

A Robust Deep Learning Framework Reveals the Spread of Multiple Invasive Plants in a Biodiversity Hotspot using Satellite Imagery

Aniruddha Adiga,¹ Surbhi Singh,¹ Ethan Choo,¹ Johnny Yang,¹ Anjana Devkota,² Bharat Babu Shrestha,² Seerjana Maharjan,² Sita Gyawali,² Sandeep Dhakal,² Pramod Kumar Jha,² Krishna Prasad Poudel,³ Rangaswamy Muniappan,⁴ Srinivasan Venkatramanan,¹ Madhav Marathe,^{1,5}

Abhijin Adiga¹

¹ University of Virginia, Biocomplexity Institute and Initiative, Charlottesville, 22904, USA

² Tribhuvan University, Central Department of Botany, Kathmandu, 44618, Nepal

³ Tribhuvan University, Geography Education Department, Kathmandu, 44618, Nepal

⁴ Virginia Tech, Blacksburg, 24061, USA

⁵ University of Virginia, Department of Computer Science, Charlottesville, 22904, USA

abhijin@virginia.edu

Abstract

Invasive alien plants cause significant damage to biodiversity, agriculture, and the economy. Remote sensing coupled with deep learning is a promising approach to map their presence. However, seasonal variations in plant biology, field survey limitations, and the need to use expensive high-resolution imagery due to sparse coverage conditions severely limit the amount of training data that can be collected. We develop a patch-based framework using convolutional neural networks (CNNs) to predict the distribution of invasive plants from four-band high-resolution satellite imagery. The framework is applied to map three major invasive plants in a biodiversity hotspot. To cope with data challenges, we employ a robust training and evaluation framework using a multiple-hold-out method and Bayesian-optimization-based parameter tuning for model selection and transfer learning. Distribution maps obtained from ensemble predictions indicate that the three invasive plants are widespread in the landscape, even at higher elevation levels in the midhills of Nepal that are considered to be unsuitable for establishment. We evaluate several CNNs with various patch sizes, band subsets, initial CNN weights, and image acquisition periods.

Introduction

Biological invasions cause unprecedented disruptions to native ecosystems, negatively impacting health and economy (Pyšek and Richardson 2010). The mean cost of global invasions is estimated to be US \$26.8 billion per year (Díagne et al. 2021). The invasive species problem is a major impediment to the achievement of several sustainable development goals (UN 2019): Good Health and Well-being, Climate Action, Life Below Water, Life on Land, to name a few. Constructing accurate species distribution maps is an essential task in ecology and biogeography and help assist in monitoring, early detection, and mitigation efforts. Increased availability of high-resolution remote-sensed data has made large-scale surveillance of invasive species a viable option (He et al. 2015). Coupled with recent developments in artificial intelligence, this approach holds considerable promise in large-scale surveillance of invasive species spread (Ball, Anderson, and Chan Sr 2017).

Deep learning methods have been applied extensively in remote sensing in the last decade. Several survey papers (Ball, Anderson, and Chan Sr 2017; Zhang, Zhang, and Du 2016; Kamilaris and Prenafeta-Boldú 2018) provide a comprehensive list of state-of-the-art models in this area. In the context of environmental health and agriculture, there are many works pertaining to land cover classification and crop monitoring (Ball, Anderson, and Chan Sr 2017; Shermeyer et al. 2020; Kamilaris and Prenafeta-Boldú 2018; Jia et al. 2017; Kussul et al. 2017). However, species identification and mapping using high-resolution satellite and unmanned aerial vehicle (UAV) imagery is an emerging area (Kattenborn, Eichel, and Fassnacht 2019; Kislov and Korznikov 2020; Korznikov et al. 2021).

The focus region for this study is Nepal's Chitwan Annapurna Landscape (CHAL), which is part of the Himalayan biodiversity hotspot. Its unique native biodiversity is threatened by the combined effects of climate change and ever-increasing human activities such as trade and tourism, a significant consequence of which is increased instances of biological invasions. There have been several works on mapping invasive alien plant species (IAPS) and developing suitability maps in the Himalayan landscape (for e.g. (Shrestha 2016)). We focus on three pervasive invasive species – *Lantana camara*, *Chromolaena odorata*, and *Parthenium hysterophorus* – well-known for their severe negative impact on agriculture, natural ecosystems, and human health and are a severe threat to biodiversity in the CHAL region.

Challenges. Training deep learning models requires large amounts of data, which is seldom available in many remote-sensing applications, including IAPS mapping, due to field survey limitations and lack of high-quality satellite imagery (Ball, Anderson, and Wei 2018). Several constraints imposed by the biology of the plants, seasonality and terrain make field survey particularly challenging. The mountainous terrain of the CHAL limits access to many areas.

Much of the survey is limited to roads, residential and urban areas, leading to fewer data points and roadside bias. Unlike crops in a field, the target species coverage might be sparse and is likely co-located with other plants (Figure S1 in the supplement). Due to such sparse coverage conditions, freely available moderate-resolution images are seldom sufficient for detection. However, high-resolution images are not openly available and are too expensive to acquire for the entire CHAL region. This means that not only is the coverage of the acquired imagery small, but it is also heterogeneous; images can differ in terms of the number of bands, satellite source, illumination, time of capture, etc. The lack of spatial coverage and spatial resolution also makes it challenging to utilize publicly available databases¹. Differences in the appearance of plants across seasons lead to further restrictions. In Nepal, September to December and February to April are relatively cloud or fog-free; therefore, most of the optical sensor data is collected during this time. The color and appearance of certain plants greatly vary between seasons (*C. odorata* and *P. hysterophorus* for example). Therefore, budget, weather, and seasonality constraints imposed by multiple invasive species drastically reduce the quantity of usable imagery.

Contributions. In this collaborative work, botanists, ecologists, GIS specialists, and computer scientists from Nepal and the US address the problem of mapping multiple invasive species in the CHAL region. We develop a general framework to construct species distribution maps from high-resolution satellite imagery using deep learning.

- We present a patch-based approach for constructing species distribution maps from satellite imagery using CNN. Square patches of satellite imagery annotated by field survey data are used to train the CNN. This approach ensures that adequate area is captured in a survey location to account for sparse coverage of the plant and inaccuracy of the GPS device. It also enables us to leverage state-of-the-art architectures designed for classification of regular images in computer vision applications. Unlike pixel-based approaches (Wang et al. 2008), CNNs can effectively learn spatial patterns and textures that characterize the target species from patches of images.
- We applied several popular CNNs developed for classifying RGB images by suitably modifying them so that they could be applied to four band satellite imagery.
- The training framework addresses the problem of inadequate labeled data and image heterogeneity in several ways. We use multiple-hold-out approach for model exploration coupled with a Bayesian optimization framework for hyperparameter tuning. We also use the popular transfer learning technique of initializing the CNNs with weights obtained from pre-training them on large labeled datasets. We account for image acquisition period by including this information as a feature.
- Classifiers were trained and analyzed for different networks, various patch sizes, combinations of frequency bands, initial filter weight, survey data subsets, and image acquisition period.

¹<http://www imapinvasives org>

- Species distribution maps were created based on ensemble averaging of the selected models. Using these maps, we analyzed the extent of invasion by district and with respect to elevation. The maps indicate that the three invasive plants are widespread in the landscape, even at higher elevation levels in the midhills of Nepal that are considered to be unsuitable for establishment.

From a methodological perspective, our work shows the feasibility of applying modern machine learning techniques to construct species distribution maps. From a biological invasions perspective, our framework not only helps map the current state of IAPS establishment, but can also aid in understanding the spread pathways. By applying the models learned to satellite imagery from different time periods, the pattern of spread can be reconstructed over time. Coupled with bioclimatic variables and information on human activities, factors contributing to the spread can be identified.

Framework

We describe the survey procedure, image dataset, and the patch-based convolutional neural network framework to map species distribution using multi-band high-resolution satellite imagery.

Survey data and imagery data. The occurrence data and location of the study species' *L. camara*, *C. odorata*, and *P. hysterophorus* was collected by on-location surveyors using hand-held GPS primarily during September-December (autumn and winter) and March-May (spring) between the years 2016 and 2020. No survey was conducted during the June-August period (rainy season). The surveyors marked locations of both presence and absence took photos at some of these coordinates. High-resolution imagery from four different satellites was used in this work. This data has panchromatic imagery comprising of 8- and 4-band multispectral imagery spanning four different satellites with spatial resolution varying from 1.24m to 1.65m. Since the imagery corresponding to the entire CHAL region is expensive, we identified subregions for acquiring images so as to capture human activity, diverse vegetation, and varying elevation levels observed in this area. A total of 53 images were acquired, with each image captured in the year 2017 or 2018 and month falling in one of the survey periods (a map with the images is provided in the supplement). The main challenge is that there are very few images that satisfy the cloud cover (< 5%) and time period criteria. We have categorized the images based on the season they were acquired. We have 28 images from the March-April period referred to as "Spring" images. The remaining are from the November-January period referred to as "Winter" images. The number of survey points corresponding to each plant varies and we note that most of the survey points belong to the Winter category. It is to be noted that there is hardly any overlap between the acquired images as our goal was to select images covering as maximum an area as possible. A detailed description of survey data, distribution of absence-presence data and satellite image data are presented in the supplementary material.

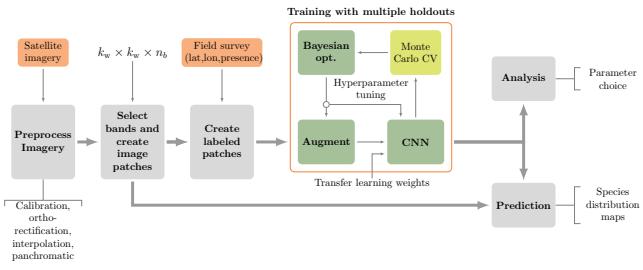


Figure 1: Modeling framework: The modules correspond to data preparation, feature vector generation, training, and model selections, followed by application of the models to generate species distribution maps and other analyses.

Workflow. The workflow is illustrated in Figure 1. Image patches of $k_w \times k_w \times n_b$ pixels with the given coordinates at the center are extracted from the satellite image. Here, k_w corresponds to the patch (or window) size and n_b corresponds to the number of satellite bands. The problem of predicting whether the target species is present in a given location is formulated as a binary classification problem, where the patch corresponding to this location is the input to the classifier. Given a patch $\mathbf{X} \in \mathbb{R}^{k_w \times k_w \times n_b}$, the i^{th} convolutional layer performs P affine transformations of the input features $\mathbf{X}^{(i)}$ followed by a nonlinear activation function τ : $\mathbf{X}_p^{(i+1)} = \tau(\mathbf{W}_p^{(i)} \mathbf{X}^{(i)} + b_p^{(i)})$, where $\mathbf{W}_p^{(i)}$ is a matrix that corresponds to the convolution operation performed by the p^{th} spatially localized filter $\mathbf{w}_p^{(i)}$ and b_p is the bias parameter. The output layer is a dense layer that provides a score $s_p^{(i)}$ between $[0, 1]$. The filters minimize the cross-entropy cost function: $\sum v_i \log s_p^{(i)} + (1 - v_i) \log (1 - s_p^{(i)})$, where $v_i \in \{0, 1\}$ is the label. We experimented with $k_w = 32, 64, 128, 256$, and four frequency bands: Near-Infrared (n), Red (r), Green (g), and Blue (b).

Models for classification. As for the classifier, we considered several CNNs – a class of shallow networks that we defined and *de facto* standard visual recognition networks such as VGG-16 (Simonyan and Zisserman 2014), InceptionV3 (Szegedy et al. 2016), and Xception (Chollet 2017) which have been designed for classifying RGB images into multiple classes. As a baseline, we used the random forest model (RF) (Breiman 2001). We had to modify each of these networks for our application. Firstly, these models were originally designed for multiclass image classification. We appended a final dense layer with softmax function as the activation function. It provides a single score between 0 and 1 indicating the likelihood of presence. Secondly, these networks are designed for RGB images, and therefore, support only three bands. For the nrgb mode, we expanded the first convolution layer to accommodate the fourth band. Details of the shallow networks is provided in the supplement.

Transfer learning and augmentation. To overcome the lack of training data, various transfer learning (Pan and Yang 2009) techniques have been applied in deep learning for

remote sensing (Ball, Anderson, and Wei 2018). Transfer learning corresponds to transferring knowledge from models trained in a feature space with abundant data and applying it to the domain of interest. Here, we follow the approach of Perez et al. (Perez et al. 2017). The VGG-16, InceptionV3 and Xception are initialized with weights from the ImageNet dataset (Russakovsky et al. 2015). Since ImageNet is a collection of RGB images, weights are only available for the RGB bands. Accordingly, for the RGB bands, we use the corresponding weights from the pre-trained model. The mean of the RGB weights is applied for the near-infrared. In addition, we use data augmentation by applying geometric transformations (flipping and rotation) to the feature vectors.

Model exploration and selection. The goal here is to obtain predictions that are robust (reduce the chance of bias) to the choice of the hold-out set. The process is outlined in Algorithm 1, Given a species and a network, the set of labeled feature vectors fed to the training framework is determined by the window size (k_w) for each patch, the bands in the satellite imagery and the data collection period. For this input, the framework uses a multiple-hold-out approach for model exploration (Monteiro et al. 2016). This is an iterative process which results in a different model for each iteration. In one iteration, we first keep aside 20% of the dataset as the *hold-out* or *test set* by randomly sampling data points ensuring a 1 : 1 presence to absence ratio. The rest of the dataset corresponds to the *optimization set*, used to train and validate models during the model exploration process. Using the optimization set, we tune the hyperparameters batch size (16, 32 and 64), learning rate (sampled from a log-uniform distribution with lower and upper bounds 0.0001 and 0.01, respectively) and augmentation choice (binary) via Bayesian optimization using Gaussian processes (Snoek, Larochelle, and Adams 2012) (50 iterations). In one iteration of the optimization process, a combination of the hyperparameter values is chosen to be evaluated. To overcome overfitting issues due to limited data, we use the Monte Carlo cross-validation (MCCV) technique (Burman 1989) for evaluating this set of hyperparameter values. In the MCCV approach, multiple train-validate splits of the optimization set are created by randomly picking the data points for each group (80-20 split of the optimization dataset). In order to evaluate the candidate set of hyperparameter values, we consider 20 train-validate splits. For each split, training is performed by minimizing the binary cross-entropy loss function using the stochastic gradient descent method. The mean validation F1 score for the 20 splits is used to score the instance. Finally, using the hyperparameter values obtained from the tuning process, we train the CNN using the optimization set and evaluate it using the hold-out test dataset. We consider five rounds of hyperparameter tuning in the multiple-hold-out process, yielding five top models for each pair. For the RF model, we apply the same process with hyperparameters being number of estimators (200 to 2000), maximum features (auto, square-root), maximum depth (10 to 100), bootstrapping (binary), and a minimum number of samples per leaf (1, 2, and 4).

Algorithm 1: Multiple hold-out model exploration with Bayesian optimization.

Input : Window size k_w , subset of bands, convolutional neural network $G(k_w, n_b)$ with pre-trained weights (if any), labeled patches $\mathbf{X} = \{(X_1, \ell_1), (X_2, \ell_2), \dots, (X_n, \ell_n)\}$, with patch $X_i \in \mathbb{R}^{k_w \times k_w \times n_b}$ and label $\ell_i \in \{0, 1\}$, number of hold-out iterations η , number of MCCV train-validate splits ω .

Output: Models $\{\mathcal{M}_{G,j} \mid 1 \leq j \leq t\}$ with performance evaluation.

```

// Multiple hold-out iterations
1 for  $j = 1, 2, \dots, \eta$  do
2   Prepare test dataset  $\mathbf{X}_j^{\text{test}}$  by randomly choosing 20% of the input with equal number of presence and absence points.
3   Let  $\mathbf{X}_j^{\text{opt}} = \mathbf{X} \setminus \mathbf{X}_j^{\text{test}}$  denote the optimization set.
4   hyperparameters = BAYESOPT-TUNE( $G, \mathbf{X}^{\text{opt}}$ , loss function: -avg. F1 score from MCCV)
5    $\mathcal{M}_{G,j} = \text{TRAIN}(G, \mathbf{X}^{\text{opt}}, \text{hyperparameters})$ 
6 end
7 for  $j = 1, 2, \dots, \eta$  do
8   eval[j] = EVALUATE( $\mathcal{M}_{G,j}, \mathbf{X}^{\text{test}}$ )
9 end

```

Results

Classification performance and choice of models. We discuss the performance of the classifiers trained on the labeled data generated from all the images and incidence data from the three surveys. Recall that five rounds of training and hyperparameter tuning were conducted for each patch size (k_w), a subset of frequency bands, and CNN. We averaged the performance of the resulting models that emerge out of model exploration in each round of the optimization process. The results for different CNNs for $k_w = 64$ and bands n_{rgb} are summarized in Table 1. For all species and networks, the standard deviation of the test F1 score for the models is low, showing consistency in performance across different rounds of training. Among CNNs, models initialized with ImageNet weights generally perform well. Further, we note that InceptionV3, InceptionV3-P, Xception and Xception-P consistently appear among the top-performing models. However, many other models, including the baseline RF, are not far behind. One observation about the RF model is the high training accuracy, indicating some overfitting. We also note that for the class of models $\mathcal{M}(\ell, \chi)$, as the number of layers ℓ and number of channels χ are increased, the performance generally increases.

Patch size. For all the three species and the three top models, we observed that as k_w is increased from 32 to 256, the average test F1 score increases (Figure 2 top row). The larger the patch size, the greater the coverage area, and therefore, the chances of not capturing the plant within the patch due to an error in recording the coordinates are reduced. Also, if the plant is shrub-like, more plants will be accounted for, leading to better detection. However, with increasing patch

size, there is a higher chance that two patches overlap due to their corresponding survey points being close to each other. If one of the coordinates is labeled 1 and the other 0, then there is a possibility that at least one of the patches might be misclassified due to the overlap. If both coordinates are labeled 1, the likelihood of correctly classifying the corresponding patches increases if the specimen is present in the overlapping region. To investigate further, we computed the Haversine distance between every pair of survey locations for each species. Then, we counted the number of instances where (a) two patches overlap and one patch is labeled 1 while the other is labeled 0 (mismatched pairs) and (b) both patches are labeled the same (matched pairs). This happens when two survey locations are at a distance at most $\sqrt{2} \cdot k_w$. This is plotted in Figure 2 (bottom row). We observe that the number of overlapping pairs (matched or mismatched) increases with patch size. There is a significant increase in such pairs from 64 to 128 and 128 to 256. Hence, we have fixed the patch size to be 64 for creating the distribution map and further analysis.

Seasonality. The imagery dataset can be categorized into two seasons: winter and spring (see Table S1 in the supplement). Most of the survey points belong to images captured in winter. To study how seasonal variations in the appearance of plants can affect classification performance, we applied two approaches. Throughout, we considered the following two models: InceptionV3-P and Xception-P. In the first approach, we used only images acquired in winter. Omission of image patches corresponding to summer can affect the results in two ways. Firstly, the variation in appearance is reduced, which could lead to an improvement in the performance. At the same time, since the number of labeled patches available for training reduces, it could lead to overfitting and a decrease in performance. The available data points relative to the entire dataset are approximately 88% for *C. odorata*, 90% for *L. camara*, and 75% for *P. hysterophorus*. The results are in Table 2 (1st column). Comparing with the results in Table 1 corresponding to all data points for InceptionV3-P and Xception-P (reproduced in the 3rd column of Table 2), we see that the performance is significantly better in the case of *P. hysterophorus*. In the second approach, we added an extra channel to the input of the CNN, which we refer to as the *season channel*. Recall that the channel dimension is $k_w \times k_w$, the spatial dimension of the image. All the elements of the season channel were set to 1 if the image was acquired in winter, 0 otherwise. The idea here is that the model accounts for the season information during the classification. The results are in the second column of Table 2. We observe that this method yields inferior performance compared to the first approach.

Species distribution maps. We generated a high-resolution grid ($\approx 50\text{m} \times 50\text{m}$) and retained only those points which were (i) present in the satellite images and (ii) near major roads using the Open Street Map database². Both major and minor roads were considered. We deliberately restricted our predictions to points close to roads due

²<https://www.openstreetmap.org>

Table 1: Average performance of representative classifiers for patch size 64 and nrgb bands across five different hold-out sets. In each case, the top mean F1 scores are highlighted for each species. Full table is in the supplement.

species	model	Train acc.		Test		
		accuracy	precision	recall	mean F1 score	std. dev.
<i>C. odorata</i>	InceptionV3	0.747	0.713	0.683	0.8	0.732
	InceptionV3-P	0.919	0.829	0.799	0.884	0.838
	VGG-16-P	0.797	0.755	0.725	0.832	0.771
	Xception	0.801	0.761	0.729	0.832	0.773
	Xception-P	0.963	0.855	0.838	0.884	0.86
	RF	1	0.676	0.641	0.811	0.715
<i>L. camara</i>	InceptionV3	0.804	0.713	0.678	0.843	0.745
	InceptionV3-P	0.863	0.788	0.768	0.828	0.795
	VGG-16-P	0.776	0.704	0.705	0.698	0.7
	Xception	0.836	0.808	0.806	0.822	0.809
	Xception-P	0.916	0.803	0.792	0.822	0.805
	RF	1	0.701	0.635	0.935	0.756
<i>P. hysterophorus</i>	InceptionV3	0.837	0.839	0.855	0.817	0.835
	InceptionV3-P	0.907	0.814	0.826	0.789	0.807
	VGG-16-P	0.851	0.763	0.795	0.703	0.746
	Xception	0.842	0.828	0.827	0.829	0.825
	Xception-P	0.909	0.817	0.842	0.793	0.81
	RF	0.994	0.758	0.777	0.714	0.744

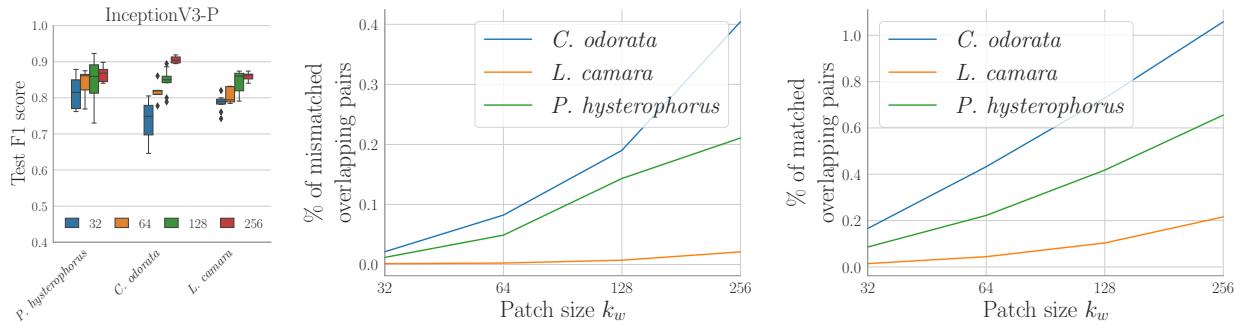


Figure 2: The first plot corresponds to performance of InceptionV3-P with increasing patch size. We have a similar result for Xception-P in the supplement. The other two plots show the extent of overlap between image patches as the patch size is increased.

to roadside bias in our field survey. With each chosen point as the centroid, we constructed patches of size $k_w = 64$. For each of the CNN models under consideration, we obtained the softmax values at each location. The average of these softmax values (bagged predictor) was used as an estimate of the probability of presence of the target species in that location (Breiman 2001). The results of the ensemble prediction are in Figure 4 for *C. odorata*. For the remaining species, the maps are in the supplement. The entire study region is divided into 14 subregions. The details of each subregion can be found in Figure 4. The aggregated prediction by administrative districts is in Figure 4 for *C. odorata* and Figure 5 for the other two species.

Our results indicate that all the species are widespread in the CHAL region on the roadsides. Generally, infestation is observed more in the southern part of the CHAL region: Nawalparasi, Chitwan, and Makwanpur. These districts are

in the Terai region, which is closer to the border with India. Northern parts of Nepal are at much higher elevations and are generally considered unsuitable for the establishment and growth of the studied IAS. The only exception is the case of *C. odorata* in Rasuwa. Our results indicate widespread prevalence in the Rasuwa district (Figure 5, Sub-region 4), which is higher than the elevation ranges accepted as suitable for *C. odorata* (Shrestha 2016). Also, upon inspection, experts on the ground believed that for the Tanahun region (Subregions 5, 6, and 7 in Figure 4), the map exhibits a lower infestation of *C. odorata* than observed. On the other hand, for the Nawalparasi region (Subregion 10), our map suggests more infestation of *L. camara* than observed on the ground.

Analysis of performance with respect to model parameters. In this analysis, we present results for InceptionV3-P and Xception-P. To study the importance of the multi-

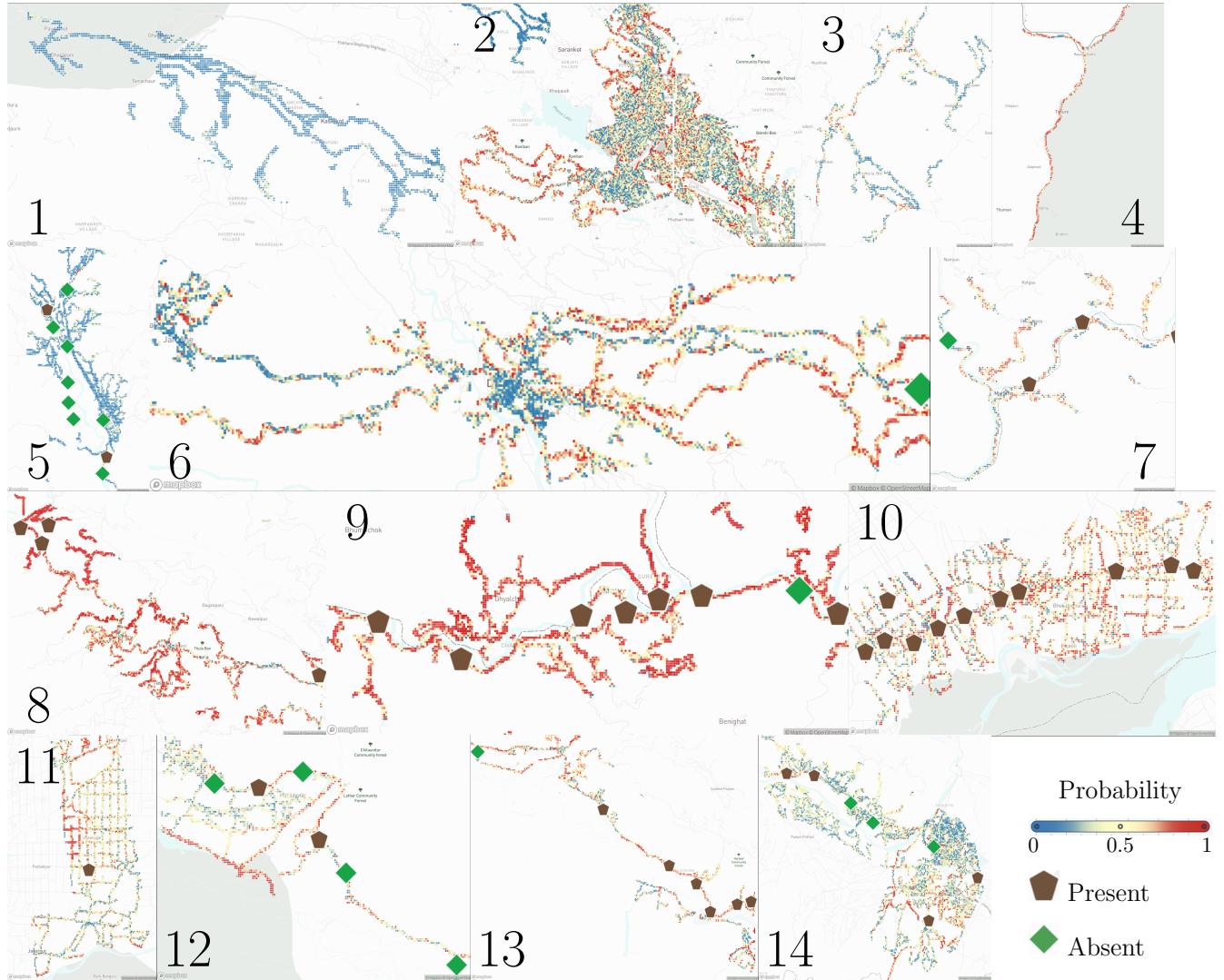


Figure 3

Figure 4: Predicted distribution and survey locations for *C. odorata*. The distribution maps for the other two species are in the supplement.

spectral bands, we performed feature ablation. We trained the classifiers by dropping the target band. The performance of these classifiers was compared to the reference model with bands=nrgb. The results are in Figure S6 in the supplement. In the case of InceptionV3-P, we see a slight degradation in performance when the r band is omitted for *C. odorata* and *L. camara*. We also analyzed the hyperparameters of the models derived from the Bayesian optimization process (Figure S7 in the supplement). We observed that for almost all models corresponding to *P. hysterophorus* and *L. camara*, the batch size was 8. For *C. odorata*, the batch size of most models was 32. For these models, we observe that the training accuracy is relatively higher than in the other cases. This is in line with the general observation that large-batch training tends to over-fit the models (Keskar

et al. 2017). We observed a large variation across models for the learning rate. Also, augmentation yielded superior performance in most cases.

Related work

The recent success of CNNs in image classification tasks has led to their widespread adoption for remote sensing data in the agricultural domain. Kiliaris and Prenafeta-Boldú (Kiliaris and Prenafeta-Boldú 2018) provide an overview of the use of CNNs in agriculture-based applications, including those that pertain to remote sensing imagery. Jia et al. (2017) use several one-class classification models based on sparse-autoencoders to identify and map major crops using high-resolution multi-spectral satellite data. Kussul et al. (2017) develop an ensemble of CNN classifiers

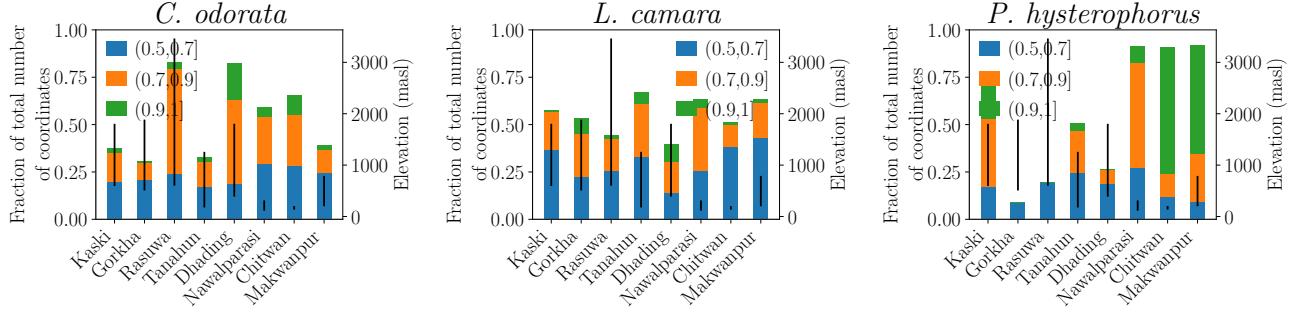


Figure 5: The extent of invasion by district. Only locations with a predicted probability > 0.5 have been accounted for. Further, among these points, the locations are divided into three bins based on their probability values: $(0.5, 0.7]$, $(0.7, 0.9]$, and $(0.9, 1]$. For reference, the elevation range in meters above sea level (masl) is plotted in black.

Table 2: Seasonality: Average test F1 score for classifiers with patch size 64. The first column corresponds to training with only winter images. The second column corresponds to using an additional channel to encode season information. The third column contains values from Table 1 for reference.

species	model	Winter images	Season channel	All
<i>C. odorata</i>	InceptionV3-P	0.81	0.771	0.838
	Xception-P	0.813	0.777	0.86
<i>L. camara</i>	InceptionV3-P	0.808	0.786	0.795
	Xception-P	0.826	0.828	0.805
<i>P. hysterophorus</i>	InceptionV3-P	0.843	0.798	0.807
	Xception-P	0.853	0.812	0.81

for identifying various crop types with the classifier trained on multitemporal multisource satellite imagery. However, unlike regularly laid out croplands, invasive plant detection is much more challenging due to a lack of regularity in their spatial extent. In recent years, landcover classification using CNNs has become popular. In early 2020, the SpaceNet 6 dataset and challenge (Shermeyer et al. 2020) were announced. The dataset comprises Synthetic Aperture Radar data and optical data from WorldView-2 (similar to what we have used). The top five models of this challenge used U-net encoder-based models. Most of these models used pre-training using the ImageNet data.

Our work highlights the importance of plant phenology as a limiting factor in identifying plants using remote-sensed images. The importance of timing of image acquisition is a vital topic. In fact, phenological differences between species are considered an effective means of identifying invasive plants (Bradley 2014). However, we would require a time series of remotely sensed imagery, to take advantage of this phenomenon. This is not always feasible due to the limited availability of such images and the considerable expense incurred when relying on commercial datasets. Also, given enough locations for which patches from both seasons are available, transductive transfer learning techniques can be

Table 3: Subregions corresponding to each satellite image are grouped together by their district. Also provided is the elevation range in meters above sea level (masl).

District	Subregions	Elevation range (masl) min	max
Kaski	1,2	594	1804
Gorkha	3	506	1885
Rasuwa	4	600	3464
Tanahun	5,6,7	176	1258
Dhading	8,9	387	1804
Nawalparasi	10	113	313
Chitwan	11	135	203
Makwanpur	12,13,14	200	790

applied to utilize models developed for one season to train models on datasets for other seasons (Pan and Yang 2009).

Conclusion

There are several directions to be explored to enhance the framework’s performance despite data limitations that were brought up in the above discussions. Like our work, most recent studies on mapping invasive species in the Himalayan region involve simultaneously surveying multiple invasive species. In such instances, inductive transfer learning and multi-task learning techniques can be applied to obtain classifiers (Pan and Yang 2009). Another direction is to consider using bioclimatic variables frequently used in ecological suitability models. One could even consider incorporating the output of ecological suitability models or mechanistic growth models. This is in line with a recent body of work on theory-guided machine learning (Karpatne et al. 2017; Fox et al. 2019), which attempts to leverage scientific knowledge for improving data-driven models. Understanding invasive species spread is critical to environmental health and social well-being. Our work demonstrates the feasibility of Modern AI methods in effectively mapping invasive species using remote sensing data, thus aiding in monitoring them and curtailing their spread.

References

- Allan, S.; Shi, B.; Adkins, S. W.; et al. 2018. Impact of Parthenium Weed on Human and Animal Health. *Parthenium Weed: Biology, Ecology and Management*, 7: 105.
- Bajwa, A. A.; Chauhan, B. S.; Farooq, M.; Shabbir, A.; and Adkins, S. W. 2016. What do we really know about alien plant invasion? A review of the invasion mechanism of one of the world's worst weeds. *Planta*, 244(1): 39–57.
- Ball, J. E.; Anderson, D. T.; and Chan Sr, C. S. 2017. Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community. *Journal of Applied Remote Sensing*, 11(4): 042609.
- Ball, J. E.; Anderson, D. T.; and Wei, P. 2018. State-of-the-art and gaps for deep learning on limited training data in remote sensing. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, 4119–4122. IEEE.
- Bhagwat, S. A.; Breman, E.; Thekaekara, T.; Thornton, T. F.; and Willis, K. J. 2012. A battle lost? Report on two centuries of invasion and management of *Lantana camara* L. in Australia, India and South Africa. *PLoS One*, 7(3).
- Bradley, B. A. 2014. Remote detection of invasive plants: a review of spectral, textural and phenological approaches. *Biological invasions*, 16(7): 1411–1425.
- Breiman, L. 2001. Random forests. *Machine learning*, 45(1): 5–32.
- Burman, P. 1989. A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods. *Biometrika*, 76(3): 503–514.
- Chollet, F. 2017. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1251–1258.
- Chollet, F.; et al. 2015. Keras. <https://github.com/fchollet/keras>.
- Diagne; et al. 2021. High and rising economic costs of biological invasions worldwide. *Nature*, 592(7855): 571–576.
- Fox; et al. 2019. Learning everywhere: Pervasive machine learning for effective high-performance computation. In *2019 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*, 422–429. IEEE.
- Grizonnet, M.; Michel, J.; Poughon, V.; Inglada, J.; Savinaud, M.; and Cresson, R. 2017. Orfeo ToolBox: Open source processing of remote sensing images. *Open Geospatial Data, Software and Standards*, 2(1): 15.
- He et al., K. S. 2015. Will remote sensing shape the next generation of species distribution models? *Remote Sensing in Ecology and Conservation*, 1(1): 4–18.
- Jarvis; A; Reuter, H. I.; Nelson, A.; and Guevara, E. 2008. Hole-filled seamless SRTM data V4, International Centre for Tropical Agriculture (CIAT). <http://srtm.csi.cgiar.org/>.
- Jia; et al. 2017. Joint sparse auto-encoder: A semi-supervised spatio-temporal approach in mapping large-scale croplands. In *2017 IEEE International Conference on Big Data (Big Data)*, 1173–1182. IEEE.
- Kamilaris, A.; and Prenafeta-Boldú, F. 2018. A review of the use of convolutional neural networks in agriculture. *The Journal of Agricultural Science*, 156(3): 312–322.
- Karpatne; et al. 2017. Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Transactions on knowledge and data engineering*, 29(10): 2318–2331.
- Kattenborn, T.; Eichel, J.; and Fassnacht, F. E. 2019. Convolutional Neural Networks enable efficient, accurate and fine-grained segmentation of plant species and communities from high-resolution UAV imagery. *Scientific reports*, 9(1): 1–9.
- Keskar; et al. 2017. On large-batch training for deep learning: Generalization gap and sharp minima. In *5th International Conference on Learning Representations, ICLR 2017*.
- Kislov, D. E.; and Korznikov, K. A. 2020. Automatic windthrow detection using very-high-resolution satellite imagery and deep learning. *Remote Sensing*, 12(7): 1145.
- Korznikov; et al. 2021. Using U-Net-Like Deep Convolutional Neural Networks for Precise Tree Recognition in Very High Resolution RGB (Red, Green, Blue) Satellite Images. *Forests*, 12(1): 66.
- Kussul; et al. 2017. Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, 14(5): 778–782.
- Liang, S. 2005. *Quantitative remote sensing of land surfaces*, volume 30. John Wiley & Sons.
- Lowe, S.; Browne, M.; Boudjelas, S.; and De Poorter, M. 2000. *100 of the world's worst invasive alien species: a selection from the global invasive species database*, volume 12. Invasive Species Specialist Group Auckland, New Zealand.
- Mack, R.; and Smith, M. 2011. Invasive plants as catalysts for the spread of human parasites. *NeoBiota*, 9: 13.
- Mather, P. M.; and Koch, M. 2011. *Computer processing of remotely-sensed images: an introduction*. John Wiley & Sons.
- Monteiro; et al. 2016. A multiple hold-out framework for Sparse Partial Least Squares. *Journal of neuroscience methods*, 271: 182–194.
- Muniappan, R.; Reddy, G.; and Lai, P.-Y. 2005. Distribution and biological control of *Chromolaena odorata*. In *Invasive plants: ecological and agricultural aspects*, 223–233. Springer.
- Pan, S. J.; and Yang, Q. 2009. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10): 1345–1359.
- Perez; et al. 2017. Poverty prediction with public landsat 7 satellite imagery and machine learning. *arXiv preprint arXiv:1711.03654*.
- Pyšek, P.; and Richardson, D. M. 2010. Invasive species, environmental change and management, and health. *Annual Review of Environment and Resources*, 35: 25–55.
- Russakovsky; et al. 2015. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3): 211–252.

Shermeyer; et al. 2020. SpaceNet 6: Multi-sensor all weather mapping dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 196–197.

Shrestha, B. B. 2016. Invasive alien plant species in Nepal. *Frontiers of botany. Central Department of Botany, Tribhuvan University, Kirtipur, Kathmandu*, 269–284.

Shrestha, B. B.; Pokhrel, K.; Paudel, N.; Poudel, S.; Shabbir, A.; and Adkins, S. W. 2019. Distribution of *Parthenium hysterophorus* and one of its biological control agents (Coleoptera: *Zygogramma bicolorata*) in Nepal. *Weed Research*, 59(6): 467–478.

Simonyan, K.; and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Snoek, J.; Larochelle, H.; and Adams, R. P. 2012. Practical bayesian optimization of machine learning algorithms. In *Advances in neural information processing systems*, 2951–2959.

Szegedy; et al. 2016. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2818–2826.

Te Beest, M.; Cromsigt, J. P.; Ngobese, J.; and Olff, H. 2012. Managing invasions at the cost of native habitat? An experimental test of the impact of fire on the invasion of *Chromolaena odorata* in a South African savanna. *Biological Invasions*, 14(3): 607–618.

Technologies, M. 2020. Datasheet for absolute radiometric calibration. <https://blog.maxar.com/earth-intelligence/2020/absolute-radiometric-calibration-is-an-essential-tool-to-imagery-science-but-what-is-it>.

Tiwari, S. 2005. *An inventory and assessment of invasive alien plant species of Nepal*. IUCN Nepal.

UN. 2019. Sustainable development goals.

Van Rossum, G.; and Drake, F. L. 2009. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace. ISBN 1441412697.

Wang, L.; et al. 2008. Invasive species spread mapping using multi-resolution remote sensing data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37: 135–142.

Zhang, L.; Zhang, L.; and Du, B. 2016. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and Remote Sensing Magazine*, 4(2): 22–40.

Supplementary material: A Robust Deep Learning Model Reveals the Spread of Multiple Invasive Plants in a Biodiversity Hotspot using Satellite Imagery

Additional details for Framework



(a) *P. hysterophorus*: (1) dense (2) moderate and (3) low.



(b) *L. camara*: (1) dense (2) moderate and (3) low.



(c) *C. odorata*: (1) dense (2) moderate and (3) low.

Figure S1: Pictures of representative survey locations with the intensity of occurrence. Here, dense means a $10m \times 10m$ coverage or more, and low is $2m \times 2m$ or less. This figure also reveals one of the challenges this variation in distribution creates for training reliable models.

Table S1: Field survey data used in the experiments. It is to be noted that this is the subset of the total survey points that overlaps with the imagery data. The first set corresponds to all such data points. These are further classified into points present in images captured during spring and during autumn and winter.

Species	All images		Spring		Winter	
	Total	Presence	Total	Presence	Total	Presence
<i>Chromolaena odorata</i>	445	236	54	46	391	190
<i>Lantana camara</i>	661	366	70	2	591	364
<i>Parthenium hysterophorus</i>	383	186	96	60	287	126

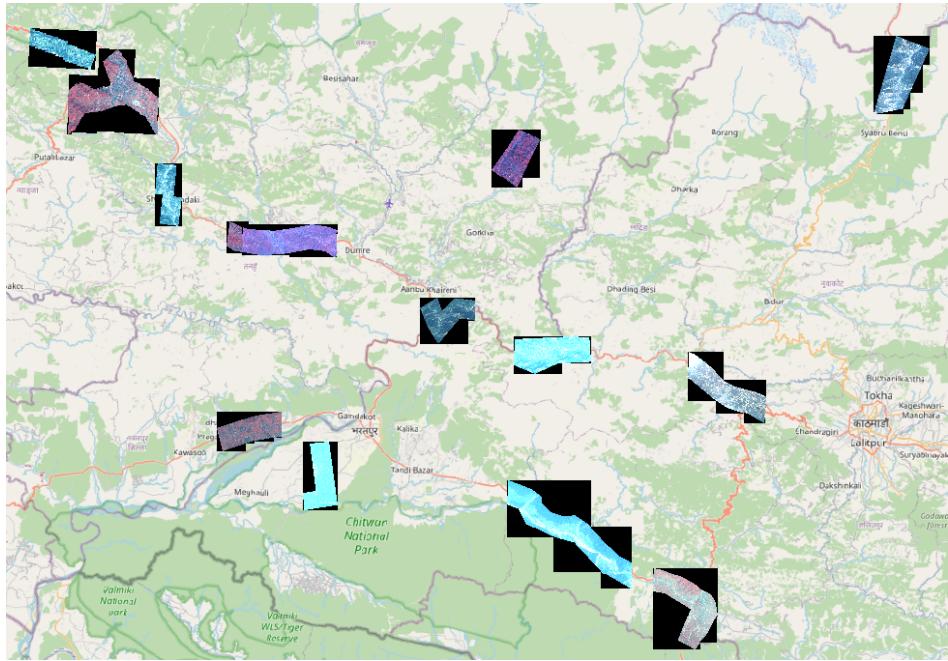


Figure S2: The CHAL area and more than 50 images constitute our image dataset (before calibration).

Table S2: Satellite imagery dataset comprises 54 images covering approximately 1,400km² across nine locations.

Satellite	Resolution	Bands	Panchrom.
WorldView-4	1.24m	4	0.31m
WorldView-3	1.24m	8	0.31m
WorldView-2	1.84m	8	0.46m
GeoEye-1	1.65m	4	0.41m

Study species. *Chromolaena odorata*, *Lantana camara*, and *Parthenium hysterophorus* are tropical plant species native to central and south America, and now widespread in Africa, Asia, and Australia. They are among the most significant global invasive alien plant species (Bajwa et al. 2016; Lowe et al. 2000; Muniappan, Reddy, and Lai 2005; Bhagwat et al. 2012) and threaten natural ecosystems, agrosystems, and human health (Allan et al. 2018). *P. hysterophorus* is a short-lived annual, diffused leafy herb germinating at any time of the year and growing rapidly to a height of 0.5–2.0m. It flowers almost year-round, particularly from March to December. It was first recorded in Nepal in 1967 (Tiwari 2005). It inhabits a wide range of habitats, typically including grazing lands, summer crops, disturbed and cultivated areas, roadsides, wastelands, cleared and grazed pastures, river-banks, and open spaces in urban areas (Shrestha et al. 2019). *C. odorata* is a fast-growing perennial shrub that reaches up to 2.5m, with flowering and fruiting occurring from December to April. It was first recorded in 1956. It forms dense patches on roadsides, fallow lands, shrub-lands, agricultural lands, grasslands, tree plantations, and forests with low foliage density and low canopy cover (Te Beest et al. 2012). *L. camara* was first recorded in Nepal in 1966 (Tiwari 2005). It is an erect or suberect shrub up to 3m high and grows well on roadsides, fallow lands, shrub lands, pastures, and forest. It flowers and fruits almost throughout the year. It is believed that many native species of grasses have disappeared after the introduction of *L. camara* (Tiwari 2005). In tropical regions, it harbors pests that affect human health by providing shelter during the day for tsetse flies which are vectors for African sleeping sickness (Mack and Smith 2011).

The shallow networks, denoted by $\mathcal{M}(\ell, \chi)$ comprise up to ℓ convolutional layers with a receptive field of size 3×3 pixels, followed by dense layers. The first convolution layer takes a patch as input, and with a convolution stride of 1 produces χ feature maps or *channels*. We progressively double the depth of the feature map in the subsequent layers. For example, $\mathcal{M}(4, 8)$ comprises four convolution layers with 8, 16, 32, and 64 channels for layers 1, 2, 3 and 4 respectively. The output of the final convolutional layer is vectorized using a flattening layer so that the features can be fed to the dense layers. The final layer is a dense layer with softmax function as the activation and outputs the likelihood of the presence of the IAPS. Since available data for training is scarce, we run the risk of overfitting the training data. To avoid overfitting, we used batch-normalization and dropout as a means of regularization. We also incorporated max-pooling layers, which downsamples the input features in the spatial extent.

Image preprocessing and data preparation. The acquired satellite data consists of standard imagery, to which relative radiometric correction has been applied. Considering that these images were acquired using multiple satellite sources and at different times in the year, they must be calibrated so that they can be physically interpreted and comparable with each other (Mather and Koch 2011; Liang 2005). We applied absolute radiometric correction and ortho-rectification. Finally, we performed pansharpening by merging the given images with the corresponding spatially high-resolution panchromatic images. For this, we used image-specific data from the XML files accompanying the images, gain-bias and solar illumination data from the most current data sheet (Technologies 2020), and digital elevation model (Jarvis et al. 2008). Due to multiple satellite sources, our images also vary in spatial resolution: 0.31m to 0.41m pixel length. We used the bivariate cubic spline interpolation method to convert coarser images to 0.31m pixel length images. The data preparation process is fairly generic and allows for easy addition of relevant information such as slope, temperature, humidity, etc., based on their availability for future work. In order to ensure that adequate area is captured in a survey location to account for sparse coverage of the plant and inaccuracy of the GPS device, we used a patch-based method for classification. A *patch* is a square window of size $k_w \times k_w$ pixels with the given coordinates at the center, where we refer to k_w as the patch (or window) size. We experimented with $k_w = 32, 64, 128, 256$. We considered four frequency bands: Near-Infrared (n), Red (r), Green (g), and Blue (b). To assess the importance of each band, in our experiments, we considered different subsets of the four bands: nrgb, rgb, nrg, nrb, and ngb. Suppose n_b is the total number of considered bands (four for nrgb and three for the rest), then each feature vector corresponds to a $k_w \times k_w \times n_b$ matrix, which we henceforth refer to as a *patch*. The data preparation process is fairly generic and allows for easy addition of relevant information such as slope, temperature, humidity, etc., based on their availability for future work.

Implementation and computing environment. The entire framework was implemented using Python 3.7 (Van Rossum and Drake 2009). The satellite image preprocessing was achieved using the Orfeo Toolbox (Grizonnet et al. 2017). The CNNs were implemented using Keras (Chollet et al. 2015). Compute nodes with 2 x Intel(R) Xeon(R) Gold 6248 CPU @ 2.50GHz processors with 20 cores per CPU; 386GB memory; and 4xNVIDIA Volta V100 GPUs with 32GB GPU memory using a Lustre filesystem on a Mellanox MT27800 Family [ConnectX-5] infiniband backend network were used for training.

Additional details for Results

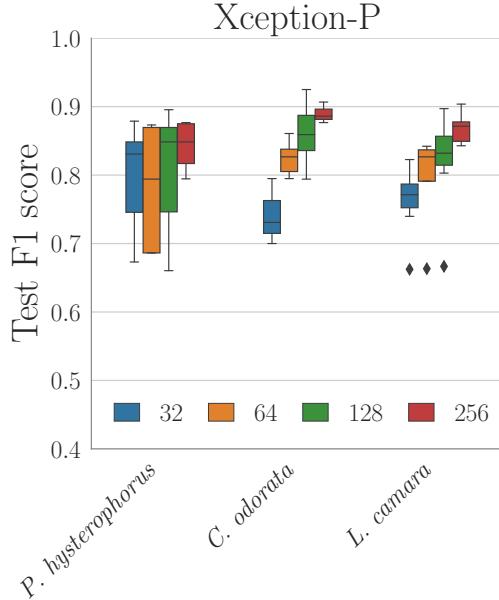


Figure S3: The performance of Xception-P with increasing patch size.

Table S3: Average performance of classifiers for patch size 64 and nrgb bands across five different hold-out sets.

species	model	Train acc.		Test		
		accuracy	precision	recall	mean F1 score	std. dev.
<i>C. odorata</i>	$\mathcal{M}(2, 16)$	0.705	0.697	0.692	0.732	0.705
	$\mathcal{M}(2, 4)$	0.641	0.666	0.642	0.789	0.703
	$\mathcal{M}(4, 16)$	0.742	0.726	0.731	0.737	0.725
	$\mathcal{M}(4, 4)$	0.727	0.721	0.744	0.689	0.712
	InceptionV3	0.747	0.713	0.683	0.8	0.732
	InceptionV3-P	0.919	0.829	0.799	0.884	0.838
	VGG-16	0.587	0.5	0.5	1	0.667
	VGG-16-P	0.797	0.755	0.725	0.832	0.771
	Xception	0.801	0.761	0.729	0.832	0.773
	Xception-P	0.963	0.855	0.838	0.884	0.86
<i>L. camara</i>	RF	1	0.676	0.641	0.811	0.715
	$\mathcal{M}(2, 16)$	0.743	0.707	0.651	0.898	0.753
	$\mathcal{M}(2, 4)$	0.722	0.678	0.618	0.917	0.739
	$\mathcal{M}(4, 16)$	0.768	0.751	0.712	0.84	0.77
	$\mathcal{M}(4, 4)$	0.754	0.71	0.664	0.852	0.744
	InceptionV3	0.804	0.713	0.678	0.843	0.745
	InceptionV3-P	0.863	0.788	0.768	0.828	0.795
	VGG-16	0.658	0.496	0.496	1	0.663
	VGG-16-P	0.776	0.704	0.705	0.698	0.7
	Xception	0.836	0.808	0.806	0.822	0.809
<i>P. hysterophorus</i>	Xception-P	0.916	0.803	0.792	0.822	0.805
	RF	1	0.701	0.635	0.935	0.756
	$\mathcal{M}(2, 16)$	0.821	0.814	0.894	0.709	0.789
	$\mathcal{M}(2, 4)$	0.803	0.796	0.895	0.664	0.762
	$\mathcal{M}(4, 16)$	0.826	0.811	0.892	0.709	0.787
	$\mathcal{M}(4, 4)$	0.819	0.814	0.91	0.697	0.786
	InceptionV3	0.837	0.839	0.855	0.817	0.835
	InceptionV3-P	0.907	0.814	0.826	0.789	0.807
	VGG-16	0.618	0.613	0.633	0.629	0.585
	VGG-16-P	0.851	0.763	0.795	0.703	0.746

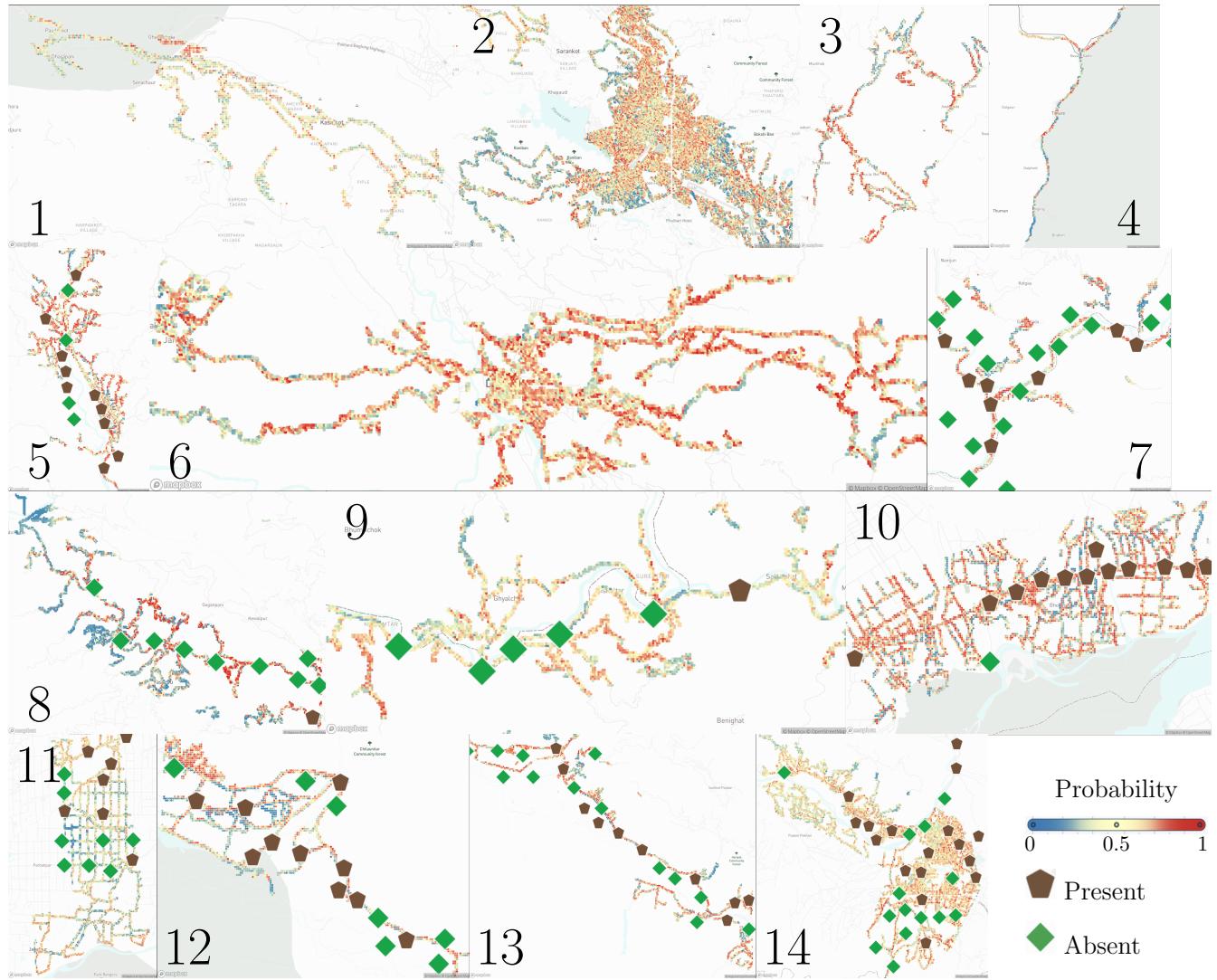


Figure S4: Predicted distribution and survey locations for *L. camara*.

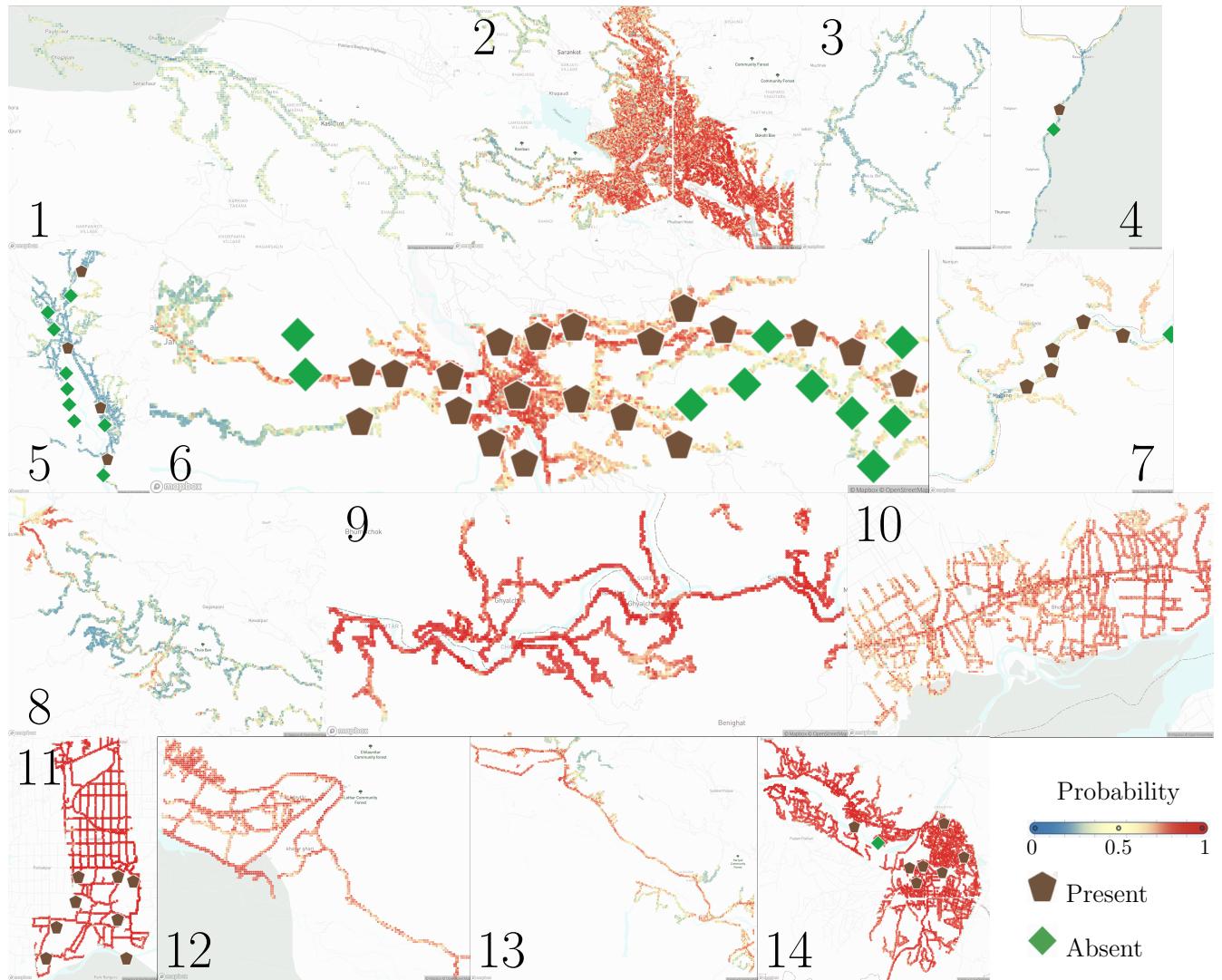


Figure S5: Predicted distribution and survey locations for *P. hysterophorus*.

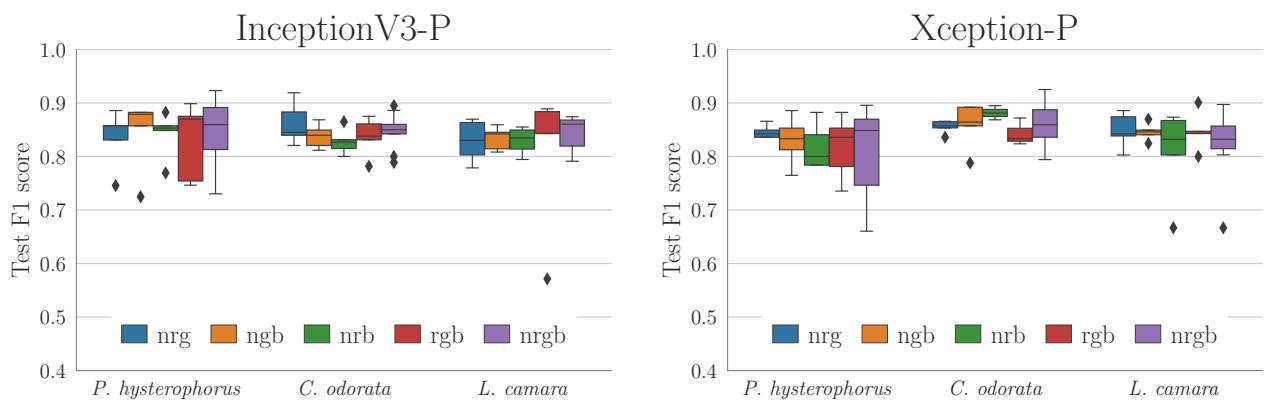


Figure S6: The performance of InceptionV3-P and Xception-P for different band subsets.

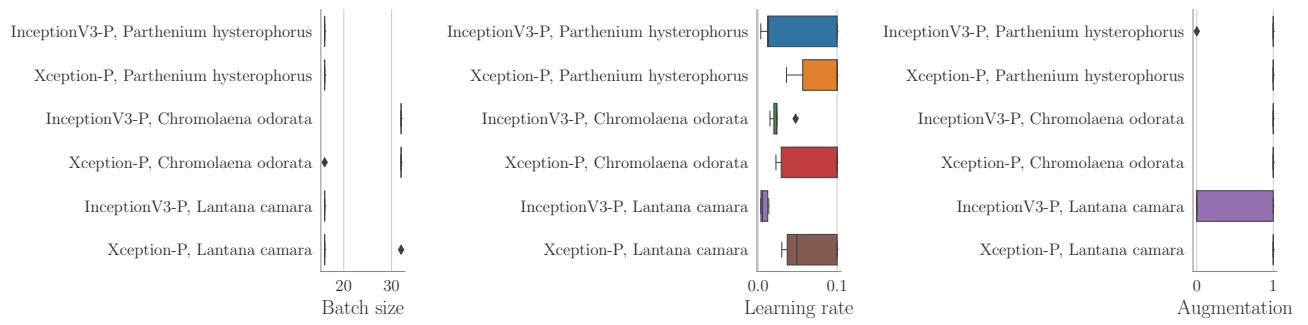


Figure S7: Hyperparameter tuning: The hyperparameters for the best models obtained through Bayesian optimization using Gaussian processes.