


An open-source image classifier for characterizing recreational activities across landscapes

Samantha G. Winder¹  | Heera Lee²  | Bumsuk Seo²  | Emilia H. Lia¹  |
Spencer A. Wood^{1,3} 

¹Outdoor R&D, University of Washington, Seattle, Washington, USA

²Institute of Meteorology and Climate Research, Atmospheric Environmental Research, Karlsruhe Institute of Technology, Garmisch-Partenkirchen, Germany

³eScience Institute, University of Washington, Seattle, Washington, USA

Correspondence

Samantha G. Winder
Email: sgwinder@uw.edu

Funding information

Alfred P. Sloan Foundation, Grant/Award Number: 3835; Gordon and Betty Moore Foundation, Grant/Award Number: 2013-10-29; Helmholtz Association; Karlsruhe House of Young Scientists of Karlsruhe Institute of Technology

Handling Editor: Richard Ladle

Abstract

1. Environmental management increasingly relies on information about ecosystem services for decision-making. Compared with regulating and provisioning services, cultural ecosystem services (CES) are particularly challenging to characterize and measure at management-relevant spatial scales, which has hindered their consideration in practice.
2. Social media are one source of spatially explicit data on where environments support various types of CES, including physical activity. As tools for automating social media content analysis with artificial intelligence (AI) become more commonplace, studies are promoting the potential for AI and social media to provide new insights into CES. Few studies, however, have evaluated what biases are inherent to this approach and whether it is truly reproducible.
3. This study introduces and applies a novel and open-source convolutional neural network model that uses computer vision to recognize recreational activities in the content of photographs shared as social media. We train a model to recognize 12 common recreational activities to map one aspect of recreation in a national forest in Washington, USA, based on images uploaded to Flickr.
4. The image classifier performs well, overall, but varies by activity type. The model, which is trained with data from one region, performs nearly as well in a novel region of the same national forest, suggesting that it is broadly applicable across similar public lands. By comparing the results from our CNN model with an on-site survey, we find that there are apparent biases in which activities visitors choose to photograph and post to social media.
5. After considering potential issues with underlying data and models, we map activity diversity and find that natural features (such as rivers, lakes and higher elevations) and some built infrastructure (campgrounds, trails, roads) support a greater diversity of activities in this region.
6. We make our model and training weights available in open-source software, to facilitate reproducibility and further model development by researchers who seek to understand recreational values at management-relevant scales—and

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *People and Nature* published by John Wiley & Sons Ltd on behalf of British Ecological Society.

more broadly provide an example of how to build, test and apply AI to understand recreation and other types of CESs.

KEYWORDS

convolutional neural network, cultural ecosystem services, environmental management, image recognition, machine learning, open source, recreational activities, social media

1 | INTRODUCTION

Ecosystem services are a widely accepted framework for incorporating and accounting for human well-being in environmental assessments and management decisions (IPBES, 2019; Mandle et al., 2021; TEEB, 2010). Beyond the material services provided by ecosystems, the framework emphasizes consideration of the non-material and cultural ecosystem services (CES) that benefit people through increased physical, mental, social and spiritual health (Fish et al., 2016; IPBES, 2019; Russell et al., 2013). However, CES are rarely considered in practice, especially compared with regulating and provisioning services such as water supply and carbon sequestration (Alix-Garcia & Wolff, 2014; Goldman-Benner et al., 2012; Gould et al., 2019). The potential of CES research has not been fully realized by managers, at least in part due to difficulties characterizing and measuring CES, as well as mismatches in the scale of CES studies and management decisions (Baumeister et al., 2020; Gould et al., 2019; Guerry et al., 2015; Hernández-Morcillo et al., 2013; La Rosa et al., 2016; Satz et al., 2013).

Recent research has proposed that challenges to conceptualizing and measuring CES can be overcome by analysing information shared on social media (Calcagni et al., 2019; Gliozzo et al., 2016; Wood et al., 2013). This idea is premised on the observation that people use social media platforms such as Twitter and Instagram to share descriptions and depictions of their cultural experiences and interactions with the surrounding environment (Di Minin et al., 2015; Ghermandi & Sinclair, 2019; Langemeyer & Calcagni, 2022). Additionally, since social media are often georeferenced—meaning they are tied to the location where the content is created—they provide a spatial record of human–ecosystem interactions. As Havinga et al. (2020) demonstrate, geotagged social media are useful for measuring many types of CES including physical activity, aesthetic appreciation, ecological meaning, the development of knowledge and spiritual importance. Indeed, studies leveraging publicly available social media have begun investigating where environments provide aesthetically pleasing landscapes (Egarter Vigl et al., 2021; Figueroa-Alfaro & Tang, 2017; Ghermandi et al., 2020), enjoyment of plants and animals (Richards & Friess, 2015), and opportunities to participate in religious (Roberts, 2017), spiritual (Oteros-Rozas et al., 2018) or recreational activities (Väisänen et al., 2021).

Most prior CES research using social media has relied on total counts of users who share content to map recreational visitation (Tenkanen et al., 2017; Wilkins et al., 2021; Wood et al., 2020) and understand how visitation varies across environments or conditions (Fisher et al., 2018; Kim et al., 2019; Levin et al., 2017). More recently,

studies have begun utilizing not only the location and volume of social media, but also the content of the posts (e.g. Clemente et al., 2019; Egarter Vigl et al., 2021; Gosal et al., 2019). In some studies, the content contained in images is classified by manually viewing and labelling images (Clemente et al., 2019; Langemeyer et al., 2018; Oteros-Rozas et al., 2018; Pickering et al., 2020; Richards & Friess, 2015; Thiagarajah et al., 2015). This method is inefficient, however, and it is difficult to scale (an average of 140 photos per researcher per hour (Richards & Friess, 2015)), so recent studies have explored automated methods using artificial intelligence (AI) provided by commercial tools such as Google Cloud Vision (Gosal et al., 2019; Richards & Tunçer, 2018) and Clarifai (Egarter Vigl et al., 2021; Lee et al., 2019). Gosal et al. (2019), for example, used Google Cloud Vision combined with a latent semantic analysis in order to map several recreational values and identify where recreation may threaten particular species within protected areas. Lee et al. (2019) used Clarifai to annotate images, then performed a network analysis to derive themes of the photos, several of which were related to cultural services. These studies were an innovative step towards using publicly shared photographs to learn about recreational, aesthetic and other cultural benefits which people derive from nature, and they illustrate the potential for AI to be applied to a wide variety of questions related to these topics.

Despite the demonstrated potential for AI such as computer vision to be used to measure aspects of CES from social media, few studies have evaluated what biases are inherent to this approach and whether it is truly reproducible. There are several potential issues. First, the AI underlying commonly used tools may not be generating consistent predictions about the content depicted in images. Yet, this is difficult to evaluate since nearly every study to date has relied on proprietary tools that are implemented in commercial software (but see Väisänen et al., 2021). These commercial tools are generally not well documented and are often modified (or decommissioned) without notice nor consultation with users. This lack of transparency makes it difficult to reproduce analyses and to apply standard methods for evaluating model accuracy and performance since the outputs of a model may differ in unknown ways over time (Lazer et al., 2014). In contrast, with open-source tools and algorithms, it is feasible to test and evaluate model performance in different situations and understand how outcomes are related to model structure and parameters. Second, even when computer vision models perform well, the content of the images that are shared as social media may not accurately represent the ways that people interact with the environment. There are likely biases in both who chooses to share images to particular platforms and in what type of content is represented in those images (Mashhadi et al., 2021; Ruths &

Pfeffer, 2014). Without comparing results from AI models to results from other methods of elucidating CES values such as participatory geographic information systems (PPGIS) (Muñoz et al., 2020; Tolvanen et al., 2020) or surveys (Heikinheimo et al., 2017; Song et al., 2020), it is impossible to say whether applying AI to social media is generating consistent, complementary or even directly contradictory information about CES.

This study explores the potential for social media and computer vision to map where ecosystems create opportunities for recreational activity, as one example of a CES. We introduce a convolutional neural network (CNN) that uses AI to recognize outdoor activities in the content of photographs posted to social media. We describe the creation of a training dataset and our development and evaluation of the model using 13 years of photographs from Flickr. To test how well the model performs, we compare model predictions to labels assigned manually by researchers, and discuss the performance overall, and by activity class, both in the original study region and in a novel region of the same forest. Then we demonstrate that there are biases in the frequency with which various activities are shared on social media, by comparing the predicted activities to activities reported by respondents to an on-site survey administered on the same public lands in Washington, USA. Finally, we present an example application of the methods by creating maps of recreational activities at two locations and relating the diversity of activities present in different parts of the region to underlying landscape characteristics. To facilitate future model development and to ensure reproducibility, we share an entirely open-source software package called *recCNNize*¹ as well as the fitted weights of our CNN model so that others can replicate our approach for identifying recreational activities in images.

2 | METHODS

In this study, we develop, test and apply a classifier to recognize recreational activities in images shared on social media from two locations on public land in Washington, USA.

2.1 | Study site

This study focuses on public lands in the Mount Baker-Snoqualmie National Forest in western Washington, USA (Figure 1). We developed our classifier using photographs from the Middle Fork Snoqualmie River Valley (Middle Fork), which is a popular recreation destination located approximately 50 km east of the Seattle metropolitan area. The Valley contains lush coniferous forest in the foothills of the Cascade Range, ranging from 150 to 1850 m above sea level. Due to its low elevation and temperate maritime climate, the valley floor and access roads remain snow-free year-round. The Middle Fork Snoqualmie River is a designated Wild and Scenic River, and approximately half of the study region is in the Alpine Lakes Wilderness. The area provides a variety of recreational opportunities, ranging from trails for hiking, biking and horseback riding to campgrounds and day-use picnic sites. Vehicle access to this valley is via a single road, making it a relatively self-contained recreation destination. We tested our classifier in the Mountain Loop Highway corridor (Mountain Loop) region of the same National Forest, located approximately 75 km northeast of Seattle, WA. This area is another common destination for visitors, though the amount and character of recreational use differs between the two regions. The Mountain Loop region has steeper

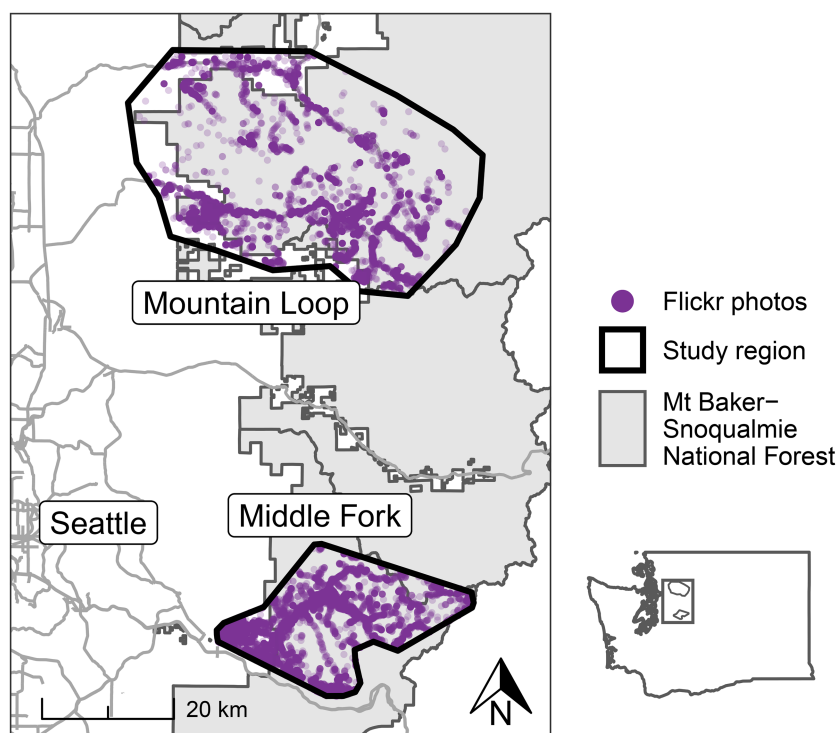


FIGURE 1 Map of the Middle Fork and Mountain Loop study regions in western Washington, USA, showing the location of all geotagged Flickr photographs posted within the boundaries between 2005 and 2018.

terrain and it includes both developed and informal campsites. As a result, it tends to draw individuals interested in physical challenge as well as a larger number of overnight visitors. Portions of the region are snow covered and inaccessible during the winter months.

We created spatially explicit boundaries around each region based on access. The Middle Fork boundary encompassed just over 450 km² around recreation sites primarily accessed via the Middle Fork Road. The Mountain Loop boundary included almost 1450 km², with recreation sites primarily accessed from the Mountain Loop Highway or from State Route 530.

2.2 | Data description

2.2.1 | Flickr data

We collected all photographs posted to Flickr within the Middle Fork and Mountain Loop regions between 2005 and 2018 by querying the Flickr application programming interface (API) in December, 2018–January, 2019. This resulted in 14,839 photographs from the Middle Fork region (from 688 unique users, max. 5940 photos from a single user) and 18,350 photos in the Mountain Loop region (from 863 unique users, max. 681 photos from a single user).

2.2.2 | On-site surveys

Researchers surveyed visitors leaving the Middle Fork Valley on 23 days between August 2 and October 10, 2018. We selected survey days throughout the survey period to include an even number of weekday (Monday–Thursday) and weekend (Friday–Sunday) days. After intercepting a group of visitors, researchers explained that they were studying outdoor recreation in the Middle Fork Valley, that participation was entirely voluntary, and that responses would be anonymous. They then verbally asked whether the visitors would participate in the study. Upon receiving verbal consent, the researchers asked a random member of the party (the adult with the next upcoming birthday) to complete a written survey in English. The intercept survey was part of a larger project to describe outdoor recreational use in the region, and among other questions it asked visitors to select the activities they participated in during their visit to the region from a predefined list (Table 1). In total, 595 visitors representing 595 parties completed a survey, and 580 of the respondents reported what recreational activities they participated in during their visit. The University of Washington Institutional Review Board reviewed our study and determined that the research met regulations for the protection of human subjects and the study carried no greater than minimal risk to those subjects. Accordingly, they granted the study exempt status (IRB ID: STUDY00005339).

2.3 | Image classification

2.3.1 | Recreational activity classes

We chose 12 of the most common recreational activities in the Middle Fork for image classification (Table 1). Additionally, we created two ‘no activity’ classes to aid in image classification—one for photographs without any people present, and the other for photographs where people were present but not obviously engaging in any particular activity.

2.3.2 | CNN classifier

We developed a CNN to measure the probability that Flickr images contained evidence of the 14 recreational activity classes (Table 1). We regarded the identification of activities in images as a multinomial classification problem. In other words, given a photograph X_i from location i , we evaluated whether the photograph reflected recreational activity a . We customized and fine-tuned Google's Inception (InceptionResNetV2) model (Szegedy et al., 2017) to recognize patterns corresponding with our focal recreational activities. We chose to use this particular model because it is widely recognized as one of the best performing CNN models, compared to competing models such as Resnet and GoogLeNet (Bianco et al., 2018; Canziani et al., 2017). While it is slower to train, we expected the deepness and complexity of this model to be more likely to capture subtle differences between activity classes. Furthermore, the training speed was not a critical concern since our study used a relatively small number of training images as compared to other work using CNNs.

The final model contained 164 layers, the first 160 of which were taken directly from the original Inception model. Because these early layers recognize lower level features, such as edges and basic shapes, we were able to take advantage of them by re-using the trained weights provided with the original CNN model (Szegedy et al., 2017), in a process known as transfer learning.

We customized the model by removing the final dense layer from the Inception model (which is trained to recognize 1000 classes in ImageNet photos), and replacing it with four custom layers which we trained using the training data described in Section 2.3.3. Specifically, we added a 2D global average pooling layer, a new dense layer with 1024 classes, a 30% dropout layer (i.e. 30% of the features are set to zero randomly) and a dense layer with softmax activation that assigns class labels to images ($n = 14$) (Chollet et al., 2018) (Figure S1). The average pooling and the dropout layers are included to avoid overfitting. Note that the dropout layer was ignored when testing at the end of each training epoch, so all features were used in predicting validation images. A consequence of using a dropout layer is that the model can yield higher validation accuracy than training accuracy. The final softmax layer calculates the probability of an image belonging to each of the possible classes. In the analysis, we

TABLE 1 The 12 activity and two no-activity classes that we trained our image classifier to recognize. The description is our criteria for deciding whether or not to assign a class during manual evaluation, and the survey categories are the closest matching activities from the predefined list on the survey instrument

Activities	Description	Survey categories
Backpacking	Person with large backpack or hiking poles, tents obviously in the backcountry	Overnight backpacking
Bird watching	Close-up images of birds, person with binoculars	Bird or wildlife viewing
Boating	Any type of boat	Boating
Camping	Person or tent in an established campground, or with vehicle visible	Camping in campground, camping along the road
Fishing	Person near a river with fishing gear, or holding a fish	Fishing
Hiking	Person with day backpack or hiking poles, in the forest, by a river or trail	Hiking or walking
Horseback riding	Person with horse trailer, riding a horse	Horseback riding
Mountain biking	Person with a bike, a bike on a trail or road	Bicycling on trails or roads
Rock climbing	Person on a rock face, rock climbing gear	Rock climbing
Swimming	Person in a bathing suit in or near a river, lake or hot spring	Swimming
Trail running	Person in running clothes, person with hydration pack	Running or jogging
Other activities	Person engaging in any other activity, including working, motorcycling, skiing, etc.	Other activities, hunting, collecting, foraging, picnicking
No activity	Images without people, such as landscapes	NA
People no activity	No obvious object to fit into an activity but a person present in the image	NA

regarded this probability output as the model's confidence on target images, and call the class with the highest probability the top-1 class label. We used the top-1 prediction in later analyses.

2.3.3 | Classifier training

We randomly sampled 6459 (44%) of the Flickr images downloaded from the Middle Fork and manually assigned each to one of the recreational activity classes described in Table 1. There was an embedded data imbalance that led to a scarcity of some activity classes in the random photographs (e.g. only one image depicting trail running and three images depicting fishing). Following an initial round of training, we added training data for some poorly performing activity classes by downloading public tagged photographs from ImageNet² and Pixabay³ (Table S1; Figure S2). These additional images were used primarily to increase the number of images included in the minor classes, as well as adding photographs for classes which were frequently confused with one another (such as mountain biking and the 'other activities' class which included motorcycling). These supplementary images were mainly in situ images relevant to the activities, but for some classes without sufficient in situ images we also included images depicting elements or objects such as backpacks that we believed to be associated with a specific activity. The full training dataset was 11,912 photographs (6459 Flickr images from the Middle Fork and 5453 images from the ImageNet and Pixabay databases; Table S1).

We used the 11,912 labelled images to train the final four custom layers in our CNN model. In other words, only the four layers were trained against the labelled Flickr photos while the early layers were unchanged or 'frozen' during training (Chollet et al., 2018). We sought to minimize the categorical cross-entropy loss function, a common function used to estimate how well predicted class probabilities match the target classes, using the Adam optimization algorithm (learning rate = $1e-5$) (Chollet, 2015). We allowed the algorithm to conduct up to 300 complete cycles through the training dataset (i.e. epochs) as it searched for the optimal solution and training weights, but instructed it to stop training if the training accuracy did not improve in 20 consecutive epochs.

We trained the model on 60% (7140) of the labelled images and withheld 40% (4772) of the images for validation during the training process. To compensate for the small number of images in some classes, we augmented our training data by applying random geometric transformations in each epoch (such as flipping, resizing, brightening and rotating, Table S2). We chose a batch size of 512, meaning that 512 of the labelled images were used to train at a time, causing the algorithm to run through 14 folds (7140/512) in each epoch. In each fold, the predictions were iteratively evaluated against the true class labels using categorical accuracy (i.e. training accuracy), and then the model weights were updated. Thus, the model was updated 14 times in each epoch. After each epoch, the algorithm evaluated model performance on the 4772 validation images which were not used for training (i.e. validation accuracy) and the optimizer routine decided whether to continue or stop training.

The model which achieved the best training accuracy was selected and used to predict the class of new photos.

We carried out the training on a workstation with an eight-core CPU (AMD Ryzen 2700x) and 64 GB memory. The classification and prediction Python scripts were based on Tensorflow and Keras APIs (Chollet, 2015; Martín Abadi et al., 2015) running on a single NVIDIA Titan Xp with 12 GB of video memory. The optimizer stopped training after 129 epochs, after training accuracy did not improve in 20 consecutive epochs. The overall training time was 12.5 h and the best validation accuracy achieved was 0.88 and the loss was 0.41 (Figure S3). The validation accuracy and loss were better than the training accuracy and loss in this study, which is likely due to the dropout layer in the CNN model.

2.3.4 | Classifier evaluation

To evaluate model performance beyond the training data, we carried out two external validations. First, we randomly selected 742 of the Middle Fork Flickr photographs that were not used in the training ($n = 8380$) and manually evaluated whether they belonged in the class with the greatest probability returned by the CNN model (the top-1 prediction). These test photographs were stratified to represent 20% of the images in each top-1 prediction class, and represented 6.2% of all Flickr photos in the Middle Fork. Additionally, we tested the model in the Mountain Loop region. We predicted the activity class of all Flickr photographs acquired in this region ($n = 18,350$), using the model trained on the Middle Fork images. We randomly sampled the predicted photographs from this site using the same sampling scheme, and manually evaluated the activity classes ($n = 491$; 2.7% of the total images).

The manual evaluation consisted of two examiners independently assigning the sampled images to their 'true' class and comparing these results to the model predictions. We evaluated these predictions by creating a confusion matrix and calculating several standard metrics (precision, recall and F_1 score) for each class, as well as summarizing overall model performance using accuracy, macro F_1 , and Cohen's unweighted Kappa (Kuhn et al., 2008).

Each of the class-level metrics that we used were calculated from a combination of the number of true positives (TP , images which belonged to an activity class and were correctly classified as belonging to that class), false positives (FP , images which did not belong to an activity class but were incorrectly assigned to that class), false negatives (FN , images which belonged to an activity class but were incorrectly assigned to a different class) and true negatives (TN , images which did not belong to an activity class and were correctly not assigned to that class). *Precision* is the positive predicted rate ($\frac{TP}{TP+FP}$), or how many of the samples predicted as positive are actually positive. It can be interpreted, for example, as 'How many of the images labelled hiking are actually images of hiking?'. *Recall* is the true-positive rate ($\frac{TP}{TP+FN}$), also known as sensitivity, and represents how many of the positive samples are captured by the positive predictions. It answers the question 'How many of the hiking images are actually captured in the hiking class predictions?' The F_1 score is

the harmonic mean of precision and recall, creating a single metric of class performance. The harmonic mean gives greater weight to lower values, so classes will only have a high F_1 score if both precision and recall are high. We chose not to focus on *specificity*, the rate at which a model correctly detects that an image does not belong to a particular class ($\frac{TN}{TN+FP}$, true-negative rate), because metrics which use TN tend to be less insightful with unbalanced datasets. This is because minor classes will tend to have very large numbers of true negatives, since most of the images belong to other classes.

We summarized overall model performance by calculating accuracy, macro F_1 score and Cohen's unweighted Kappa. Accuracy is the proportion of all predictions that are correct. The macro F_1 score is the average of each individual class' F_1 score, giving equal weight to every class. Finally, to confirm whether the model performs better than an uninformed guess, we measured the model's performance using Cohen's unweighted Kappa and carried out a one-sided test of the overall accuracy, which compared the overall accuracy to the rate of the largest class (Kuhn et al., 2008).

2.4 | Survey comparison

To examine how well images of activities shared on Flickr, as classified by our model, represent actual rates of activity participation in the region, we compared our model predictions to the empirical survey data. For this comparison, we subset our Flickr photographs to only include those which were geotagged within recreation areas which we believed to be accessed primarily by the Middle Fork Road, excluding some locations which could be more easily accessed from a nearby highway. This resulted in 8396 photos. To control for users who may have posted multiple photographs of a single activity, we calculated activity photo-user-days (APUD) (following Wood et al., 2013) as the number of unique users posting images that were classified (using the top-1 prediction) as being in each activity, each day. APUD is not directly analogous to PUD, since a single user's photographs may represent multiple activities on a single day. In total, we found that the 8396 photos reflected 2321 APUD, 1076 of which represented recreational activities (1245 were classified into one of the no activity classes). We chose to include all photographs ever posted to Flickr in the region because there were too few photographs posted during the months of the survey for a meaningful comparison (only 15 APUD representing recreational activities in August–October 2018). We compared total APUD per activity to the number of survey respondents who reported participating in each activity and measured correspondence between the two datasets by calculating Pearson's correlation.

2.5 | Case study

Finally, to demonstrate one potential use of this model, we used all of our classified images from both regions to create maps of the frequency of recreational activities shared on Flickr. Due to the relatively small number of photographs representing some of the

activities, we focused on the diversity of activities found across the landscape. We did not include fishing, trail running or horseback riding in our diversity calculation because these classes were poorly represented in our data. We divided each study area into 2 km grid cells, and for each grid cell we calculated the total APUD for each activity, based on the top-1 prediction from the CNN. We then calculated and mapped the number of distinct activities represented in each grid cell. Because we removed the three data-poor classes above and the two no activity classes, the maximum number of activities possible in a grid cell was nine.

We related the number of distinct activities in a grid cell to underlying landscape features to learn about the features of the landscape that drive a greater diversity of activities in each region. For this step, we pooled the two regions and used negative binomial regression, where the number of activities was modelled as a function of whether or not the grid cell contained a campsite, lake, picnic area, river or trail, as well as the minimum distance to the nearest major road, the elevation of the centre of the grid cell and the proportion of the grid cell which is designated wilderness. All variables were scaled to fall between 0 and 1 so that coefficients could be compared directly. To avoid issues with multicollinearity, we checked for strong correlations between the variables (all were $< |0.4|$). We compiled landscape data from

the US Forest Service⁴ (campsites, picnic areas, trails and wilderness boundaries), Washington State Department of Transportation⁵ (roads), Washington State Geospatial Open Data Portal⁶ (rivers and lakes) and USGS Elevation Point Query Service⁷ (elevation), accessed using the `EL-EVATR` package in R (Hollister, 2018). We chose to use a negative binomial GLM because the number of activities occurring in a cell is a discrete, rather than continuous variable, and our data were overdispersed so a Poisson model was not appropriate. We measured the predictive power of our model using a pseudo- R^2 metric (Zuur et al., 2009).

3 | RESULTS

3.1 | CNN classifier

3.1.1 | Classifier evaluation

The classifier performed well in the Middle Fork region, achieving overall accuracy of 0.71 (95% CI: 0.67–0.74; macro $F_1 = 0.61$; Cohen's unweighted Kappa = 0.59) on the randomly sampled test images (Figure 2, Table 2). It classified images into recreational activities better than would have been expected by chance (one-sided

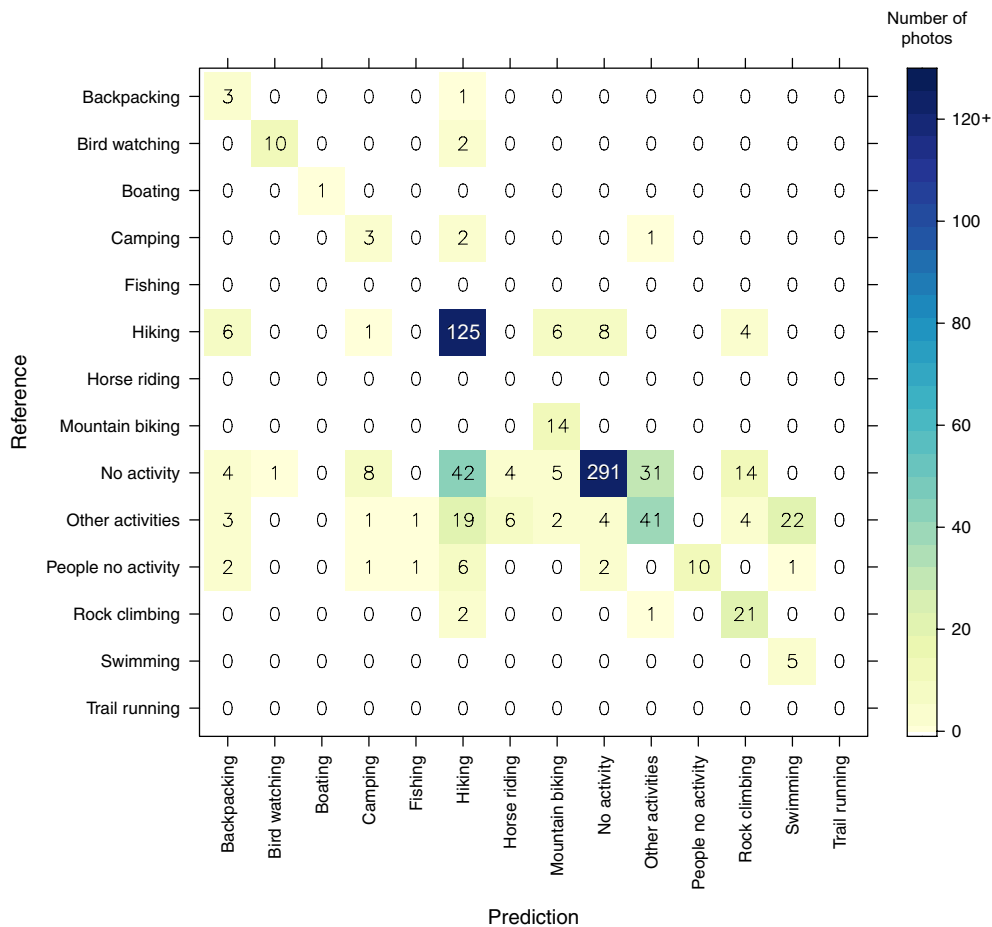


FIGURE 2 Confusion matrix for top-1 predictions in the Middle Fork region using photographs randomly sampled with stratification ($n = 742$)

TABLE 2 Model performance for top-1 predictions in the Middle Fork region using photographs randomly sampled with stratification ($n = 742$). The Support column gives the number of test photographs which truly belong to each activity, not the number of photographs which were selected for manual evaluation

	F_1 score	Precision	Recall	Support (n)
Backpacking	0.27	0.17	0.75	4
Bird watching	0.87	0.91	0.83	12
Boating	1.00	1.00	1.00	1
Camping	0.30	0.21	0.50	6
Fishing	–	0.00	–	0
Hiking	0.72	0.63	0.83	150
Horse riding	–	0.00	–	0
Mountain biking	0.68	0.52	1.00	14
No activity	0.83	0.95	0.73	400
Other activities	0.46	0.55	0.40	103
People no activity	0.61	1.00	0.43	23
Rock climbing	0.63	0.49	0.88	24
Swimming	0.30	0.18	1.00	5
Trail running	–	–	–	0

test of the overall accuracy $p < 0.0001$), with large variability in the model's performance across activity classes (Table 2, Figure 3). The model performed well on images from the bird watching, hiking and no activity classes, with relatively high F_1 scores driven by high recall (indicating that the model was able to successfully identify most images which truly represented these classes) and high precision (indicating that most of the photographs classified in these activities were correct). Note that the perfect F_1 score for boating is due to an extremely small test sample (support = 1), so further testing is necessary to accurately judge the performance of this minor class. The swimming and backpacking classes both had high recall but low precision, indicating that the model was able to successfully identify photographs that included these activities, but that it also incorrectly assigned many other photographs to these classes. Because of the low precision of these classes, their F_1 scores were correspondingly low. The other activities class had relatively low recall and precision, likely due to the broadness of the category. Some of the minor classes (fishing, horseback riding and trail running) did not have F_1 scores; fishing and horseback riding were never classified correctly in our test set (precision = 0, despite evaluating two photographs which our model classified as fishing, and 10 images classified as horseback riding), while trail running was not selected as the top-1 activity for any of the photographs in the Middle Fork (Table 2; Figure 2). Between activity classes, there are also notable differences in the confidence of the top-1 class assignments (Figure S5). Classes such as bird watching, no activity and camping were frequently assigned with probability > 0.9 , representing the high confidence of the CNN model. On the contrary, backpacking and other activities were assigned a lower probability even when they were selected as the most likely class (top-1).

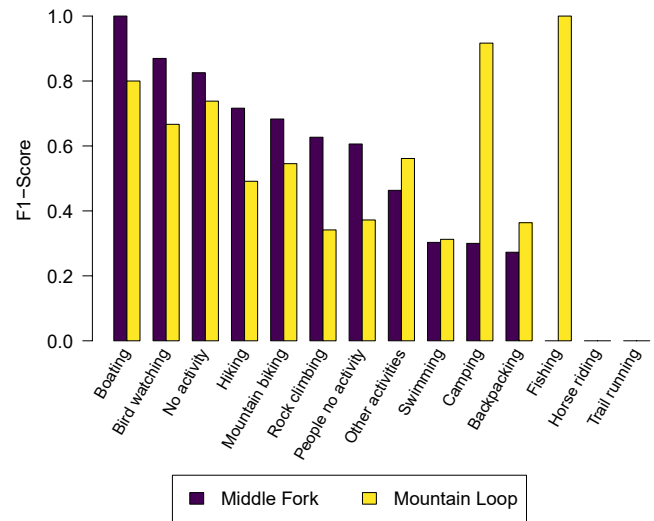


FIGURE 3 Model performance by activity class in the Middle Fork (accuracy = 0.71) and in the Mountain Loop (accuracy = 0.60) regions. Activity classes are ordered by their F_1 score in the Middle Fork region; values are provided in Table 2 and Table S3. Minor classes horseback riding and trail running have no values as none of the test images were correctly classified into these categories. Fishing was evaluated only in the Mountain Loop region ($n = 1$)

3.1.2 | Regional test of classifier

The model performed almost as well in the novel Mountain Loop region as in the Middle Fork, with an overall accuracy of 0.60 (95% CI: 0.56–0.65; macro $F_1 = 0.59$; Kappa = 0.53; $p < 0.0001$; Figure 3, Table S3, Figure S4). Compared to the Middle Fork region, the model performance in the Mountain Loop decreased by 12.7% (overall accuracy), 3.3% (macro F_1 score) and 10.2% (Kappa). While class performance also varied widely in the Mountain Loop, the classes which performed well were not always the ones which performed well in the Middle Fork (Figure 3). In particular, camping performed much better in the Mountain Loop (F_1 score = 0.92) than in the Middle Fork (F_1 score = 0.30). This may be due to a substantially larger number of photos of this activity in the Mountain Loop region as compared to the Middle Fork. The swimming, backpacking and other activities classes each performed slightly better in the Mountain Loop, while all other classes performed somewhat worse in the novel region. Note that fishing only had one test photo in the Mountain Loop, so as with boating in the Middle Fork performance metrics are not indicative of true performance.

3.2 | Survey and model prediction comparison

We did not find a strong correlation between the log number of APUD and the log number of survey respondents who reported participating in each activity across the Middle Fork ($r = 0.32$; 95% CI: -0.31 to 0.75 , Figure 4). There were far more survey respondents who reported trail running or fishing, in particular, than there

were photos classified as belonging to these activities. While we do not have F_1 scores for these classes (due to a lack of successful classifications in our test set), we do know that each of these classes was extremely uncommon in our labelled photos from the Middle Fork (one and three photos, respectively, out of 6459 labelled images; Table S1). When we removed these two classes, as well as horseback riding, which also had no successful classifications in the manual evaluation, we found a much higher correlation between log APUD and log survey respondents ($r = 0.73$; 95% CI: 0.12–0.94; dashed line in Figure 4). Still, the relationship is not perfectly linear, with some activities (such as swimming and bird watching) apparently under-represented in the photos and others (rock climbing, boating) perhaps over-represented in the photos. These patterns are not explained by model performance, with both high-performing and low-performing activities falling on both sides of the line (Figure 4).

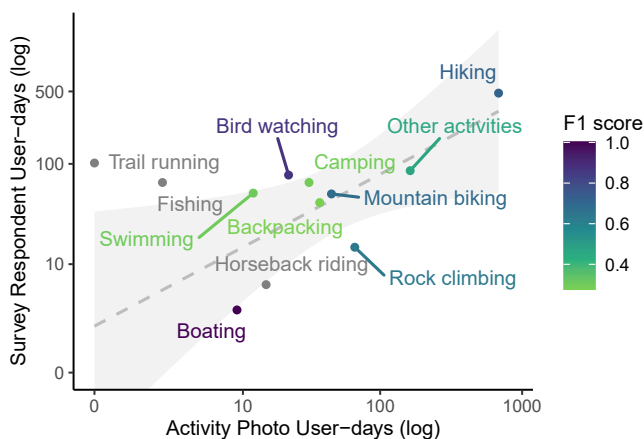


FIGURE 4 The number of participants in each of 12 recreational activities across the entire Middle Fork study region. Colours represent the F_1 score for each class based on the manual evaluation in the region. Classes without an F_1 score are shown in grey. The dashed line shows the relationship between surveys and APUD for activities with an F_1 score, with the light grey ribbon showing the 95% confidence interval

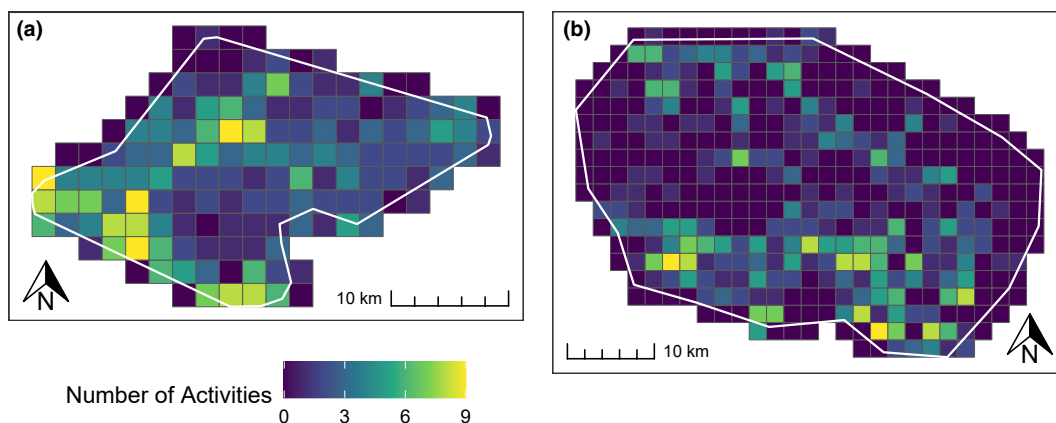


FIGURE 5 Maps showing the number of distinct activities by 2 km grid cell across the Middle Fork (a) and the Mountain Loop (b) regions, according to Flickr images classified by the CNN model. Study region boundaries are shown in white

3.3 | Spatial variability in activities

We found hotspots of recreational activities in both of our study regions (Figure 5). In these areas, we found evidence of visitors participating in up to nine distinct activities. These nine activities include all of the recreational activity classes that our CNN model was able to identify with confidence (not including the minor classes trail running, fishing or horseback riding, nor the two no activity classes). Other portions of the landscape appear to support fewer recreational activities. The number of distinct activities occurring in a grid cell was highly correlated with total APUD ($r = 0.70$), indicating that a greater number of activities occur in more popular areas.

The landscape features which we tested in the negative binomial regression model explained some of the variability in activity diversity across the landscape (pseudo- $R^2 = 0.22$). In particular, a greater number of activities occurred in areas near roads, at higher elevations, and in areas with trails, rivers, lakes or campgrounds. Of these landscape features, proximity to roads had the greatest effect size, with the number of activities present in a grid cell dropping off quickly as distance to the nearest road increased. The presence of either rivers or lakes was significantly positively correlated with the number of activities, though the effect size was smaller than the effect of the infrastructure variables (roads, trails and campgrounds). We found no significant relationship with picnic areas nor the proportion of wilderness in a cell (Figure 6).

4 | DISCUSSION

This study introduces and applies a reproducible approach for studying spatial variability in nature-based recreational activities based on content in social media images. Our CNN model is able to classify images according to 12 distinct recreational activities and two no activity classes. This is, to our knowledge, the first open-source solution to be evaluated for this particular classification problem. While demonstrating that it is possible to identify individual recreational activities in images, we also show that there are apparent biases in

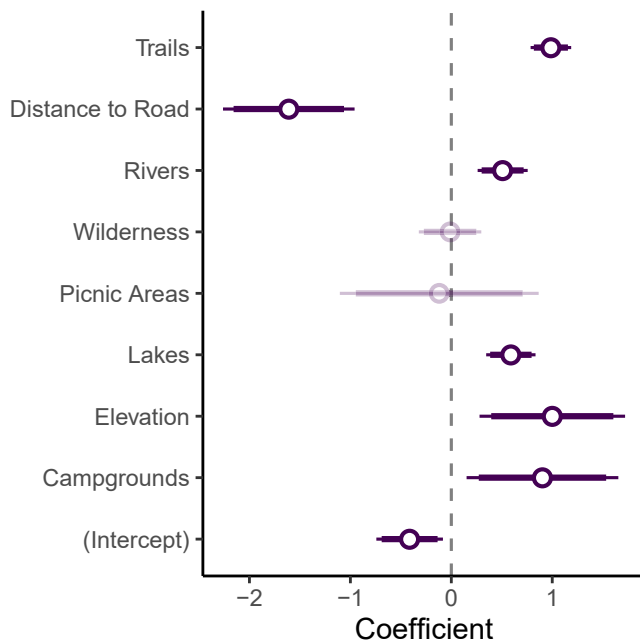


FIGURE 6 Coefficients from the model relating diversity of activities to underlying landscape features in two study regions ($n = 554$, psuedo- $R^2 = 0.22$). Coefficients which are insignificant at $\alpha = 0.05$ are partially transparent

which activities visitors choose to photograph and post to social media. Two popular recreational activities (fishing and trail running) in our study area in western Washington, for example, are almost entirely absent from Flickr photographs taken within the region. Yet, with careful consideration of underlying biases, there is potential for this approach to identify how recreational activities covary with features of the landscape—thereby explicitly measuring how the CES is supported by several features of the natural and built environments. More broadly, this study serves as an example of how to build, distribute and apply AI to understand novel aspects of interactions between people and ecosystems.

4.1 | Classifier performance

While the accuracy and confidence of our classifier is high overall, it varies substantially across the activity classes in our study, and this finding has broader implications for CES research using computer vision. Our CNN is best at recognizing popular or clearly defined classes such as boating and bird watching that are associated with recognizable objects (Table 2, Figure 3). The model struggles, meanwhile, to distinguish some activity classes that are visually similar to each other, such as hiking and backpacking which involve similar equipment and both occur on trails (Figure S6). There are also important regional differences in model accuracy for some classes such as camping. Despite being trained with images from the Middle Fork region, images of camping are more identifiable in the Mountain Loop. These regional and inter-class difference in performance and accuracy—which are inherent to all image classifiers, including

commercial and proprietary ones—highlight the importance of measuring, reporting and addressing the potential for uncertainty in AI to generate biased results.

Our comparison between activity participation according to Flickr photographs and the on-site survey supports previous research concluding that visitors share social media about their participation in a multitude of recreational activities (Hartmann, 2019; Heikinheimo et al., 2017; Norman et al., 2019; Väisänen et al., 2021). Looking more closely at the relative frequency of each activity, we find that some activities are more popular among survey respondents than among social media users, and vice versa. Trail running, fishing and swimming are less prevalent in social media images compared to the survey responses, whereas rock climbing is more popular on social media. We suspect that this is largely because visitors are less likely to share photographs of certain activities on social media relative to other activities (as suggested by Ghermandi & Sinclair, 2019; Tenerelli et al., 2016; Wood et al., 2013). The measured differences may also be due to the timing of the survey. Swimming is more common in the summer when the survey occurred, and one survey day coincided with a trail running event, meaning that these activities are likely over-represented in the survey data. However, trail running events are not uncommon, and the fact remains that of the 6459 images we manually labelled, only one was of trail running. Additionally, the disconnect may be due partly to differences in time-scales of the two datasets. Due to the relatively small number of photographs posted to Flickr during the survey period (15 APUD representing recreational activities were posted between August and October 2018), we chose to include all photographs posted over the life span of the platform (2005–2018).

The general correspondence in the CNN model predictions between the two regions indicates that there is potential for the model to be applied with relative confidence to characterize recreational activities across larger landscapes. However, the disconnects between the survey results and the images uploaded to social media also suggest that there are sampling biases that should be taken into account before the current model is generalized widely. If some activities are better represented in social media photographs than other activities, there is potential for activities by some user groups to be underestimated or missed entirely if this method is applied without caution. If differences are due to prediction errors by the CNN model, then the approach would benefit from additional training data in the form of labelled images. While we found no significant relationship between F_1 scores and the number of training images, three classes were so uncommon in the Flickr images that we were unable to calculate F_1 scores for them. Future studies working at sites with higher visitation and thus more photographs, or incorporating more popular social media platforms such as Instagram, will be able to improve our open-source classifier by retraining the CNN model. To support CNN model development—and to advance the use of image recognition in CES research, more broadly—future studies should report common metrics for evaluating performance, identifying data imbalances and guiding future training. In this study, we

include our confusion matrices and report three standard metrics for machine learning: namely F_1 score, precision and recall. These key statistics facilitate open research and development of tools that are transparent, understandable and improvable.

4.2 | Case study

After accounting for potential issues with the underlying social media data and image classifier, we were able to create maps of recreational activity diversity in two distinct regions of the Mount Baker-Snoqualmie National Forest, for the nine activities that our CNN model can identify with confidence. Our observation is that there are 'hotspots' that support a larger diversity of recreational activities than their surroundings (Figure 5). These hotspots also have more visitors over time, evidenced by a correlation between the total number of activities and APUD. This may be due to more popular areas encouraging a greater diversity of activities, or because areas which allow a greater diversity of activities draw more people. Locations that support a large diversity of activities may also support a greater number of other CES, although this work does not directly measure these other types of interactions with the ecosystem.

Our analysis of activity patterns in Mount Baker-Snoqualmie National Forest serves to demonstrate one of many potential applications of fine-grained maps of recreational activities. Specifically, we show how information about the spatial variability in recreation can be leveraged to understand how the CES is related to the ecosystem and elements of the built environment—similar to previous studies which have explored a variety of methods for using social media and PPGIS for this purpose (Bagstad et al., 2016; Fisher et al., 2019; Kliskey, 2000; Levin et al., 2017; Sherrouse et al., 2011). In the Mount Baker-Snoqualmie National Forest, hotspots of recreational activity occur most often in areas with natural features (rivers, lakes, higher elevations) supported by built infrastructure such as campgrounds and access via trails and roads. Of the landscape features which we considered, proximity to a road had the greatest impact on the diversity of recreational activities. While we chose to consider the diversity of recreational activities, due largely to the relatively low number of images depicting many of our more minor classes, future work could apply a similar model and framework to model individual activities. Additionally, the 'no activity' class could be further divided into various other CES types such as aesthetic and heritage value and analysed similarly (Havinga et al., 2020).

Specific knowledge of activity locations and how they depend on the underlying ecosystems can support sustainable recreation planning. Activity maps can help managers identify areas with high potential for environmental degradation and conflict between user groups that need mitigation or supporting infrastructure. Information on the relative importance of natural and built environments for recreational use allows managers to target spending on ecological restoration projects and the maintenance or construction of infrastructure improvements. In the Mount Baker-Snoqualmie National Forest, our analyses indicate that by improving access to

areas which are currently undeveloped—particularly areas with natural features such as lakes or rivers—managers in the region could increase the likelihood of attracting visitors who will interact with sites in a wider variety of ways. Furthermore, knowing which types of environmental features support a particular recreation activity could help practitioners argue for the value of those landscapes. This explicit valuation of public lands as 'spatial assets' (Jepson et al., 2017) gives managers and conservationists another tool with which to justify continued protection of these areas.

4.3 | Limitations and precautions

This study highlights several important precautions when using social media for CES research. As our results show, not all recreational activities are represented proportionally in images posted to Flickr. Beyond the conclusion that certain types of activities are better suited to photographs than others, Flickr users are unlikely to be a representative sample of visitors to the study region (Ruths & Pfeffer, 2014). In the Middle Fork in particular, more than one-third of the photographs posted to Flickr were uploaded by a single user. While we mitigated this impact by calculating APUD rather than working directly with total numbers of photographs, this user's choice about what content to upload clearly impacted our results. This is an extreme case, but it is important to remember that images from any location may be biased towards particular types of activities that are popular with the dominant Flickr user-group. Future studies would be wise to temper this by working with images from multiple social media platforms (Wood et al., 2020).

Studies that store and analyse images containing people should follow practices that protect the privacy of individuals who appear in the images (Di Minin et al., 2021). Automated analyses may provide a slight advantage in this regard, as fewer of the photographs are actually observed by human researchers (Väisänen et al., 2021). However, creating training data and validating results still requires researchers to directly observe some photographs containing individuals who have not explicitly opted in to this research, or necessarily consented to having their likeness shared publicly on social media in the first place. For this study, we stored Flickr images on a restricted disk and then limited the number of researchers who directly viewed the images. Then, we chose to make our training weights, but not the training images themselves, public. This approach poses little risk to privacy. Future CES research using computer vision would be best served by a public repository of images and CNN models to facilitate model building with common training data and benchmarks. Furthermore, such a repository could address ethical and privacy concerns if images were crowd-sourced from individuals who voluntarily submitted content with explicit permissions and restrictions on the uses of those images (Mashhadi et al., 2021). Individuals could optionally self-report their recreational activities, along with other information about their experience and interactions with the environment. This public repository would be invaluable for researchers applying computer vision to questions about CES.

4.4 | Technological advances

By developing an open-source and reproducible approach, this study overcomes several limitations of previous studies that used proprietary tools such as Google Vision and Clarifai (e.g. Egarter Vigl et al., 2021; Lee et al., 2019; Richards & Tunçer, 2018). Among the many issues with proprietary tools are that researchers do not know how closed-source models are constructed, how they are trained, when they change, how long they will be available, and at what price. In contrast, an open-source model facilitates science that is reproducible, testable, improvable and accessible. The CNN model weights that we have provided could be applied in research on these activities in other regions, though we also recommend that future studies expand and improve the CNN by retraining it with additional images that capture a broader range of situations. Furthermore, a community of practice focused on the goal of testing and training models for recognizing different types of recreational activities could ultimately create more accurate and trusted tools for management compared to proprietary tools. Freely available tools that are properly evaluated, improved and run by the community could additionally build capacity within organizations that would otherwise be unable to afford commercial solutions.

5 | CONCLUSION

Together, computer vision and volunteered geographical information from social media have the potential to help overcome challenges to conceptualizing and measuring CES over large geographies. It is important to recognize that these techniques can suffer from biases created during both the data generation and analysis stages of a study, illustrating the danger of relying on unvalidated data sources and models when drawing conclusions about visitors to public lands. Yet, we conclude that carefully applying AI to user-generated content allows researchers and practitioners to learn about patterns of recreational activities across larger spatial scales than would otherwise be feasible. In our study region, by leveraging this technique, we were able to demonstrate that a greater diversity of activities occur in parts of the landscape which exhibit certain natural and built features. We believe that this study and our open-source image classifier are important steps towards creating reproducible and actionable information about recreation, offering researchers a new tool to characterize one type of CES at scales which are relevant to environmental decision-making.

AUTHORS' CONTRIBUTIONS

S.G.W., H.L., B.S., E.H.L. and S.A.W. co-designed the study; E.H.L. and S.A.W. collected the data; E.H.L., H.L. and B.S. created the training data and conducted the manual evaluation; B.S. built, trained and tested the CNN model; S.G.W. compared the CNN results to the survey data; S.G.W., H.L., B.S., E.H.L. and S.A.W. contributed to the interpretation of results and authorship of the manuscript.

ACKNOWLEDGEMENTS

We thank Mountains to Sound Greenway Trust for assistance with the on-site survey, and Anne Guerry, Lesley Miller and Eric White for comments on the manuscript. This research was supported by a Data Science Environments project award from the Gordon and Betty Moore Foundation (Award #2013-10-29) and the Alfred P. Sloan Foundation (Award #3835) to the University of Washington eScience Institute. This research was also supported by the Recruiting Initiative of the Helmholtz Association, and Karlsruhe House of Young Scientists of Karlsruhe Institute of Technology in the form of a Networking Grant which enabled HL to visit the University of Washington in 2019. BS gratefully acknowledges the computing resources provided by IMK-IFU, KIT and Leibniz Supercomputing Centre (<https://www.lrz.de>; project #pn69tu) and the NVIDIA GPU Grant program for the donation of GPUs used for this research (<https://www.nvidia.com>).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The customized CNN model (called recCNNize) is publicly available at our Github repository <https://github.com/OutdoorRD/recCNNize>. The training weights, survey data and all scripts used in this manuscript are available on OSF at <https://doi.org/10.17605/OSF.IO/SNHWJ> (Winder et al., 2022). To protect the privacy of individuals who uploaded images to Flickr without directly consenting to this research, we did not archive identifiable information about Flickr users or images.

ORCID

Samantha G. Winder  <https://orcid.org/0000-0002-7620-6916>

Heera Lee  <https://orcid.org/0000-0002-3869-0698>

Bumsuk Seo  <https://orcid.org/0000-0002-9424-9784>

Emilia H. Lia  <https://orcid.org/0000-0002-4046-1050>

Spencer A. Wood  <https://orcid.org/0000-0002-5794-2619>

ENDNOTES

¹ <https://github.com/OutdoorRD/recCNNize>

² <http://image-net.org>

³ <http://pixabay.com>

⁴ <https://data.fs.usda.gov/geodata/edw/datasets.php>

⁵ <https://www.wsdot.wa.gov/mapsdata/geodatacatalog/default.htm>

⁶ <https://geo.wa.gov/>

⁷ <http://ned.usgs.gov/epqs/>

REFERENCES

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., ... Zheng, X. (2015). *TensorFlow: Large-scale machine learning on heterogeneous systems*.

- Alix-Garcia, J., & Wolff, H. (2014). Payment for ecosystem services from forests. *Annual Review of Resource Economics*, 2, 361–380.
- Bagstad, K. J., Semmens, D. J., Ancona, Z. H., & Sherrouse, B. C. (2016). Evaluating alternative methods for biophysical and cultural ecosystem services hotspot mapping in natural resource planning. *Landscape Ecology*, 32, 77–97.
- Baumeister, C. F., Gerstenberg, T., Plieninger, T., & Schraml, U. (2020). Exploring cultural ecosystem service hotspots: Linking multiple urban forest features with public participation mapping data. *Urban Forestry & Urban Greening*, 48, 126561.
- Bianco, S., Cadene, R., Celona, L., & Napolitano, P. (2018). Benchmark analysis of representative deep neural network architectures. *IEEE Access*, 6, 64270–64277.
- Calcagni, F., Maia, A., Connolly, J., & Langemeyer, J. (2019). Digital co-construction of relational values: Understanding the role of social media for sustainability. *Sustainability Science*, 14, 1309–1321.
- Canziani, A., Paszke, A., & Culurciello, E. (2017). An analysis of deep neural network models for practical applications. *arXiv:1605.07678 [cs]*.
- Chollet, F. (2015). *Keras*. <https://keras.io>
- Chollet, F. (2018). *Deep learning with python*. Manning.
- Clemente, P., Calvache, M., Antunes, P., Santos, R., Cerdeira, J. O., & Martins, M. J. (2019). Combining social media photographs and species distribution models to map cultural ecosystem services: The case of a Natural Park in Portugal. *Ecological Indicators*, 96, 59–68.
- Di Minin, E., Fink, C., Hausmann, A., Kremer, J., & Kulkarni, R. (2021). How to address data privacy concerns when using social media data in conservation science. *Conservation Biology*, 35, 437–446.
- Di Minin, E., Tenkanen, H., & Toivonen, T. (2015). Prospects and challenges for social media data in conservation science. *Frontiers in Environmental Science*, 3, 1–6.
- Egarter Vigl, L., Marsoner, T., Giombini, V., Pecher, C., Simion, H., Stemle, E., Tasser, E., & Depellegrin, D. (2021). Harnessing artificial intelligence technology and social media data to support cultural ecosystem service assessments. *People and Nature*, 3, 673–685.
- Figuerola-Alfaro, R. W., & Tang, Z. (2017). Evaluating the aesthetic value of cultural ecosystem services by mapping geo-tagged photographs from social media data on Panoramio and Flickr. *Journal of Environmental Planning and Management*, 60, 266–281.
- Fish, R., Church, A., & Winter, M. (2016). Conceptualising cultural ecosystem services: A novel framework for research and critical engagement. *Ecosystem Services*, 21, 208–217.
- Fisher, D. M., Wood, S. A., Roh, Y.-H., & Kim, C.-K. (2019). The geographic spread and preferences of tourists revealed by user-generated information on Jeju Island, South Korea. *Land*, 8, 73.
- Fisher, D. M., Wood, S. A., White, E. M., Blahna, D. J., Lange, S., Weinberg, A., Tomco, M., & Lia, E. (2018). Recreational use in dispersed public lands measured using social media data and on-site counts. *Journal of Environmental Management*, 222, 465–474.
- Ghermandi, A., Camacho-Valdez, V., & Trejo-Espinosa, H. (2020). Social media-based analysis of cultural ecosystem services and heritage tourism in a coastal region of Mexico. *Tourism Management*, 77, 104002.
- Ghermandi, A., & Sinclair, M. (2019). Passive crowdsourcing of social media in environmental research: A systematic map. *Global Environmental Change*, 55, 36–47.
- Gliozzo, G., Pettorelli, N., & Haklay, M. M. (2016). Using crowdsourced imagery to detect cultural ecosystem services: A case study in South Wales, UK. *Ecology and Society*, 21, 6.
- Goldman-Benner, R. L., Benitez, S., Boucher, T., Calvache, A., Daily, G., Kareiva, P., Kroeger, T., & Ramos, A. (2012). Water funds and payments for ecosystem services: Practice learns from theory and theory can learn from practice. *Oryx*, 46, 55–63.
- Gosal, A. S., Geijzendorffer, I. R., Václavík, T., Poulin, B., & Ziv, G. (2019). Using social media, machine learning and natural language processing to map multiple recreational beneficiaries. *Ecosystem Services*, 38, 100958.
- Gould, R. K., Morse, J. W., & Adams, A. B. (2019). Cultural ecosystem services and decision-making: How researchers describe the applications of their work. *People and Nature*, 1, 457–475.
- Guerry, A. D., Polasky, S., Lubchenco, J., Chaplin-Kramer, R., Daily, G. C., Griffin, R., Ruckelshaus, M., Bateman, I. J., Duraiappah, A., Elmqvist, T., Feldman, M. W., Folke, C., Hoekstra, J., Kareiva, P. M., Keeler, B. L., Li, S., McKenzie, E., Ouyang, Z., Reyers, B., ... Vira, B. (2015). Natural capital and ecosystem services informing decisions: From promise to practice. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 7348–7355.
- Hartmann, M. (2019). *Using a model trained on text and image data from social media content to predict and map leisure activities*. MS Swiss Federal Institute of Technology Zurich.
- Havinga, I., Bogaart, P. W., Hein, L., & Tuia, D. (2020). Defining and spatially modelling cultural ecosystem services using crowdsourced data. *Ecosystem Services*, 43, 101091.
- Heikinheimo, V., Minin, E. D., Tenkanen, H., Hausmann, A., Erkkonen, J., & Toivonen, T. (2017). User-generated geographic information for visitor monitoring in a National Park: A comparison of social media data and visitor survey. *ISPRS International Journal of Geo-Information*, 6, 85.
- Hernández-Morcillo, M., Plieninger, T., & Bieling, C. (2013). An empirical review of cultural ecosystem service indicators. *Ecological Indicators*, 29, 434–444.
- Hollister, J. (2018). *Elevatr: Access elevation data from various APIs*.
- IPBES. (2019). *Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the intergovernmental science-policy platform on biodiversity and ecosystem services*. Technical Report, IPBES Secretariat.
- Jepson, P. R., Caldecott, B., Schmitt, S. F., Carvalho, S. H. C., Correia, R. A., Gamarra, N., Bragagnolo, C., Malhado, A. C. M., & Ladle, R. J. (2017). Protected area asset stewardship. *Biological Conservation*, 212, 183–190.
- Kim, Y., Kim, C.-K., Lee, D. K., Lee, H.-W., & Andrada, R. I. T. (2019). Quantifying nature-based tourism in protected areas in developing countries by using social big data. *Tourism Management*, 72, 249–256.
- Kliskey, A. D. (2000). Recreation terrain suitability mapping: A spatially explicit methodology for determining recreation potential for resource use assessment. *Landscape and Urban Planning*, 52, 33–43.
- Kuhn, M., et al. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, 28, 1–26.
- La Rosa, D., Spyra, M., & Inostroza, L. (2016). Indicators of cultural ecosystem services for urban planning: A review. *Ecological Indicators*, 61, 74–89.
- Langemeyer, J., & Calcagni, F. (2022). Virtual spill-over effects: What social media has to do with relational values and global environmental stewardship. *Ecosystem Services*, 53, 101400.
- Langemeyer, J., Calcagni, F., & Baró, F. (2018). Mapping the intangible: Using geolocated social media data to examine landscape aesthetics. *Land Use Policy*, 77, 542–552.
- Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google flu: Traps in big data analysis. *Science*, 343, 1203–1205.
- Lee, H., Seo, B., Koellner, T., & Lautenbach, S. (2019). Mapping cultural ecosystem services 2.0—Potential and shortcomings from unlabeled crowd sourced images. *Ecological Indicators*, 96, 505–515.
- Levin, N., Lechner, A. M., & Brown, G. (2017). An evaluation of crowdsourced information for assessing the visitation and perceived importance of protected areas. *Applied Geography*, 79, 115–126.
- Mandle, L., Shields-Estrada, A., Chaplin-Kramer, R., Mitchell, M. G. E., Bremer, L. L., Gourevitch, J. D., Hawthorne, P., Johnson, J. A., Robinson, B. E., Smith, J. R., Sonter, L. J., Verutes, G. M., Vogl, A. L.,

- Daily, G. C., & Ricketts, T. H. (2021). Increasing decision relevance of ecosystem service science. *Nature Sustainability*, 4, 161–169.
- Mashhadi, A., Winder, S. G., Lia, E. H., & Wood, S. A. (2021). No walk in the park: The viability and fairness of social media analysis for parks and recreational policy making. *Proceedings of the International AAAI Conference on Web and Social Media*, 15, 409–420.
- Muñoz, L., Hausner, V. H., Runge, C., Brown, G., & Daigle, R. (2020). Using crowdsourced spatial data from Flickr vs. PPGIS for understanding nature's contribution to people in southern Norway. *People and Nature*, 2, 437–449.
- Norman, P., Pickering, C. M., & Castley, G. (2019). What can volunteered geographic information tell us about the different ways mountain bikers, runners and walkers use urban reserves? *Landscape and Urban Planning*, 185, 180–190.
- Oteros-Rozas, E., Martín-López, B., Fagerholm, N., Bieling, C., & Plieninger, T. (2018). Using social media photos to explore the relation between cultural ecosystem services and landscape features across five European sites. *Ecological Indicators*, 94, 74–86.
- Pickering, C., Walden-Schreiner, C., Barros, A., & Rossi, S. D. (2020). Using social media images and text to examine how tourists view and value the highest mountain in Australia. *Journal of Outdoor Recreation and Tourism*, 29, 100252.
- Richards, D. R., & Friess, D. A. (2015). A rapid indicator of cultural ecosystem service usage at a fine spatial scale: Content analysis of social media photographs. *Ecological Indicators*, 53, 187–195.
- Richards, D. R., & Tunçer, B. (2018). Using image recognition to automate assessment of cultural ecosystem services from social media photographs. *Ecosystem Services*, 31, 318–325.
- Roberts, H. (2017). Using twitter data in urban green space research: A case study and critical evaluation. *Applied Geography*, 81, 13–20.
- Russell, R., Guerry, A. D., Balvanera, P., Gould, R. K., Basurto, X., Chan, K. M., Klain, S., Levine, J., & Tam, J. (2013). Humans and nature: How knowing and experiencing nature affect well-being. *Annual Review of Environment and Resources*, 38, 473–502.
- Ruths, D., & Pfeffer, J. (2014). Social media for large studies of behavior. *Science*, 346, 1063–1064.
- Satz, D., Gould, R. K., Chan, K. M. A., Guerry, A., Norton, B., Satterfield, T., Halpern, B. S., Levine, J., Woodside, U., Hannahs, N., Basurto, X., & Klain, S. (2013). The challenges of incorporating cultural ecosystem services into environmental assessment. *Ambio*, 42, 675–684.
- Sherrouse, B. C., Clement, J. M., & Semmens, D. J. (2011). A GIS application for assessing, mapping, and quantifying the social values of ecosystem services. *Applied Geography*, 31, 748–760.
- Song, X. P., Richards, D. R., He, P., & Tan, P. Y. (2020). Does geo-located social media reflect the visit frequency of urban parks? A city-wide analysis using the count and content of photographs. *Landscape and Urban Planning*, 203, 103908.
- Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. In *Thirty-first AAAI conference on artificial intelligence*, Vol. 31.
- TEEB. (2010). *The economics of ecosystems and biodiversity*. Technical Report.
- Tenerelli, P., Demšar, U., & Luque, S. (2016). Crowdsourcing indicators for cultural ecosystem services: A geographically weighted approach for mountain landscapes. *Ecological Indicators*, 64, 237–248.
- Tenkanen, H., Minin, E. D., Heikinheimo, V., Hausmann, A., Herbst, M., Kajala, L., & Toivonen, T. (2017). Instagram, Flickr, or twitter: Assessing the usability of social media data for visitor monitoring in protected areas. *Scientific Reports*, 7, 17615.
- Thiagarajah, J., Wong, S. K., Richards, D. R., & Friess, D. A. (2015). Historical and contemporary cultural ecosystem service values in the rapidly urbanizing city state of Singapore. *Ambio*, 44, 666–677.
- Tolvanen, A., Kangas, K., Tarvainen, O., Huhta, E., Jäkäläniemi, A., Kyttä, M., Nikula, A., Nivala, V., Tuulentie, S., & Tyrväinen, L. (2020). The relationship between people's activities and values with the protection level and biodiversity. *Tourism Management*, 81, 104141.
- Väisänen, T., Heikinheimo, V., Hiippala, T., & Toivonen, T. (2021). Exploring human–Nature interactions in national parks with social media photographs and computer vision. *Conservation Biology*, 35, 424–436.
- Wilkins, E. J., Wood, S. A., & Smith, J. W. (2021). Uses and limitations of social media to inform visitor use Management in Parks and Protected Areas: A systematic review. *Environmental Management*, 67, 120–132.
- Winder, S. G., Lee, H., Seo, B., Lia, E. H., & Wood, S. A. (2022). Data for: An open-source image classifier for characterizing recreational activities across landscapes.
- Wood, S. A., Guerry, A. D., Silver, J. M., & Lacayo, M. (2013). Using social media to quantify nature-based tourism and recreation. *Scientific Reports*, 3, 2976.
- Wood, S. A., Winder, S. G., Lia, E. H., White, E. M., Crowley, C. S. L., & Milnor, A. A. (2020). Next-generation visitation models using social media to estimate recreation on public lands. *Scientific Reports*, 10, 15419.
- Zuur, A., Ieno, E. N., Walker, N., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R* (1st ed.). Springer-Verlag New York.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Winder, S. G., Lee, H., Seo, B., Lia, E. H., & Wood, S. A. (2022). An open-source image classifier for characterizing recreational activities across landscapes. *People and Nature*, 00, 1–14. <https://doi.org/10.1002/pan3.10382>