossiter@cdu.edu.au).

Samantha A. Setterfield is with the School of Biological Sciences, The University of Western Australia, Perth, WA 6009, Australia (e-mail: samantha.setterfield@uwa.edu.au).

Shaun R. Levick is with Land and Water, Commonwealth Scientific and Industrial Research Organisation, Darwin 0828, NT, Australia (e-mail: shaun.levick@csiro.au).

Digital Object Identifier 10.1109/JSTARS.2020.3013663

4-band imagery in red, green, blue (RGB) and near-infrared (NIR) wavelengths. At present, the most advanced, publicly available VHR satellite sensor is WorldView-3 having 17 bands [including panchromatic (PAN) band] in visible/NIR (VNIR) and short-wave infrared (SWIR) wavelengths (450–2365 nm) [19]. The benefit of these additional spectral bands of VHR satellite multispectral sensors for mapping invasive weeds requires investigation. It was previously reported that imagery with additional spectral bands beyond RGB+NIR wavelengths did not improve the discrimination of invasive weeds [15], [17]. While textural features extracted from VHR satellite imagery have been previously reported to capture the components of vegetation structure [20], no research has investigated the benefits of textural features for classifying invasive weeds.

The aim of this study was to test the suitability of VHR WorldView-3 imagery for mapping the presence of gamba grass over large landscapes (>200 km^2). To evaluate the sensitivity of WorldView-3 imagery in discriminating gamba grass, we tested the utility of the following.

1) Spectral bands.
2) Textural features.
3) Normalized difference spectral indices (NDSIs).
4) All predictors in combination.

We also investigated whether the additional 12 spectral bands that WorldView-3 offers beyond standard RGB+NIR wavelengths improves the discrimination of invasive weeds.

In this study, an extreme gradient boosting (XGBoost) classification algorithm [21] was employed to take advantage of the data dimensionality. To date, the most popular classification algorithms for mapping invasive weeds were random forest [12], [16], maximum likelihood [11], [15], and spectral angle mapper [11], [13]. However, in this study, we explored a decision tree boosting algorithm, which has showed good potential in multiple classification benchmarks [22], [23]. The main advantages of XGBoost are as follows:

1) high computational speed due to parallel processing of data;
2) generally better performance in comparison to other decision tree-based models (if hyperparameters are tuned properly);
3) nonreliance on missing value imputation, scaling, and normalization of the input data; and
4) in-built regularization terms that can be used to control the complexity of the model and avoid overfitting [24].

II. Methodology

A. Study Area

The 205-km^2 area of interest (AOI) was located near the township of Batchelor, which is approximately 100 km south of Darwin, in the Northern Territory of Australia (see Fig. 1). The study area is largely under private ownership for pastoral lease or semirural development, with other significant areas owned by local communities or under government ownership [9]. This area experiences a tropical climate with distinct wet and dry seasons. The annual rainfall is 1535 mm with the heaviest falls occurring during November to April [25]. The major vegetation type

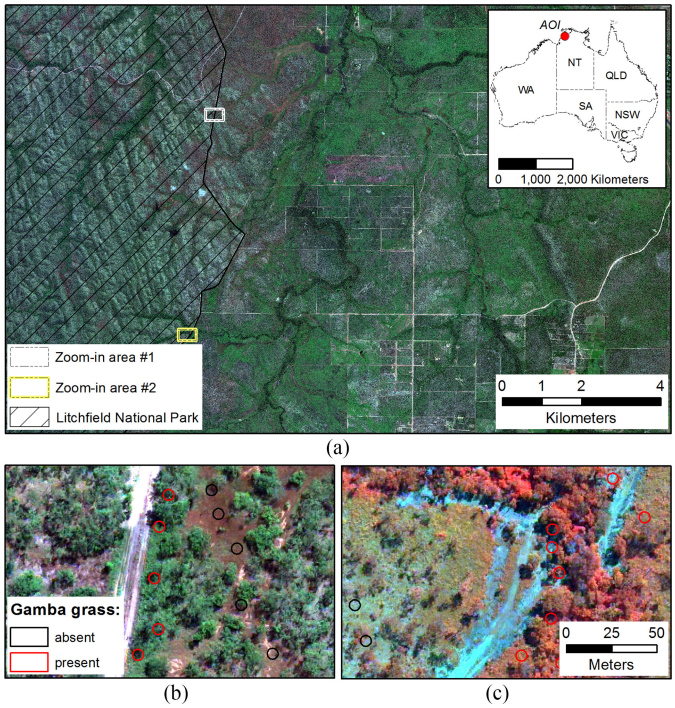

Fig. 1. WorldView-3 image of (a) AOI in relation to the extent of Australia with a (b) zoom-in area #1 in true-color (5, 3, 2) and (c) zoom-in area #2 in false-color (8, 4, 1) (see Table II for spectral band notations).

Fig. 2. Examples of live gamba grass (*Andropogon gayanus*) as seen from the (a) ground level and (b) air in April 2019.

is savanna woodland dominated by *Eucalyptus miniata* Cunn. Ex Schauer (Darwin woollybutt) and *Eucalyptus tetrodonta* F. Muell (Darwin stringybark), with a grass understory dominated by native annual grass species, such as *Sorghum* spp., and perennial species such as *Heteropogon contortus* (L.) Roem. & Schult and *Alloteropsis semialata* (R. Br.) Hitchc., and invasive perennial species such as *Andropogon gayanus* Kunth. (see Fig. 2).

Gamba grass could be structurally (i.e., in terms of height and density) similar to *Sorghum* spp., which is the most dominant annual grass species in the study area. However, gamba grass has a different phenological cycle compared to native savanna grasses [3], [7], and remains tall (up to 4 m) and photosynthetically active into the mid dry season (June–July), by which time

TABLE I
WORLDVIEW-3 ACQUISITION PARAMETERS

Scene name	Sun θ (°)	Sun α (°)	Satellite θ (°)	Satellite α (°)	Off-nadir angle (°)	Time (GMT)
P001	44.0	61.1	356.4	79.5	9.5	01:57:53
P002	43.9	61.1	197.5	71.7	16.7	01:58:35

θ = elevation angle; α = azimuth angle; GMT = Greenwich Mean Time.

the native grasses have senesced (usually March–May) and have "collapsed" to form a low (∼0.5–1 m) grass layer. Consequently, the early to middle dry season is the best seasonal time to accurately detect gamba grass, when it is distinct and clearly visible in the landscape.

B. Field Measurements

The field measurements were collected in multiple surveys conducted on 21–22 March, 16–17 April, 5–6 July, 6–7 August, and 17–18 September 2019. A Leica GS16 GNSS smart antenna and CS20 controller [26] were used in conjunction with Leica's precise point positioning (PPP) service to enable centimeter accuracy of field data. In the first survey, circular plots [radius (r) = 3 m and area (A) ≈ 28.3 m^2] were mapped with a homogeneous representation of (1) live (green) gamba grass, (2) burnt gamba grass, (3) herbicide-sprayed gamba grass, (4) senesced *Sorghum* spp., (5) burnt *Sorghum* spp., and (6) other live and senescing native grass species. In the second survey, circular plots were only classed as gamba or non-gamba due to phenological changes that occur as the dry season progresses. Additional classes were defined after the fieldwork by delineating homogeneous circular plots (r = 3 m) of (7) trees and (8) water bodies in WorldView-3 imagery. For the purpose of this study, gamba grass classes (i.e., 1, 2, and 3) and non-gamba classes (i.e., 4, 5, 6, 7, and 8) were grouped into two classes resulting in an imbalanced (≈1:2) training dataset of 187 and 355 samples, respectively.

C. WorldView-3 Imagery

The WorldView-3 imagery was tasked to be acquired on the 11th of April 2019 and two scenes were collected within 42 s of each other (see Table I).

Each scene consisted of PAN, VNIR, and SWIR bands in a wavelength range 400–2365 nm (see Table II). April is the transition between wet and dry season and was selected to task the satellite imagery acquisition because *Sorghum* spp. and other native grasses have commenced senescence and browning in color, while gamba grass remains photosynthetically active and green, and as such, is recognizable from the air (see Fig. 2) [2].

The satellite imagery was provided as an Ortho Ready 2A product, which was radiometrically corrected to ground reflectance, and projected to a plane using UTM 52S projection. Radiometric correction was accomplished using atmospheric compensation (AComp) algorithm developed by DigitalGlobe for their WorldView sensor series imagery [27].

TABLE II
WORLDVIEW-3 SPECTRAL BANDS

#	Name	Wavelength (nm)	Resolution (m)
1	Coastal (COA)	400 – 450	1.2
2	Blue (BLU)	450 – 510	1.2
3	Green (GRE)	510 – 580	1.2
4	Yellow (YEL)	585 – 625	1.2
5	Red (RED)	630 – 690	1.2
6	Red Edge (REE)	705 – 745	1.2
7	NIR-1 (NI1)	770 – 895	1.2
8	NIR-2 (NI2)	860 – 1040	1.2
9	SWIR-1 (SW1)	1195 – 1225	7.5
10	SWIR-2 (SW2)	1550 – 1590	7.5
11	SWIR-3 (SW3)	1640 – 1680	7.5
12	SWIR-4 (SW4)	1710 – 1750	7.5
13	SWIR-5 (SW5)	2145 – 2185	7.5
14	SWIR-6 (SW6)	2185 – 2225	7.5
15	SWIR-7 (SW7)	2235 – 2285	7.5
16	SWIR-8 (SW8)	2295 – 2365	7.5
17	Panchromatic (PAN)	450 – 800	0.3

NIR = near-infrared, SWIR = short-wave infrared.

As a first preprocessing step, both VNIR and SWIR bands were pan-sharpened to 0.3 m resolution based on the PAN band and using Zhang [28] algorithm implemented in PCI Geomatica software [29]. Then, pan-sharpened VNIR and SWIR bands as well as the PAN band were orthorectified using rational polynomial coefficients (RPC) and 30 m resolution SRTM-derived digital elevation model (DEM) using PCI Geomatica OrthoEngine. Finally, PAN, VNIR, and SWIR spectral bands were stacked and merged into a single mosaic with 0.3 m spatial resolution.

From the PAN band, 12 textural features were derived using a filter window of 3 × 3 pixels [30]–[32] (see Table III). Previous research suggests that using a filter window of 3 × 3 pixels results in the highest accuracy of vegetation parameter estimation in VHR satellite imagery [32]. Similarly, from VNIR and SWIR bands, 120 NDSIs were derived (i.e., $n!/r! \times (n-r)!$, where n represents the total number of bands and r represents the number of bands used to calculate one NDSI). NDSIs were calculated in succession from Coastal Blue to SWIR-8 (see Table II) spectral bands as follows:

$$\mathrm{NDSI}_{(i,j)} = (R_i - R_j) / (R_i + R_j)$$

where R is the spectral reflectance, and i and j are numbers indicating the wavelengths (nm). Throughout this article, each NDSI is denoted as combination of three-letter acronyms in Table II (e.g., $\mathrm{NDSI}_{(\mathrm{COA,SW8})}$ is denoted as COASW8).

Finally, we calculated predictor variables from WorldView-3-derived spectral bands, textural features, and NDSIs. For this we extracted statistics (see Table IV) in 17 × 17 pixels windows (i.e., ≈26 m^2) to match the areas of field measurements (i.e., ≈28.3 m^2) for each field sampling plot, resulting in total of 1036 predictor variables.

The examples of WorldView-3 imagery derived features and predictor variables are presented in Fig. 3.

TABLE III
TEXTURAL FEATURES EXTRACTED FROM THE PANCHROMATIC BAND

Name	Equation
Homogeneity (HOM)	$\sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1+(i-j)^2}$
Contrast (CON)	$\sum_{i,j=0}^{N-1} P(i,j) \times (i-j)^2$
Dissimilarity (DIS)	$\sum_{i,j=0}^{N-1} P(i,j) \times \lvert i-j \rvert$
Mean (MEA)	$Mean_i = \sum_{i,j=0}^{N-1} i \times (P(i,j))$
Standard Deviation (STD)	$Var_i = \sum_{i,j=0}^{N-1} (P(i,j) \times (i - Mean_i)^2$ $Std_i = \sqrt{Var_i}$
Entropy (ENT)	$\sum_{i,j=0}^{N-1} -P(i,j) \times \log_e(P(i,j))$, assuming that $0 \times \log_e 0 = 0$
Angular Second Moment (ASM)	$\sum_{i,j=0}^{N-1} P(i,j)^2$
Correlation (COR)	$\sum_{i,j=0}^{N-1} \frac{P(i,j) \times (i - Mean_i) \times (j - Mean_j)}{\sqrt{Var_i \times Var_j}}$
GLDV Angular Second Moment (GAM)	$\sum_{k=0}^{N-1} V(k)^2$
GLDV Entropy (GEN)	$\sum_{k=0}^{N-1} -(V(k) \times \log_e(V(k))$, assuming that $0 \times \log_e 0 = 0$
Inverse Difference (IND)	$\sum_{i,j=0}^{N-1} \frac{P(i,j)}{\lvert i-j \rvert^2}$, for $i \,!= j$
Edge (EDG)	$CT = \sum_{i=1}^{n} \frac{\lvert CT - a_i \rvert}{n}$, where a_i are gray levels of each pixel in the filter window, CT is a gray-level value of the central pixel (e.g. n=8 for a filter window of 3×3)

GLCM is a gray-level co-occurrence matrix; GLDV is a gray-level difference vector derived from a GLCM; N is the number of gray levels; P is the normalized symmetric GLCM of dimension $N \times N$; V is the normalized gray-level difference vector of dimension N; $P(i,j)$ is the normalized co-occurrence matrix such that $\sum_{i,j=0}^{N-1} P(i,j) = 1$; $V(k)$ is the normalized gray-level difference vector, $V(k) = \sum_{i,j=0}^{N-1} P(i,j)$, where $\lvert i-j \rvert = k$ [30], [33].

TABLE IV
PREDICTOR VARIABLES EXTRACTED FROM WORLDVIEW-3-DERIVED SPECTRAL BANDS, TEXTURAL FEATURES, AND NDSIS

Name	Description
max	maximum value of all pixels within 17×17 pixels window
min	minimum value of all pixels within 17×17 pixels window
avg	average value of all pixels within 17×17 pixels window
std	standard deviation of all pixels within 17×17 pixels window
p25	25th percentile of all pixels within 17×17 pixels window
p50	50th percentile of all pixels within 17×17 pixels window
p75	75th percentile of all pixels within 17×17 pixels window

Note: Predictor variables are denoted as a combination of the name of the statistics and feature (e.g., p50_COASW8 stands for 50th percentile of all pixel values in 17 × 17 pixels window of $\text{NDSI}_{(\text{COA,SW8})}$ feature).

D. Machine Learning

In this study, we used XGBoost algorithm [21], [34] to differentiate gamba grass from other vegetated and nonvegetated areas. XGBoost is an ensemble learning method that combines the predictive power of multiple linear models or decision trees using a boosting algorithm. In boosting, a decision tree or linear regression that improves the model most is added to an ensemble at each iteration until the set number of estimators (i.e., *n_estimators*) has been achieved. In contrast to bagging techniques such as random forest, in which trees are grown to their maximum extent, boosting makes use of shallow trees with fewer splits.

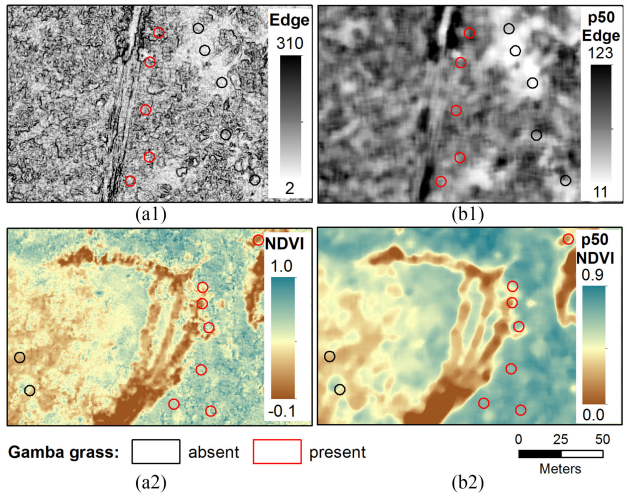

Fig. 3. Examples of (a) features within zoom-in area #1 (see Fig. 1) and (b) predictor variables within zoom-in area #2 (see Fig. 1) extracted from WorldView-3 imagery: (a1) *Edge* (see Table III), (a2) (*p50_Edge*, Table IV), (b1) normalized difference vegetation index (NDVI) (here, NDVI is an inverse of $\text{NDSI}_{(\text{RED, NI2})}$), and (b2) *p50_NDVI* (i.e., 50th percentile of all pixel values in 17 × 17 pixels window of $\text{NDSI}_{(\text{RED,NI2})}$, Table IV).

The training data were shuffled and split into *"train"* (75% of the data) and *"test"* (25% of the data) sets. Then, XGBoost classifier with a binary logistic loss function and a tree booster was used to predict the presence of gamba grass in two stages. First, randomized search on hyperparameters was performed using the *"train"* dataset with a stratified five–fold cross validation (*cv*) (training:testing ratio of 75:25). Six hyperparameters were optimized including *max_depth* in range from 1 to 11 with an increment of 1, *learning_rate* of 0.001, 0.01, 0.1, 0.5, 1, and 2, *subsample* in range from 0.2 to 1.0, with an increment of 0.1, *min_child_weight* in range from 1 to 21 with an increment of 1, *gamma* of 0, 0.25, 0.5, and 1, and *n_estimators* of 100, 500, and 1000. We used expert knowledge to specify reasonable ranges and increments of the hyperparameters. The best model according to a *cv* score was used to extract the importance of predictor variables in terms of *gain*, which is a relative contribution of the corresponding predictor to the model [35].

Second, using predictors ranked according to their importance, another XGBoost classification using step-forward predictor selection (SFPS) was performed [36]. In each iteration, the predictor which previously best improved the model performance was added until addition of new predictors did not improve the performance in terms of a *cv* score.

Eight XGBoost models were trained and optimized using predictor variables extracted from: (1) four spectral bands (2, 3, 5, and 7 in Table II), (2) eight spectral bands (1 to 8 in Table II), (3) 16 spectral bands (1 to 16 in Table II), (4) textural features derived from a PAN band, (5) six NDSIs derived from four spectral bands (2, 3, 5, and 7 in Table II), (6) 28 NDSIs derived from eight spectral bands (1 to 8 in Table II), (7) 120 NDSIs derived from 16 spectral bands (1 to 16 in Table II), and (8) all predictor variables in combination.

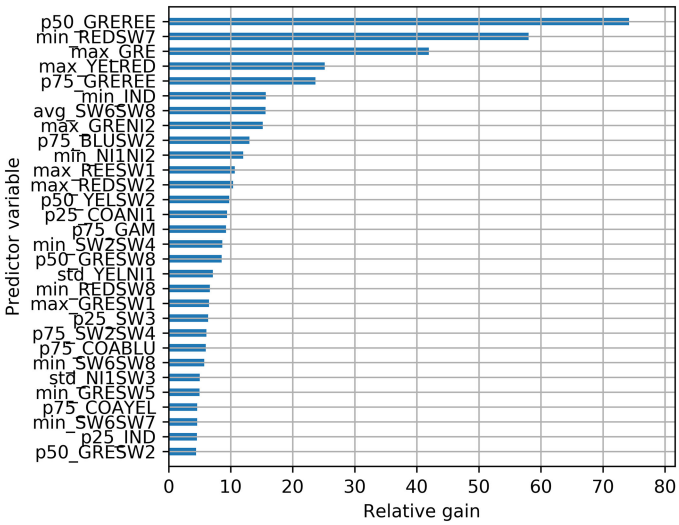

Fig. 4. Ranked (from highest to lowest) importance of 30 predictor variables for the model trained using all predictor variables in combination.

Multiple recent studies of VHR satellite imagery utilized object-based image analysis (OBIA) prior to classifying invasive weeds [10], [15], [18]. While OBIA has previously shown to improve accuracies of classification as compared to per-pixel classification [37], it usually relies on segmentation algorithms that are difficult to validate and correct, if applied poorly. In this study, to utilize the power of OBIA approaches without relying on segmentation algorithms, we trained XGBoost models using predictor variables calculated within field-measured areas ($\approx$28.3 m^2) and applied it to predictor variables calculated within 17×17 pixels neighborhood area ($\approx$26 m^2).

To optimize and evaluate our models during cross validation and test stages, we used balanced accuracy (BA) metric, which avoids inflated performance estimates on imbalanced datasets. It is the macroaverage of recall scores per class or, equivalently, raw accuracy where each sample is weighted according to the inverse prevalence of its true class, and is defined as

$$\mathrm{BA} = \left(\frac{\mathrm{TP}/P + \mathrm{TN}/N}{2} \right)$$

where TP is a true positive (i.e., correctly classified as positive), TN is a true negative (i.e., correctly classified as negative), P is a positive, and N is a negative.

III. Results

A. Importance of Predictor Variables

The top 30 most important predictors for the model trained using all predictor variables in combination is shown in Fig. 4. Here, importance is expressed in terms of a relative *gain* (i.e., the relative contribution of the corresponding predictor to the model).

Predictor importance-identified spectral bands offering greatest capacity for discrimination of gamba grass were those covering 510–745 nm and 2185–2365 nm wavelength ranges (see Table II).

TABLE V
Balanced Accuracy of Gamba Grass Presence Classification Using All Predictors and Predictors Selected Using SFPS

#	Model (b)	n	BA (%), all predictors		n	BA (%), SFPS	
			cv	*test*		*cv*	*test*
1	Bands (4)	28	80.2	77.2	12	82.4	77.9
2	Bands (8)	56	81.3	81.0	9	83.8	84.2
3	Bands (16)	112	82.5	81.1	20	85.0	82.7
4	Texture (1)	84	77.9	79.5	16	80.1	76.6
5	NDSIs (4)	42	79.4	78.9	37	80.5	79.4
6	NDSIs (8)	196	85.1	83.2	7	86.9	85.6
7	NDSIs (16)	840	85.7	88.6	80	90.6	91.3
8	Combined (17)	1036	87.3	86.9	78	91.2	91.2

BA is balanced accuracy; n = number of predictor variables; *cv* = fivefold cross validation; b = number of spectral bands used for extracting predictor variables.

B. Step-Forward Predictor Selection

The SFPS procedure generally resulted in an improved performance of XGBoost classifier with a BA increase on a *"test"* set of up to 4.3% (Model #8 in Table V), which is in line with previous findings by Robinson *et al.* [15]. The SFPS also reduced the number of necessary predictors for best performance by 57% (Model #1)–96% (Model #6). Although the combined use of all predictors (Model #8) in SFPS procedure led to a BA increase from 86.9% to 91.2%, it was still no better than that of NDSIs-derived model (Model #7).

In contrast to previous studies [15], [17], additional spectral bands did improve classification accuracy by 4.8% when going from a 4-band (Model #1) to a 16-band setup (Model #3). The improved performance became even more pronounced in an NDSIs scenario with a BA improvement between Model #5 and #7 of 11.9% (see Table V). While the model trained using textural features only (i.e., Model #4) showed the worst performance with BA of 76.6%, it was still comparable to that of Model #1 utilizing RGB+NIR bands (BA of 77.9%).

For Model #8, only three predictors (i.e., *p50_GREREE*, *min_REDSW7*, and *max_GRE*) were necessary to generate gamba grass presence classification with BA = 85% (see Figs. 4 and 5).

According to Fig. 5, cross-validated BA stopped improving after using the top 78 predictors. Nonetheless, the BA improvement between models trained using $\approx$10 and 83 predictor variables could be considered marginal.

The final map of gamba grass presence at 0.3 m spatial resolution generated using a model with hyperparameters from Model #8 (see Table V), trained using all training data (i.e., *"train"* and *"test"* sets combined) and top 78 predictor variables, is presented in Fig. 6.

Fig. 6(a) shows that only the western part (i.e., hilly to rugged ridges of the Litchfield National Park) of the study area was mostly free of gamba grass. Interestingly, shaded areas were occasionally misclassified for gamba presence [see Fig. 6(c)].

IV. Discussion

Gamba grass is a high impact weed invading northern Australia. Developing an accurate and cost-effective method to map

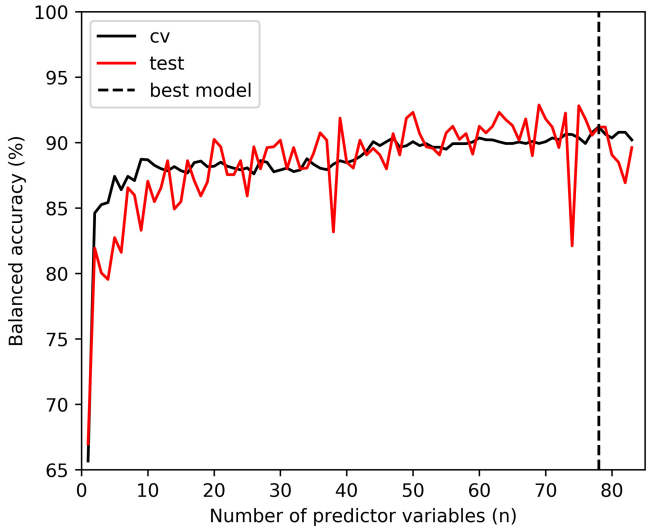

Fig. 5. SFPS for Model #8 using all predictor variables in combination. (Note: Only 83 predictor variables were used in SFPS, as 953 predictors with zero importance were excluded.)

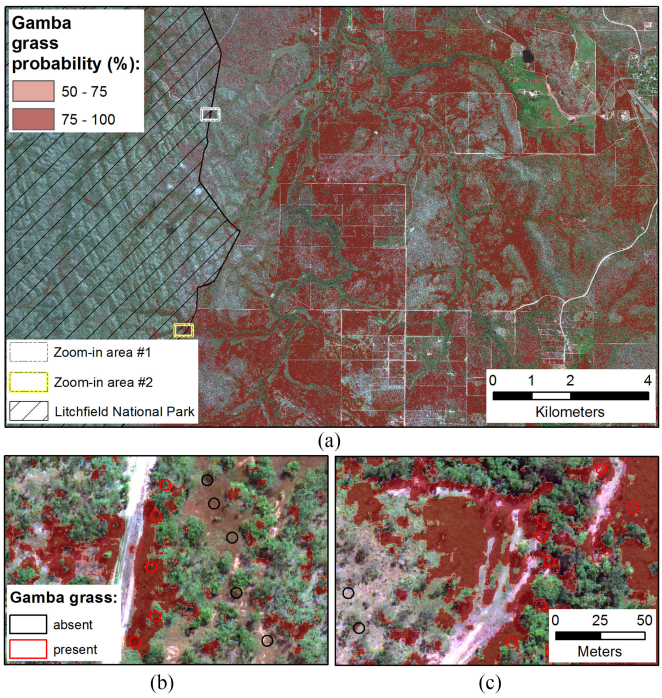

Fig. 6. (a) Final map of gamba grass presence probability (11th of April 2019) with (b) zoom-in area #1 and (c) zoom-in area #2.

gamba grass is critical for better management of this weed. Our results demonstrate that VHR WorldView-3 imagery can be used to differentiate gamba grass from other vegetated and nonvegetated areas, with accuracies of up to 91.3%. This methodology is scalable to larger areas as it relies exclusively on readily accessible VHR satellite imagery, not airborne LiDAR data [9]. This study is a significant advancement for stakeholders, as the accuracy of gamba grass mapping is sufficient to inform landscape management in northern Australia. Using WorldView-3 imagery was more affordable than airborne LiDAR or airborne hyperspectral imagery. The data acquisition costs (USD) of WorldView-3 imagery were \$0.24/ha and \$0.48/ha for 8-band and 16-band imagery, respectively, compared to ∼\$1/ha and ∼\$1.1/ha for airborne hyperspectral and LiDAR, respectively (cost based on a >10 000 ha survey) [38], [39].

WorldView-3-derived NDSIs provided more separability when classifying gamba grass compared to individual spectral bands or textural features. Additional WorldView-3 VNIR bands (i.e., Coastal, Yellow, Red Edge, and NIR-2) provided 6.2% increase in classification accuracy, while the addition of SWIR bands improved classification accuracy by another 5.7%. However, given the spectral and structural differences of different vegetation types, this result might be not applicable when mapping other weed species [15], [17]. Interestingly, textural features extracted from the PAN band provided a satisfactory classification result with BA of 76.6%, which is relevant when considering the use of single-band sensors (e.g., WorldView-1) for mapping invasive weeds.

An occasional misclassification of gamba grass presence occurred in areas shaded by tree crowns. These were characterized by change in the spectral shape, and were occasionally similar in terms of spectral and NDSI signal to unshaded gamba grass. This problem could be alleviated by collecting additional training samples in shaded areas to aid further discrimination.

VHR satellite imagery is reportedly more accurate than medium-resolution satellite imagery in mapping invasive weeds [40]. Nevertheless, there are numerous studies that evaluated medium-resolution satellite imagery for invasive weed mapping and achieved accuracies of up to 90% [40]–[42]. While being able to cover large areas and usually available for free, medium-resolution satellite imagery requires extensive field surveys at multiple time steps to achieve the above-mentioned accuracies. Therefore, we propose using the generated gamba grass map from the VHR WorldView-3 imagery to inform medium-resolution satellite imagery (e.g., Sentinel-2 [43]) in order to upscale the extent of gamba grass maps in time and space. Previous research suggests that such upscaling could be achieved through the use of convolutional neural networks [44].

V. Conclusion

The ability to map the spread of invasive weed species is critical for effective environmental management. Our results showed that XGBoost has the ability to map invasive weed presence in WorldView-3 imagery with high accuracy (BA = 91.3%). The use of WorldView-3-derived NDSIs significantly improved the detection of gamba grass as compared to raw spectral bands. Since gamba grass is rapidly invading landscapes of the northern Australia, we recommend that future studies should focus on multitemporal mapping of this weed using freely available medium-resolution, multispectral satellite imagery that is trained with the type of results shown in this study.

Acknowledgment

The authors would like to thank the Northern Territory Government, Parks, Wildlife, and Heritage Division for supporting our fieldwork in Litchfield National Park. The

authors would also like to thank J. Schatz and P. Grandclement for assistance in the field.

REFERENCES

[1] S. A. Setterfield, N. A. Rossiter-Rachor, and V. M. Adams, "Navigating the fiery debate: The role of scientific evidence in eliciting policy and management responses for contentious plants in northern Australia," *Pacific Conservation Biol.*, vol. 24, no. 3, pp. 318–328, 2018.

[2] A. M. Petty, S. A. Setterfield, K. B. Ferdinands, and P. Barrow, "Inferring habitat suitability and spread patterns from large-scale distributions of an exotic invasive pasture grass in north Australia," *J. Appl. Ecology*, vol. 49, no. 3, pp. 742–752, 2012.

[3] S. A. Setterfield, N. A. Rossiter-Rachor, L. B. Hutley, M. M. Douglas, and R. J. Williams, "Turning up the heat: The impacts of *Andropogon gayanus* (gamba grass) invasion on fire behaviour in northern Australian savannas," *Diversity Distrib.*, vol. 16, no. 5, pp. 854–861, 2010.

[4] J. Beringer *et al.*, "Fire in Australian savannas: From leaf to landscape," *Global Change Biol.*, vol. 21, no. 1, pp. 62–81, Jan. 2015.

[5] B. Bowden, "Studies on *Andropogon gayanus* Kunth: III. An outline of its biology," *J. Ecology*, vol. 52, pp. 255–271, 1964.

[6] N. A. Rossiter, S. A. Setterfield, M. M. Douglas, and L. B. Hutley, "Testing the grass-fire cycle: Alien grass invasion in the tropical savannas of northern Australia," *Diversity Distrib.*, vol. 9, no. 3, pp. 169–176, 2003.

[7] N. A. Rossiter-Rachor, S. A. Setterfield, M. M. Douglas, L. B. Hutley, G. D. Cook, and S. Schmidt, "Invasive *Andropogon gayanus* (gamba grass) is an ecosystem transformer of nitrogen relations in Australian savanna," *Ecological Appl.*, vol. 19, no. 6, pp. 1546–1560, Sep. 2009.

[8] S. A. Setterfield *et al.*, "Adding fuel to the fire: The impacts of non-native grass invasion on fire management at a regional scale," *Plos One*, vol. 8, no. 5, May 14 2013, Art. no. e59144.

[9] S. R. Levick, S. A. Setterfield, N. A. Rossiter-Rachor, L. B. Hutley, D. McMaster, and J. M. Hacker, "Monitoring the distribution and dynamics of an invasive grass in tropical savanna using airborne LiDAR," *Remote Sens.*, vol. 7, no. 5, pp. 5117–5132, May 2015.

[10] B. A. Bradley, "Remote detection of invasive plants: A review of spectral, textural and phenological approaches," *Biol. Invasions*, vol. 16, no. 7, pp. 1411–1425, Jul. 2014.

[11] C. H. Yang and J. H. Everitt, "Mapping three invasive weeds using airborne hyperspectral imagery," *Ecological Inform.*, vol. 5, no. 5, pp. 429–439, Sep. 2010.

[12] R. L. Lawrence, S. D. Wood, and R. L. Sheley, "Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (RandomForest)," *Remote Sens. Environ.*, vol. 100, no. 3, pp. 356–362, Feb. 2006.

[13] L. W. Lass, T. S. Prather, N. F. Glenn, K. T. Weber, J. T. Mundt, and J. Pettingill, "A review of remote sensing of invasive weeds and example of the early detection of spotted knapweed *(Centaurea maculosa)* and babysbreath *(Gypsophila paniculata)* with a hyperspectral sensor," *Weed Sci.*, vol. 53, no. 2, pp. 242–251, Mar./Apr. 2005.

[14] J. H. Everitt, C. Yang, and C. J. Deloach, "Remote sensing of giant reed with QuickBird satellite imagery," *J. Aquatic Plant Manage.*, vol. 43, pp. 81–85, Jul. 2005.

[15] T. P. Robinson, G. W. Wardell-Johnson, G. Pracilio, C. Brown, R. Corner, and R. D. van Klinken, "Testing the discrimination and detection limits of WorldView-2 imagery on a challenging invasive plant target," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 44, pp. 23–30, Feb. 2016.

[16] K. Peerbhay, O. Mutanga, R. Lottering, and R. Ismail, "Mapping Solanum mauritianum plant invasions using WorldView-2 imagery and unsupervised random forests," *Remote Sens. Environ.*, vol. 182, pp. 39–48, Sep. 2016.

[17] V. Marshall, M. Lewis, and B. Ostendorf, "Do additional bands (coastal, NIR-2, red-edge and yellow) in Worldview-2 multispectral imagery improve discrimination of an invasive tussock, buffel grass *(Cenchrus ciliaris)*?," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. 39-B8, pp. 277–281, 2012.

[18] F. Alvarez-Taboada, C. Paredes, and J. Julian-Pelaz, "Mapping of the invasive species *Hakea sericea* using unmanned aerial vehicle (UAV) and WorldView-2 imagery and an object-oriented approach," *Remote Sens.*, vol. 9, no. 9, Sep. 2017, Art. no. 913.

[19] Maxar Technologies. *WorldView-3 Satellite Sensor*. (2020). [Online]. Available: https://www.maxar.com/

[20] E. M. Wood, A. M. Pidgeon, V. C. Radeloff, and N. S. Keuler, "Image texture as a remotely sensed measure of vegetation structure," *Remote Sens. Environ.*, vol. 121, pp. 516–526, Jun. 2012.

[21] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016, pp. 785–794.

[22] R. S. Olson, W. La Cava, Z. Mustahsan, A. Varik, and J. H. Moore, "Data-driven advice for applying machine learning to bioinformatics problems," *Pac. Symp. Biocomput.*, vol. 23, pp. 192–203, 2018, *arXiv:1708.05070*.

[23] G. J. Briem, J. A. Benediktsson, and J. R. Sveinsson, "Multiple classifiers applied to multisource remote sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 10, pp. 2291–2299, Oct. 2002.

[24] E. Pricop, J. Fattahi, N. Dutta, and M. Ibrahim, *Recent Developments on Industrial Control Systems Resilience*. Cham, Switzerland: Springer, 2020.

[25] Australian Bureau of Meterology. *Climate Statistics for Australian Locations*. (2020). [Online]. Available: http://www.bom.gov.au/climate/

[26] Leica Geosystems AG. *Leica Viva GS16—Self-Learning GNSS Smart Antenna*. (2020). [Online]. Available: https://leica-geosystems.com/products/gnss-systems/smart-antennas/leica-viva-gs16

[27] F. Pacifici, "An automatic atmospheric compensation algorithm for very high spatial resolution imagery and its comparison to FLAASH and QUAC," in *Proc. Joint Agency Commercial Imagery Eval. Workshop,* Saint Louis, MO, USA, 2013, pp. 16–18.

[28] Y. Zhang, "Problems in the fusion of commercial high-resolution satellite as well as Landsat 7 images and initial solutions," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. 34, no. 4, pp. 587–592, 2002.

[29] PCI Geomatica. *A Remote Sensing Desktop Software Package for Processing Earth Observation Data*. (2020). [Online]. Available: www.pcigeomatics.com

[30] PCI Geomatica. *Texture Analysis*. (2019). [Online]. Available: https://www.pcigeomatics.com/geomatica-help/references/pciFunction_r/python/P_tex.html

[31] A. Baraldi and F. Parmiggiani, "An investigation of the textural characteristics associated with gray-level cooccurrence matrix statistical parameters," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 2, pp. 293–304, Mar. 1995.

[32] J. J. Zhou *et al.*, "The effects of GLCM parameters on LAI estimation using texture values from Quickbird satellite imagery," *Sci. Rep.*, vol. 7, no. 1, pp. 1–12, Aug. 2017.

[33] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.

[34] XGBoost. *XGBoost Documentation*. (2020). [Online]. Available: https://xgboost.readthedocs.io/

[35] XGBoost. *Introduction to Boosted Trees*. (2020). [Online]. Available: https://xgboost.readthedocs.io/en/latest/tutorials/model.html

[36] A. Miller, *Subset Selection in Regression*. London, U.K.: Chapman & Hall, 2002.

[37] T. G. Whiteside, G. S. Boggs, and S. W. Maier, "Comparing object-based and pixel-based classifications for mapping savannas," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 13, no. 6, pp. 884–893, Dec. 2011.

[38] A. Held, S. Phinn, M. Soto-Berelov, and S. Jones, *AusCover Good Practice Guidelines: A Technical Handbook Supporting Calibration and Validation Activities of Remotely Sensed Data Products*, vol. 352. Canberra, ACT, Australia: TERN AusCover, 2015.

[39] R. Hunt, R. Hamilton, and J. Everitt, "Mapping weed infestations using remote sensing," USDA Forest Service Remote Sens. Appl. Center, Salt Lake City, UT, USA, 2005. [Online]. Available: https://www.fs.fed.us/eng/rsac/invasivespecies/documents/mapping.pdf

[40] T. N. Matongera, O. Mutanga, T. Dube, and M. Sibanda, "Detection and mapping the spatial distribution of bracken fern weeds using the Landsat 8 OLI new generation sensor," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 57, pp. 93–103, May 2017.

[41] P. H. Evangelista, T. J. Stohlgren, J. T. Morisette, and S. Kumar, "Mapping invasive Tamarisk (Tamarix): A comparison of single-scene and time-series analyses of remotely sensed data," *Remote Sens.*, vol. 1, no. 3, pp. 519–533, Sep. 2009.

[42] G. I. Gavier-Pizarro *et al.*, "Monitoring the invasion of an exotic tree *(Ligustrum lucidum)* from 1983 to 2006 with Landsat TM/ETM+ satellite data and support vector machines in Córdoba, Argentina," *Remote Sens. Environ.*, vol. 122, pp. 134–145, 2012.

[43] Copernicus. *Copernicus—European Union's Earth Observation Programme*. (2020). [Online]. Available: https://www.copernicus.eu/en

[44] Y. Rist, I. Shendryk, F. Diakogiannis, and S. Levick, "Weed mapping using very high resolution satellite imagery and fully convolutional neural network," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 9784–9787.

Yuri Shendryk (Member, IEEE) received the B.Sc. degree in geology from the Taras Shevchenko National University of Kyiv, Kyiv, Ukraine, in 2009, the M.Sc. degree in geography from Lund University, Lund, Sweden, in 2013, and the Ph.D. degree in geography and remote sensing from the University of New South Wales, Sydney, NSW, Australia, in 2017.

Since June 2017, he has been a Postdoctoral Fellow with the Commonwealth Scientific and Industrial Research Organisation (CSIRO), Brisbane, ACT, Australia, specializing in the application of GIS, remote sensing, and machine learning for natural resource management. His current research interests include the integration of airborne- and satellite-based remote sensing for forest monitoring and precision agriculture.

Natalie A. Rossiter-Rachor received the Ph.D. degree in ecology from Charles Darwin University, Casuarina, NT, Australia, in 2008.

She is a Senior Lecturer with Charles Darwin University. Over the past 20 years, her research has focused on improving the understanding of the impacts of weed species on savanna ecosystems across northern Australia. Her current research interests include the development of management strategies to help minimize the ecosystem impacts of invasive grass weeds, including reducing flammability of invaded ecosystems.

Samantha A. Setterfield received the B.Sc. Hons. (First Class) degree from Monash University, Melbourne, VIC, Australia, in 1988, and the Ph.D. in plant ecology from Charles Darwin University, Casuarina, NT, Australia, in 1997.

From 1996 to 2016, she was an Academic Staff Member with Charles Darwin University. Since 2016, she has been an Associate Professor of Plant Ecology with The University of Western Australia, Perth, WA, Australia. She has more than 25 years' experience in tropical ecology research, particularly the impact, detection, and management of invasive plant species and is the author of more than 70 peer-reviewed research articles.

Dr. Setterfield is an Associate Editor for the *Australian Journal of Botany*.

Shaun R. Levick received the Ph.D. degree in landscape ecology from the University of the Witwatersrand, Johannesburg, South Africa, in 2008.

He is currently a Principal Research Scientist with the Commonwealth Scientific and Industrial Research Organisation (CSIRO), Darwin, NT, Australia, Australia's national science agency. He specializes in the development and application of remote sensing technologies to address environmental challenges. His research interests include savanna landscapes, particularly, the improvement of sustainable land management and biodiversity conservation.