Ecological Indicators 149 (2023) 110165

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Original Articles

Assessing forest recreational potential from social media data and remote sensing technologies data

Federico Lingua[a,*], Nicholas C. Coops[a], Verena C. Griess[b]

[a] Department of Forest Resources Management, Faculty of Forestry, University of British Columbia, Forest Sciences Centre, 2424 Main Mall, Vancouver, BC V6T 1Z4, Canada

[b] Institute of Terrestrial Ecosystems, Department of Environmental System Sciences, ETH Zurich CHN K72.2, Universitätstrasse 16, 8092 Zurich, Switzerland

A R T I C L E I N F O

Keywords:
Forest Recreation
Cultural Ecosystem Services
Flickr
Big Data
Remote Sensing

A B S T R A C T

Conventionally, forest management plans have focused on ensuring a continuous provision of wood. In recent years, political agendas worldwide have recognized the importance of forests' cultural ecosystem services, such as recreation. However, the inclusion of such values in management plans is challenging, and forest managers require novel methodologies and indicators to characterize forest recreation demand and provision. To this end, in this study, we combine remote sensing technologies and crowdsourced social media data to map and value the forest recreational potential of BC's provincial parks system. We trained and deployed convolutional neural networks to automatically classify the content of over 60,000 Flickr images, we then created a random forest model to identify the variables that influence the visitors' choice of recreational activity. These models allowed us to map the most likely recreational activities to occur in BC's provincial parks and perform a spatially explicit assessment of the consumer surplus that these activities generate with a benefit transfer approach. Our findings suggest that the most influential variables in determining the choice of forest recreational activities are topographic, while anthropogenic impacts and forest biometrics variables have less effect. Furthermore, the outcomes of our study support the proposition that the integration of social media and remote sensing technologies allow, in the future, park managers to tailor the management of recreational services to forest visitors' needs.

1. Introduction

Following the industrial revolution humans have become, and still are, the main cause of forest cover loss (Ritchie and Roser, 2021). In response to this phenomenon, sustainable forest management plans have been developed to secure a continuous provision of wood for current and future generations. More recently, the focus is shifting from management plans being solely driven by timber needs, towards ensuring the seamless provision of all the ecosystem services that forests provide (Ritter and Dauksta, 2013). In particular, forests' cultural ecosystem services (CES) are of recognized growing importance worldwide, and in the last decade, an average of over 186,000 ha of the world forests have been allocated for recreation, tourism, and education every year (FAO, 2020).

However, the inclusion of CES in forest management plans has inherent challenges: CES are troublesome to quantify (Bettinger et al., 2016), map (Termansen et al., 2013), and, given their nature of nonmarket goods, value (Paracchini et al., 2014). Traditionally, the study of forest CES has been undertaken with the use of *on-situ* surveys (Pleasant et al., 2014), interviews (Plieninger et al., 2013), or working with focus groups (Norton et al., 2012). Among these traditional approaches, the public participatory GIS (PPGIS) is a framework that has been applied numerous times to map forest cultural ecosystem services. PPGIS uses geospatial technologies to promote the participation of marginalized populations in the decision-making process regarding ecosystem services management (Brown, 2017). In particular, PPGIS aims to provide an alternative to the economic-based valuation of ecosystem services that some authors -e.g. Potschin and Haines-Young (2013)- argue that it provides a more faithful representation of natural capital. These methods are useful for capturing detailed information on CES, however, they also have two major limitations that hinder their application in forest management planning. First, traditional methods of CES characterization are expensive and time-consuming to undertake and second, they can only be applied across small study areas (Richards and Tunçer, 2018).To tackle these limitations, researchers have turned their attention to novel data sources, and most recently to the use of

* Corresponding author.
E-mail addresses: federico.lingua89@gmail.com (F. Lingua), nicholas.coops@ubc.ca (N.C. Coops), verena.griess@usys.ethz.ch (V.C. Griess).

https://doi.org/10.1016/j.ecolind.2023.110165
Received 21 November 2022; Received in revised form 7 February 2023; Accepted 17 March 2023
Available online 22 March 2023

crowdsourced social media data (Ghermandi and Sinclair, 2019).

Crowdsourced social media data are digital data (text, images, video, etc.) that are shared by social media users, and then retrieved by researchers *via* social media application programming interfaces (API). Analyses of the metadata of geotagged social media images have been undertaken across various ecosystems around the world for differing objectives including (*i*) identifying temporal trends and hotspots of recreational activities (Schirpke et al., 2018), (*ii*) estimating the number of visits into natural areas (Tenkanen et al., 2017), and (*iii*) assessing the monetary value of recreational sites (Ghermandi, 2018; Sinclair et al., 2018). These studies have demonstrated how social media image metadata can provide valuable quantitative information for CES management. However, metadata alone cannot shed light on the reasons why people choose to visit an area or the activities that recreationists carry out. To answer these qualitative questions, researchers have adopted two strategies. The first is the integration of crowdsourced images metadata and remote sensing technologies data, and the second is the inspection of social media crowdsourced images content.

Remote sensing technologies provide information about the physical characteristics of the Earth's surface. The integration of crowdsourced social media data with remote sensing data is becoming more commonplace in the study of forest CES. For example, Bernetti et al. (2019) used the metadata of Flickr images to investigate the relationship between forest and topographic variables, obtained *via* various remote sensing technologies, and picture density in Tuscany (Italy). Similarly, Ciesielski et al. (2021) used Flickr data and boosted regression tree models to determine which forest variables influenced picture acquisition. These variables were obtained by combining various sources including landcover (Hościło and Tomaszewska, 2014) and the Shuttle Radar Topography Mission (Rodriguez et al., 2006). Lastly, You et al. (2022) assessed the spatial–temporal dynamics of forest recreation values in the Zhejiang Province (China) integrating remote sensing data and text-mining of social media comments. In all the above studies, remote sensing technologies data were used to obtain topographic, socioeconomic, and forest biometrics data coincident with the location (and in some cases time) the social media images were acquired, hence identifying which of these variables had a positive effect on forest recreation.

To derive the activities occurring on the crowdsourced social media images, various techniques have also evolved. In early studies, analyses of social media images were conducted by visually inspecting, and manually classifying, the images based on the type of CES depicted in them (Pastur et al., 2016; Richards and Friess, 2015). However, this approach is tedious and time-consuming and therefore it cannot be easily applied to large study areas. Deep-learning, and in particular convolutional neural networks (CNN), offers a solution to this problem (Ghermandi et al., 2022). A CNN is a deep-learning neural network used to process structured data arrays, such as digital images. CNNs are currently considered the state of the art for automated image classification (Howard and Gugger, 2020) and are increasingly being used in computer vision. CNNs are made of two essential components: an architecture, and a set of weights. The architecture is the fixed structure of the CNN, composed of a variable number of interconnected layers. Weights are parameters (numbers), that are modified during the training process of the CNN. These numbers describe the strength of the connections between the nodes composing the different layers. Even though to the best of our knowledge, studies that adopt pre-trained commercial CNN for characterizing specifically forest CES are still lacking, several authors have used such models (e.g. Clarifai and Google Cloud Vision) to classify images acquired in the natural landscape (e.g. (Havinga et al., 2021; Payntar et al., 2021). Commercial pre-trained CNNs allow researchers to save time, however, their use is not for free, and the categories that they can identify cannot be modified according to the researchers' requirements. Transfer learning, a recent development of the deep learning field, is receiving a growing interest because it has the potential of tackling these limitations. Transfer learning allows researchers to adapt freely available pretrained CNNs, such as the ones trained on the ImageNet (Deng et al., 2009) and Places365 databases (Zhou et al., 2017), to new classification tasks. Recently, (Cardoso et al., 2022; Lingua et al., 2022a) have used transfer-learning to create CNNs purposely designed for characterizing the CES provided by natural parks located in the Iberian Peninsula (Spain and Portugal) and British Columbia (Canada) respectively. The use of transfer learning appears to be a promising approach to obtain quick, inexpensive, and detailed data that could be useful for forest CES management.

Despite the rapidly growing number of studies that use social media data for CES research, this discipline is still in its infancy, and its potential to inform ecosystem management is not fully understood (Ghermandi and Sinclair, 2019). In particular little is known about how to transform the information that can be extracted from crowdsourced social media data into useful outputs for the inclusion of CES in ecosystem management plans. Furthermore, remote sensing data and the output of automated image content analysis have not often been combined before, and the potential of their integration is not been fully explored. The overarching objective of this study, therefore, is to address these knowledge gaps, developing a methodology to effectively integrate CES in forest management plans, relying on remote sensing and crowdsourced social media data. To do so, geotagged social media images were downloaded and classified, using CNNs, based on the forest recreational activities depicted in them. The outcomes of the image classification process were then paired with their corresponding topographic, socioeconomic, and forest biometrics variables obtained using remote sensing technologies. Results provide insights into which recreational activities are most popular in the observed forests and which natural and social attributes may potentially drive their popularity. Maps of the recreational potential of forests as well as of forest recreation economic value are produced and the limitations and future perspectives of our approach are discussed.

2. Methods

2.1. Study area

The forested land within British Columbia's (BC) provincial park system (Fig. 1) covers approximately 68,000 km^2 (~7% of the entire Province), from temperate rainforests to alpine tundra biomes, and provides a wide array of ecosystem services. Among these, recreation plays an important role with BC parks offering 6000 km of hiking trails, and over 10,000 vehicle-accessible campsites, and are visited by more than 20 million recreationists annually.

2.2. Data and methods

The methodology applied in this study covers five main steps. First, the geotagged images acquired across BC forested parks were downloaded from Flickr's API. Second, the images were classified and filtered based on the recreational activities depicted in them, using convolutional neural networks (CNNs). Third, each relevant image were attributed, based on its geographic coordinates: topography (slope and elevation), forest biometrics (canopy height, forest cover, gross stem volume, total biomass), and anthropogenic impacts (global human influence index) variables. The global human influence index (GHII) is a numerical index ranging from 1 (no anthropic impacts) to 44 (maximum anthropic impacts), that synthesize information from different sources (taking into account the human population pressure, land use, the presence of infrastructure, and human access). GHII (Wildlife Conservation Society, 2005) and provides a map of anthropogenic impacts at 1 km^2 resolution. Fourth, a random forest classifier model was trained to identify the most likely forest recreational activity to occur in an area, based on the above-mentioned variables. Finally, two types of maps were produced, recreational potential maps and recreational value maps. The recreational potential maps were created using a random

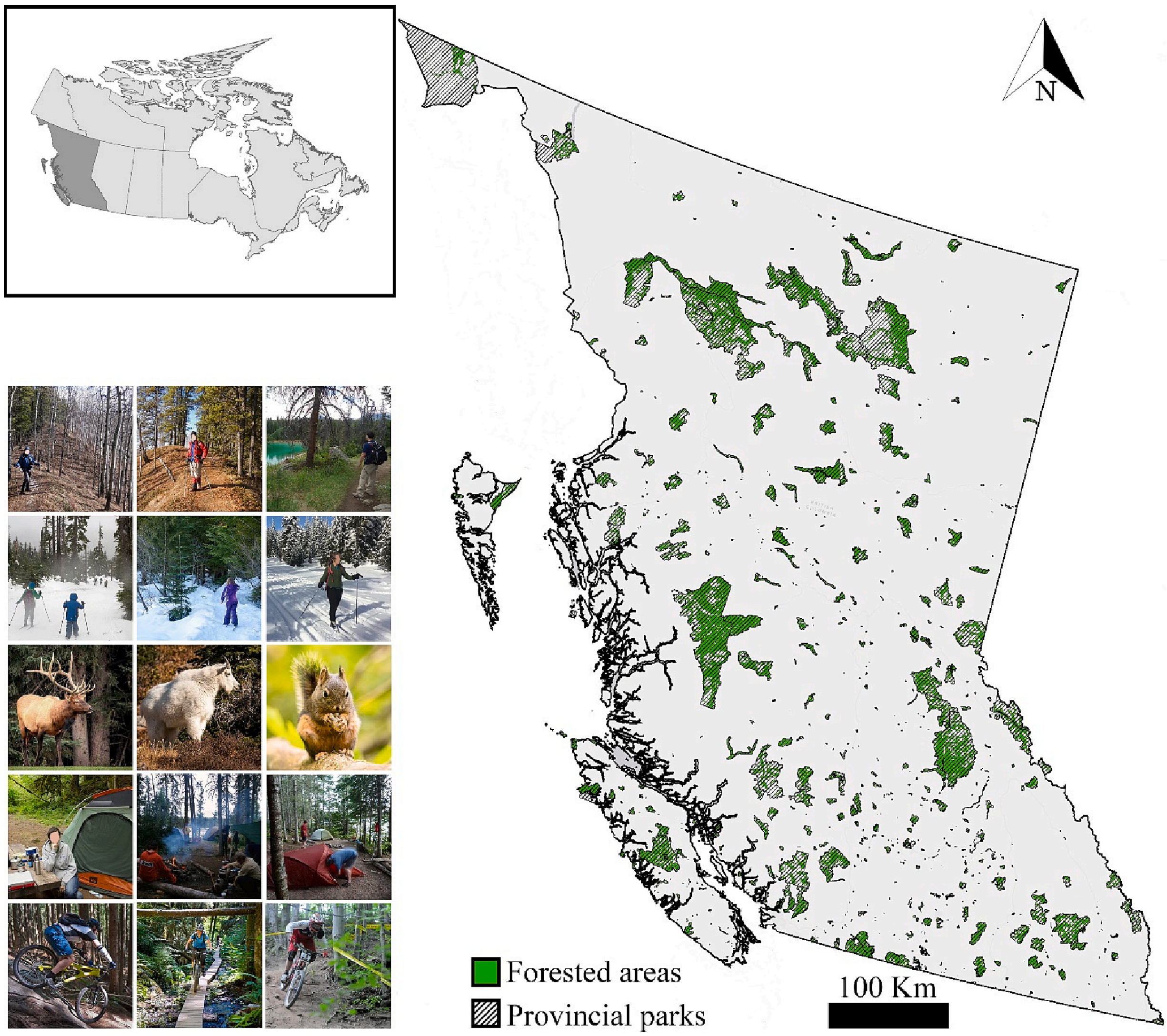

Fig. 1. Map of the study area, forests protected by the BC provincial park system. Pictures in the bottom left corner are a sample of the ones included in the study.

forest classifier model to predict the most suitable recreational activity for points regularly distributed in a grid with a spacing of 0.005 km. The recreational value maps were created by applying a crowdsourced benefit transfer approach. In the following sections, further information on each of these steps will be provided.

2.3. Image gathering

The images used in this study were downloaded from Flickr's API. Flickr is the most frequently used social media website in the study of CES, because it has an easily accessible API, and it offers the longest time series (Ghermandi and Sinclair, 2019). Images from Flickr were downloaded, using a purposely designed Python application, for two different objectives: (*i*) to explore the CES provided by the BC forested parks, and (*ii*) to train the CNN used in the image classification process.

Images used for CES exploration were obtained by downloading all the geotagged images acquired in BC between 2005-1-1 and 2020-12-31, together with their metadata. The metadata included: (*i*) image identification code; (*ii*) author identification code; (*iii*) date on which the image was taken; (*iv*) coordinates of where the image was taken.

The images used for the CNN training were obtained by combining the tags and bbox arguments of the "*flickr.photos.search*" function of Flickr's API. The tags argument allows the user of Flickr's API to query for images that were tagged with specific words by the author, the bbox argument allows the user of Flickr's API to query for images that were acquired in a minimum bounding box specifying its coordinates. The tags used are reported in the supplementary material (Table A1), with the coordinate bounding box (49.0 N, −114.0 E; 60.0 N, −70.0 E).

2.4. Image classification

The image classification process applied in this study makes use of three CNNs: the *relevance CNN*, the *cultural ecosystem services CNN*, and the *recreational CNN*. All three CNNs used the ResNet-152 architecture (He et al., 2016), a residual neural network architecture, 152 layers deep, that employs residual learning units. The *relevance CNN* was ResNet-152, trained by Zhou et al. (2017) on the Places365 database. Instead, the *cultural ecosystem services CNN* and the *recreation CNN* were purposely created by adopting a transfer learning approach. Transfer learning is an innovative methodology of the deep learning field that allows for the training of state-of-the-art CNN even when only a small training set is available (Howard and Gugger, 2020). The general idea of transfer learning is to fine-tune existing CNN to adapt it to a new classification task (Torrey and Shavlik, 2010). To do so, the last layers of the original CNN are discarded and replaced by new layers designed for the intended classification task. Lastly, the modified CNN was re-trained with a training set of images opportunely selected. In this study the CNN used as a starting point for the benefit transfer approach was

ResNet-152 trained on the ImageNet database (Deng et al., 2009).

The image classification process can be summarized in the following steps. First, each image was fed into the *relevance CNN* which was divided into two classes: relevant images and not-relevant images. Then each relevant image was fed into the *cultural ecosystem service CNN*, which classified the images into two classes: aesthetic ecosystem service and recreational ecosystem service. Lastly, the images classified as belonging to the recreational ecosystem services class were fed into the *recreational CNN*, to assign the following activities: skiing, hiking, camping, wildlife viewing, and biking. To account for the fact that among the gathered images there were also depicted scenes that did not belong to the above-mentioned categories, all the images classified by the three CNN with a confidence lower than 75% were excluded from the analyses.

To evaluate the performances of the image classification process, the following metrics were used: (*i*) accuracy (*ii*) precision; (*iii*) recall; (*iv*) F1-score (harmonic mean of precision and recall). These metrics were estimated on a randomly selected subset of 1,000 images.

2.5. Relevance CNN

The *relevance CNN* consists of the ResNet-152 architecture (He et al., 2016), pre-trained on the Places365 database. Places365 is an image database containing over 1.8 million classified images from 365 scene categories (e.g. rainforest, pier, restaurant, etc.) that belong to three macro-categories: indoor, nature, and urban. The *relevance CNN* was used to assess if images were relevant in the context of CES research. Images tagged with classes belonging to the nature category were classified as "relevant", while images tagged with classes belonging to the indoor and urban categories were classified as "not relevant". The images classified as relevant were then fed into the *cultural ecosystem service CNN*.

2.6. Cultural ecosystem services CNN

The *cultural ecosystem services CNN* uses the ResNet-152 architecture, pre-trained on the ImageNet database as a starting point, but this time the transfer learning approach is applied to adapt it to classify images in aesthetic and recreational. To do so a database, composed of 7,000 images (3,500 depicting aesthetic experiences and 3,500 depicting recreational experiences) was used. This database was created using *ad-hoc* training images obtained using the tags and bbox arguments of the "*flickr.photos.search*" function of the Flickr API.

The distinction between aesthetic and recreation experience adopted in this study broadly follows the one introduced by Richards and Friess (2015) between "Landscape" and "Social Recreation": In the aesthetic category, were included: (*i*) pictures depicting natural landscapes and skyscapes (*ii*) pictures depicting vegetation; (*iii*) pictures depicting natural features (e.g. mountains, lakes, streams, waterfalls, etc.). In the recreation category, were included: (*i*) pictures depicting people engaging in recreational activities; (*ii*) pictures depicting equipment related to recreational activities, but not people; (*iii*) pictures depicting wildlife and wildlife viewing equipment.

2.7. Recreation CNN

The *recreation CNN* uses the ResNet-152 architecture, trained on an image database composed of 3,500 pictures depicting various forest recreational activities. Again these training images were downloaded by combining the tags argument and bbox argument of the "*flickr.photos.search*" function of the Flickr's API. The recreational activities that were considered are the following: hiking, skiing, camping, biking, and wildlife viewing; the training set contained 700 images for each activity. The list of forest recreational activities was obtained by adapting the one from Rosenberger et al. (2017). The "hiking" category includes all the images depicting people hiking or posing in nature; the "skiing" category includes people skiing or skiing equipment; the "camping" category includes people camping or camping equipment; the "biking" category includes people biking or biking equipment; the "wildlife viewing" category includes images depicting wildlife in its natural habitat or wildlife viewing equipment.

2.8. Variables attribution and remote sensing technologies data

To assign to each image its corresponding topographic, anthropogenic impacts, and forest biometrics variables, the Python package "rasterstats" (https://pythonhosted.org/rasterstats/) was used. The temporal variable was directly extracted from the metadata of the image.

Topographic data have previously been found to influence the provision of forest recreation (e.g. (Roovers et al., 2002); (Abildtrup et al., 2013). In particular two topographic variables were used in this study: elevation and slope both derived from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) digital elevation model (GDEM V2, 30 m) (Tachikawa et al., 2011).

Several authors have shown the impacts that forest biometrics data have on forest recreation (e.g. (Agimass et al., 2018; Carvalho-Ribeiro and Lovett, 2011). To examine the influence of forest biophysical conditions on the provision of ecosystem services, we used four continuous wall-to-wall, 30-m forest structure metrics: (i.e., Lorey's height, basal area, volume, and above-ground biomass) based on annual composites of Landsat satellite imagery using the imputation method described in Matasci et al. (2018a,b). This method used airborne laser scanning (ALS) and field plot data to estimate forest structure metrics from topographic and the Landsat spectral predictors, using a k-Nearest Neighbor approach. In their work, the authors used a predictive modeling approach to map forest attributes in Canada's boreal forests. The model was calibrated using over 80,000 plots of a Canada-wide survey in which ALS was used to derive measurements of forest vertical structure. Specifically, the authors combined the forest structure variables surface reflectance composites derived from Landsat Thematic Mapper and Enhanced Thematic Mapper Plus imagery, obtaining r2 values ranging from 0.49 to 0.61. Lastly, previous studies conducted in the same study area have shown how forest recreation patterns are influenced by the passing of the seasons (Lingua et al., 2022; Lingua et al., 2022b). For this reason, an integer representing the day of the year in which the images were acquired was extracted from the metadata (1 for 1 January, 2 for 2 January, etc.).

2.9. The random forest classifier model

A random forest classifier was used to identify if and how the independent variables affect the use of forested areas for recreation purposes and assess the recreational potential for each forest recreational activity in Cypress park. The random forest model was created using the "RandomForestClassifier" module of the scikit-learn library.

To train the random forest classifier, the entire dataset of images depicting recreational activities was used. The dependent variable of the model was the five recreational activities identified by the recreation CNN (hiking, skiing, camping, biking, and wildlife viewing) while the independent variables were the topographic, forest biometrics, anthropic impact, and seasonal data previously illustrated. The dataset was divided randomly selecting 30% of the samples, used as a training set, while the remaining 70% was used as the validation set. The model included 30 fully-grown decision trees.

To identify which variables influenced the outcome of the model the "feature_importances_" attribute of the "RandomForestClassifier" module was used. The feature importance is the estimate of how much of the predicting power of the model is given by a variable and it is calculated as the decrease in node impurity weighted by the likelihood of reaching that node. The Gini impurity of a node can be defined as the likelihood of a random datum being misclassified if it were attributed to a random class (according to the class distribution in the dataset). Gini impurity can be calculated using the following formula

$$Giniimpurity = \sum_{i=1}^{C} f_i(1 - f_i)$$

Where f_i is the frequency of label i at a node and C is the number of unique labels.

To explore the relationship between the dependent variable and independent variables partial dependency plots were created. Partial dependency plots show the marginal effect that a variable has on the outcome of the model. These plots were created for each recreational activity and variable pair.

2.10. Mapping the recreational potential

To demonstrate the potential of the method, the random forest classifier model was used to map the recreational potential of Cypress provincial park across the four seasons. Cypress park is one of the most visited provincial parks in BC (BC Parks statistics).

To do so a grid of 200,000 regularly distributed points and a fishnet made of hexagonal cells of 1.5 ha each were created. First to each point were assigned the independent variables: elevation, slope, anthropic impact (GHII), canopy height, basal area, gross stem volume, and total biomass. To do so the Python module "rasterstats" was used (https://pythonhosted.org/rasterstats/). Second, the random forest classifier model described above was deployed to predict the probabilities for each forest recreational activity to be undertaken in each point for each season. Third, the probabilities estimated for points falling inside the same cell were averaged. Lastly, suitability maps for each recreational activity in each season were created based on the average probabilities estimated by the model in each cell.

2.11. Mapping the value of recreation

To estimate the monetary value of the recreational service provided by the forests included in the BC provincial parks system two data sources were used the consumer surpluses for forest recreational activities reported by Rosenberger et al. (2017), and the yearly visitation statistics by BC Parks. Rosenberger et al. (2017) reported the consumer surpluses that people enjoy from participating in forest recreational activities in North America, estimated by performing a meta-analysis of 342 recreation economic studies. For the forest recreational activities included in this study, the following consumer surpluses (expressed in 2022 C$) were used: hiking 96.3 C$, skiing 79.3 C$, camping 40.6 C$, wildlife viewing 68.5 C$, biking 98.9 C$. BC Parks estimated the number of daily visitors in all of BC Parks (from 2012 to 2018). These estimates are reported as the number of yearly visitors available in the end-of-the-year reports (https://bcparks.ca/research/). Cypress park, the most visited provincial park in BC, received more than 1.5 million visitors each year.

The approach used for the monetary valuation was the crowdsourced benefit transfer. When applying the benefit transfer method, the consumer surplus generated in a year by a recreational site is calculated using the following formula (Eq. (1))

$$CS = \sum_{i=0}^{n} (value\,act._i \times (avg.n^{\circ}\ annual\,visits \times ratio\,activity_i)) \tag{1}$$

Where $valueact._i$ is the CS associated with the recreational activity i; $avg.n^{\circ} annualvisitors$ is the average number of annual visits; $ratioactivity_i$ is the ratio between the number of people engaging in activity i and the total number of people visiting the site.

In this study, this formula was applied in each cell of the same grid used for mapping the recreational potential of Cypress provincial park. The first step of the valuation process was to estimate the average number of visits in each cell, in every season. To do so, an assumption that there is a linear relationship between the number of visitors and the number of images acquired was made. Therefore, the number of visitors in each cell, for every season, was estimated using the following formula (Eq. (2))

$$avg.n^{\circ} annualvisits = \frac{n^{\circ} imgincellduringseason}{totn^{\circ} imginpark} \times avg.n^{\circ} vistorsinpark \tag{2}$$

The ratios of the various recreational activities were estimated by grouping together the images acquired by the same user, during the same day, in activity user day (AUD). The activity assigned to each AUD (primary activity) was the most frequently depicted one among the images composing the same AUD. Then in each cell of the grid was estimated the ratio for each recreational activity, using the following formula:

$$ratioactivity_i = \frac{n^{\circ} AUDsactivity_i}{n^{\circ} totalAUDs} \tag{3}$$

Lastly, Eq. (2) was applied in each cell of the grid, obtaining the annual consumer surplus generated.

3. Results

During the image gathering process, 1,398,737 images taken in BC between 2005/01/01 and 2020/12/31 were downloaded *via* Flickr API. Among these images, 91,819 (6.6%) were acquired within the boundaries of a provincial park and of these, only 64,670 (4.6%) were acquired in forested areas. The outputs of the image classification process are shown in Fig. 2. The relevance CNN classified as "not relevant" 3,070 images (4.8%) that were excluded from further analyses. The CES CNN classified as "aesthetic ES" 39,660 (61.4%) images that were also excluded, while 21,863 (33.8%) were classified as "recreation ES" and fed into the level 2 CNN. The recreation CNN classified the remaining pictures as follows: 1,334 (2.1%) as "biking", 2,359 (3.5%) as "camping", 7,222 (11.2%) as "hiking", 2,919 (4.5%) as "wildlife viewing", 3,791 (5.9%) as "skiing". Pictures classifies as "climbing" and "water-related activities" were in total 4,238 (6.6%) and were excluded from the analysis.

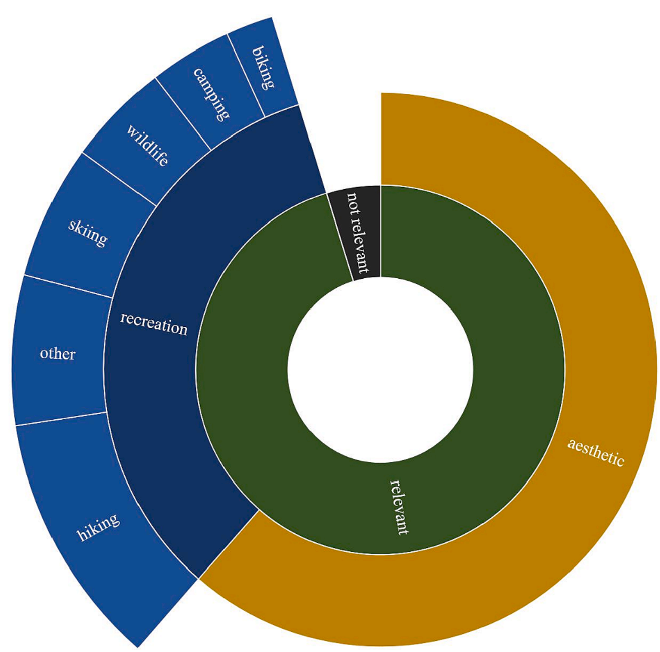

Fig. 2. Outcomes of the image classification process undertaken on Flickr images acquired in the BC parks system from 2005 to 2020. The inner layer represents Relevance CNN, the mid layer represents CES CNN, external layer represents recreation CNN.

The overall performances of the image classification process are reported in Table 1. The overall accuracy of the CNN is 0.84, however, the performances vary for the different classes. In particular, the relevance CNN that identified and excluded from further analyses the non-relevant images had the poorest performances with a F1-score of 0.62. The CES CNN had the strongest performance, with a F1-score of 0.94. Lastly the recreation CNN performed well in the classification of images depicting hiking, wildlife and skiing, and biking with F1-scores higher than 0.7, while images depicting camping had F1-scores of 0.56.

Combining the outcomes of the image classification process with the variables assigned to each image, it was possible to estimate the averages and standard deviations of the topographic variables, anthropic impacts, temporal variables, and biometrics variables within each recreational activity. In Table 2 are shown the averages and standard deviations, while in Table 3 are shown the differences between the averages of the activity and their statistical significance assessed with the ANOVA test.

“biking” and “skiing” are the activities characterized by the highest GHII, while “hiking”, “camping” and “wildlife viewing” had the lowest GHII. Furthermore, images depicting “hiking” and “skiing” were acquired on average at the highest elevations and on the steepest slopes.

Images depicting “skiing” and “wildlife viewing” were acquired in less dense forests (lower basal area, total biomass, and gross stem) while there are no statistically significant differences in the canopy height variable.

The random forest classifier was trained on the 17,805 images classified as “hiking”, “skiing”, “camping”, “biking” or “wildlife viewing” by the level 2 CNN and had an accuracy of 74%. As shown in Fig. 3, the most important variable in improving the accuracy of the predictions is seasonality (i.e. the day of the year in which the picture was acquired) contributing to ~25% of the decrease in Gini impurity. Topographic variables are the second most important variables, contributing to a decrease of ~18% (elevation) and ~13% (slope). Anthropic impacts (estimated using GHII) contributed to ~12% of the decrease in Gini impurity, while all the forest biometrics variables had a less important contribution (<10%).

Fig. 4 shows the partial dependency plots generated from the random forest classifier model, for the four most influential variables. Only the variables that contributed the most to the model performances were included (Gini impurity reduction >10 %). The graphs illustrate how the variables influence the likelihood of an image being classified as depicting one of the recreational activities. Specifically, these plots depict the relationship between the input variables and the predictions, showing how the predictions partially depend on the values of the input variables of interest. The plots in Fig. 4 are 1-way plots since they showcase how the random forest model’s predictions depend on a single input.

As shown in Fig. 4, the images acquired during the summer months (from 172 to 265) are more likely to depict “hiking” and “camping”, while the likelihood of an image to depict “wildlife viewing” activities increases during spring (80–171) and autumn (266–355). As expected, the likelihood of an image depicting “skiing” decreases rapidly during summer months and peaks during winter (355–365 and 1–79).

Most of the forest recreational activities are more likely to occur in forests at low elevations, except for “skiing” whose likelihood peaks around 1,000 m. Furthermore, forests characterized by low slopes favor “camping” and “wildlife viewing”, intermediate slopes favor “biking”, and high slopes favor “hiking”.

Lastly, anthropic impacts show a less clear trend, “biking” activity is more likely to occur where the anthropic impact on the forest is highest, while “camping” activities have the opposite behavior.

In Fig. 5 are presented the recreational potential for hiking and skiing in Cypress provincial park across the seasons while Fig. 6 are shown the recreational activities with the highest potential to occur across the four seasons in Cypress provincial park. Hiking is by far the activity with the highest potential in spring, summer, and fall, while skiing has the highest potential during winter. In particular, skiing has the highest potential in the southeast portion of the park, especially during spring and fall, while during summer it is absent. Other recreational activities, such as wildlife viewing, camping, and biking have low potential in this provincial park.

In the last step of the analysis (Fig. 7) an assessment of the monetary value of the recreational ecosystem service provided by Cypress park was undertaken. Using a crowdsourced benefit transfer approach, it was possible to estimate in every season, for each cell, the value of recreation expressed in C$/ha/day. As shown in Fig. 5, for Cypress park it is possible to identify two hotspots of recreational value, one in the southeast of the park and one in the central part of the park where most of the infrastructures (parking lots, hiking trails, Nordic ski, and sledding areas, etc.).

4. Discussions and conclusion

Crowdsourced social media images are a valuable data source to explore CES provision and consumption, however, to date, many of the existing studies consider exclusively the image’s metadata, not taking full advantage of the possibilities that crowdsourced social media data offer. In this study, carried out in BC forests included in the provincial parks system, Flickr images were automatically classified with purposely developed CNNs and coupled with topographic, socioeconomic, and forest biometrics variables obtained *via* remote sensing technologies. This approach allowed us to examine if and how these variables influence the popularity of various forest recreational activities and map both the recreational potential and the value of the recreation CES.

4.1. Outcomes and performances

The performances of the CNNs adopted to automatically classify the images used in this study align with the ones obtained by the previous applications of transfer learning to the study of CES (Cardoso et al., 2022; Gosal and Ziv, 2020). The performances however were not homogenous for all the classes. In particular, the classification process performed poorly when identifying not relevant images and images depicting people biking. Based on the automated image classification process, the most popular recreational activity in BC forests is hiking, followed (in descending order of popularity) by skiing, wildlife viewing, camping, and biking. These results compare well with those of a conventional survey administrated to BC residents by Kux and Haider (2014) where BC residents were asked to indicate in which recreational activities they participated during 2012 and found that the most popular recreational activity among BC residents is hiking, followed by skiing, fishing, and biking. Although the options given to the respondents of the survey differ from the activities considered in this analysis the two most popular activities (hiking and skiing) match.

Previous studies suggested that the recreational attractiveness of forested areas is influenced by: (*i*) topographic variables such as slope

Table 1
Overall performances of the image classification process.

Class	Precision	Recall	F1-score	N° images
Not relevant	0.71	0.52	0.62	23
Aesthetic	0.93	0.95	0.94	536
Biking	1	0.49	0.74	17
Camping	0.58	0.54	0.56	41
Hiking	0.66	0.76	0.71	125
Wildlife	0.76	0.96	0.85	49
Skiing	0.82	0.77	0.8	48
Others	0.65	0.54	0.59	61
accuracy			0.85	900
macro avg	0.76	0.69	0.73	900
weighted avg	0.84	0.84	0.83	900

Table 2
Values of topographic, anthropic impacts, seasonal, and forest biometrics data between the various forest recreational activities. Standard deviations are reported in brackets.

Activity	GHII	Slope (°)	Elevation (m)	Canopy height (m)	Basal Area (m^2/ha)	Gross Stem (m^3/ha)	Total Biomass (tonnes/ha)
Hiking	19.4 (16.0)	16.3 (13.8)	1034.1 (809.2)	25.5 (8.1)	45.0 (22.7)	621.8 (454.1)	241.0 (153.4)
Biking	23.6 (14.1)	9.1 (9.7)	731.2 (605.2)	24.9 (7.7)	43.4 (22.3)	588.4 (429.7)	240.0 (154.3)
Skiing	24.8 (13.7)	13.3 (12.7)	1325.7 (523.1)	23.6 (7.4)	36.7 (21.0)	468.0 (378.1)	194.0 (132.8)
Camping	20.0 (13.8)	9.1 (11.1)	611.4 (614.1)	23.7 (8.1)	40.8 (22.0)	535.5 (416.9)	221.2 (150.5)
Wildlife	19.7 (12.3)	12.9 (10.0)	633.6 (617.6)	22.4 (8.0)	37.3 (22.3)	471.4 (408.0)	199.3 (150.3)

Table 3
Differences in absolute value between the averages and statistical significance according to ANOVA test. Values in bold are statistically significant.

Activity 1	Activity 2	GHII	Slope (°)	Elevation (m)	Canopy height (m)	Basal Area (m^2/ha)	Gross Stem (m^3/ha)	Total Biomass (tonnes/ha)
Skiing	Wildlife	**5.0**	**3.8**	**692.1**	1.8	**4.2**	**86.3**	1.0
Skiing	Camping	**4.7**	**3.8**	**714.3**	3.0	**7.7**	**153.8**	18.8
Skiing	Hiking	**5.4**	**3.4**	**291.6**	1.9	1.6	**150.3**	**40.7**
Skiing	Biking	1.2	0.5	**594.4**	0.5	**8.3**	33.4	**46.0**
Wildlife	Camping	0.3	0.0	22.2	1.2	**3.5**	**67.5**	**19.8**
Wildlife	Hiking	0.3	**7.2**	**400.5**	0.1	2.6	**64.0**	**41.8**
Wildlife	Biking	**3.8**	**4.3**	97.7	1.3	**4.1**	52.9	**47.0**
Camping	Hiking	0.6	**7.2**	**422.7**	1.1	**6.1**	3.5	**21.9**
Camping	Biking	**3.5**	**4.3**	**119.9**	2.5	0.6	**120.4**	**27.2**
Hiking	Biking	**4.2**	**2.9**	**302.8**	1.4	**6.7**	**116.9**	5.3

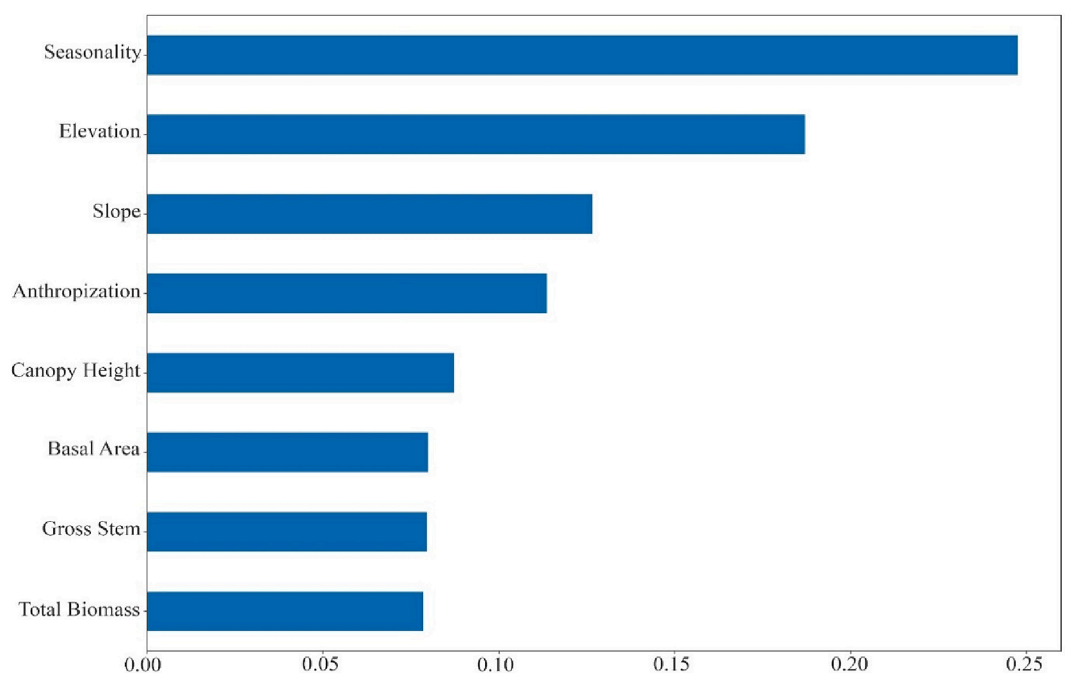

Fig. 3. Variables importance, on the x-axis is represented the percentual decrease in node impurity.

and elevation (Abildtrup et al., 2013), (*ii*) tree stand characteristics such as tree stock and crown closure (e.g. (Weller and Elsasser, 2018) (Filyushkina et al., 2017), and anthropic impacts in the area (Roovers et al., 2002). While these variables seem to affect the number of visitors in forests, not all of them had a significant effect on the type of recreational activities that people engage in. Specifically, it appears that the variables that have the most influence in determining the type of forest recreation are temporal and topographic, while forest biometrics variables play a less important role.

The recreational values maps show how most of the consumer surplus provided by the park's forests is concentrated in two hotspots (one in the southeast zone and one in the central zone), while in most of the cells created it was not possible to estimate a value since no images were acquired. This could suggest that even in forests that are managed for promoting recreational activities, visitors tend to coagulate in specific areas.

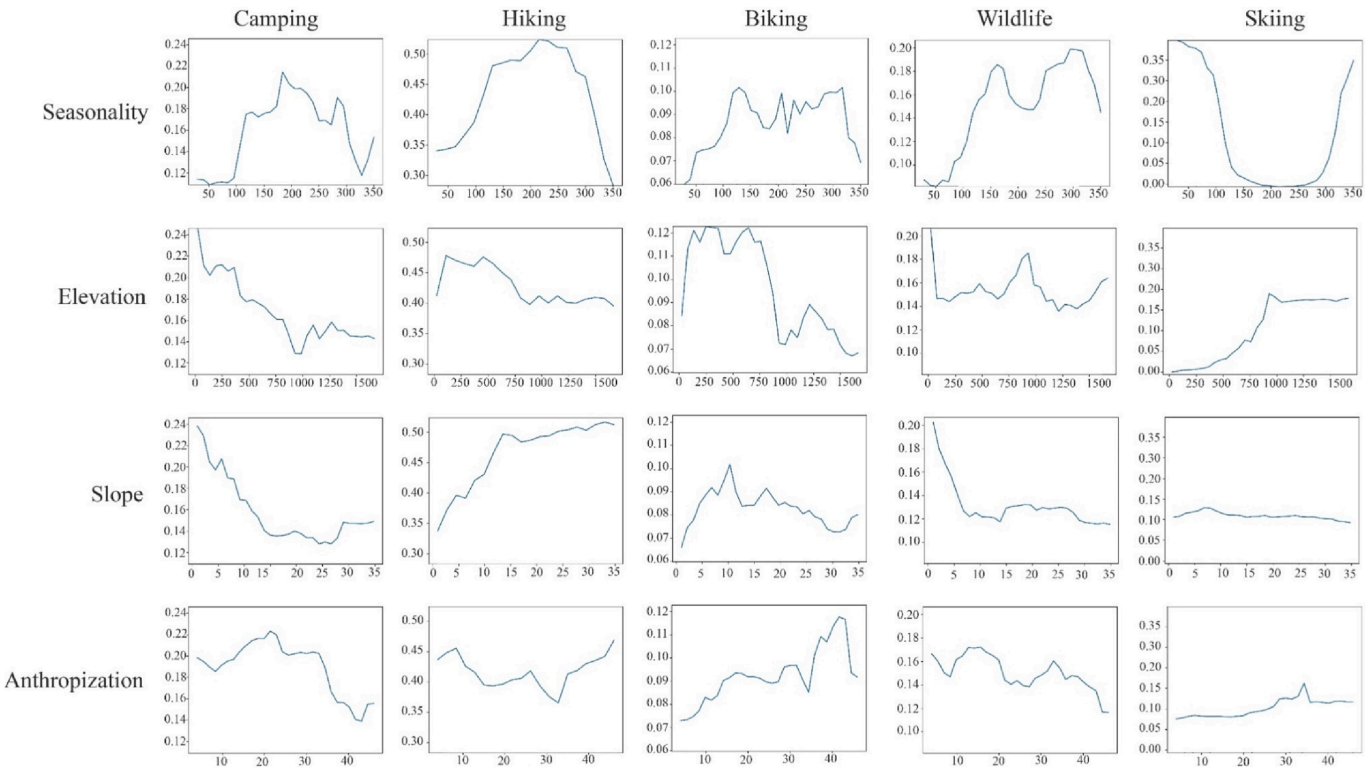

Fig. 4. Partial dependency plots showcase how the probability predictions of the model partially depend on the values of the input variables of interest. On the x-axes are the different variables (Julian day of the year for seasonality, m above sea level for elevation, degrees for slope, and GHII for anthropic impacts) on the y-axes is instead the predicted probability of the image to be classified as belonging to the class.

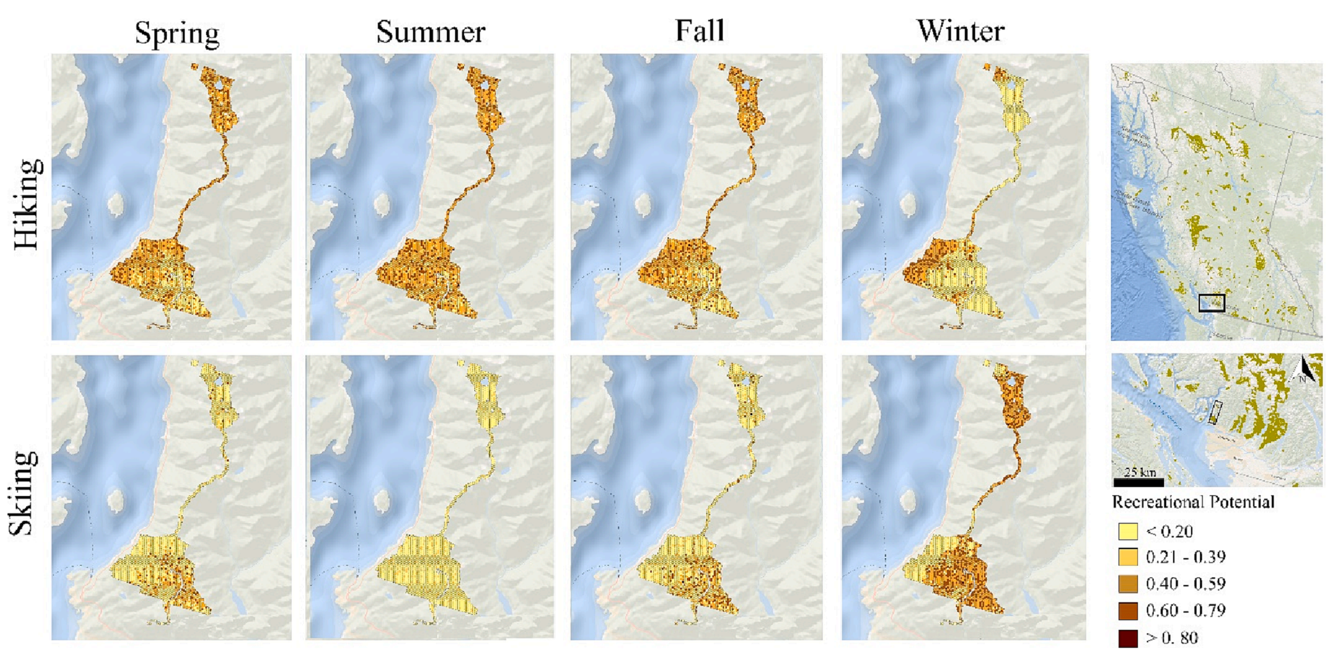

Fig. 5. Recreational potential estimated for hiking and skiing in Cypress park across the seasons. The values reported should be interpreted as a 0–1 suitability score.

4.2. *Innovativeness, Limitations, and future perspectives*

The literature around the use of social media data for CES exploration has been focusing on their use to describe their demand and provision. Less attention has been given to how these data could be used at an operational level in the management of the environment. In this study, combining social media data with remote sensing data, and applying machine learning techniques, it was possible to obtain fine-grained information on how forest and landscape variables influence the type of recreation in which forest visitors engage. Previously, to obtain such data the most common approach was to resort to *in-situ* surveys which are costly and time-consuming (Richards and Tunçer, 2018). Instead, the methodology used in this study has shown the potential to be a viable alternative to such surveys. Furthermore, the

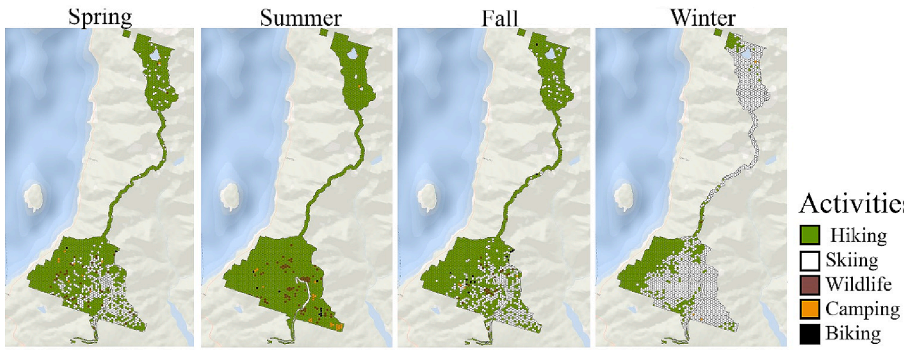

Fig. 6. Recreational activities with the highest probability to be carried out in the forested areas of Cypress park.

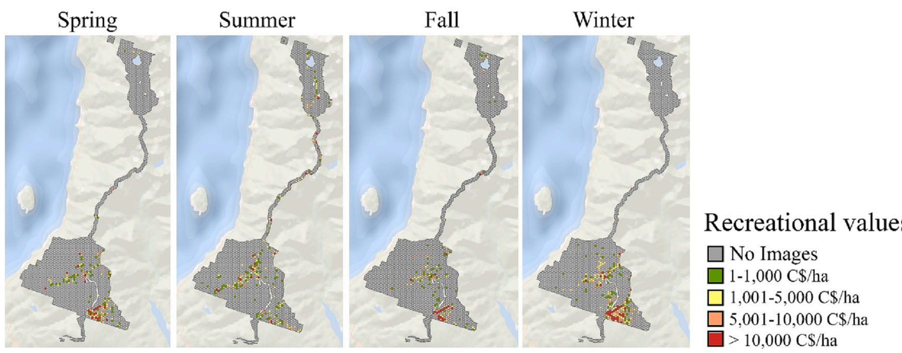

Fig. 7. Economic value of the recreational ecosystem service provided by Cypress park.

produced maps provide useful insights to forest managers through the recreational potential map and a recreational values map. Recreational potential maps could be used to plan and locate the implementation of forest recreational infrastructures such as hiking or biking trails, ensuring that the chosen forested area is suitable for that activity. Recreational values maps could be useful in providing detailed insights into the potential costs of partial park closure or the impacts of forest disturbances.

Despite the promise of the approach adopted in this study, future in-field applications of it have some limitations. Ciesielski and Stereńczak (2021) have argued that social media users' demographic characteristics could substantially differ from the ones of forest recreationists, causing the selection of a nonrepresentative sample. However, Flickr data appear to be less prone to this bias compared to other social media data (Hausmann et al., 2018). Furthermore, the various forest recreational activities could be characterized by the different frequencies of image acquisition, for example, hikers could be more inclined to take pictures than bikers. This could limit the availability of images portraying certain forest recreational activities, and affect the ratio of activities used in the assessment of recreational values. Also, the crowdsourced benefit transfer approach that was used in this study does not allow the assessment of the value of recreation in areas in which no images were acquired. Therefore, this method will rarely be applicable in difficult-to-access forested areas, as highlighted by the fact that most of the cells in Cypress park had no images. In addition, the possibility of freely accessing social media data is not granted, since changes in social media platform policies can happen abruptly. Lastly, even though crowdsourced social media data are often referred to as "big data", the share of relevant images among the ones gathered is only 4.6%. This indicates that social media data should not be seen as a panacea for the study of forests CES, but rather as a useful source of insights on CES provision and demand, especially in forests with high recreational value and peri-urban forests.

So far, the research around the use of social media data in the study of forest CES, has been mostly focused on the development of methodologies and approaches to extracting spatial and quantitative information on recreational fluxes. This study suggests that social media data and in particular images represent often overlooked opportunities for exploring the demand for forest CES from a qualitative point of view. However, to unlock the full potential of social media data for forest management, additional studies are needed. In particular, we believe that future research should focus on two objectives. The first objective is to obtain a better understanding of the relationship between *in-situ* and crowdsourced social media data. Is the ratio of activities estimated *via* social media data coherent with the one obtained *via in-situ* surveys? If biases are introduced by analyzing social media images, are they consistent in different study areas? Answering these questions would allow forest managers to confidently use the automated analyses of social media images to gather detailed, and almost real-time information on the value of the recreational ecosystem service that the forests provide by applying the crowdsourced benefit transfer approach adopted in this study. The second objective is the creation and sharing among researchers of databases of images depicting forest recreational activities. These databases would allow research groups worldwide to train new and more accurate CNN for the classification of images based on the depicted recreational activities. Ultimately, this would allow for the application of the methodology applied in this study in new contexts in which alternative approaches are difficult to apply, such as developing countries and remote areas.

CRediT authorship contribution statement

Federico Lingua: Conceptualization, Methodology, Software, Data curation, Writing – original draft, Visualization. **Nicholas C. Coops:** Writing – review & editing, Supervision. **Verena C. Griess:** Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecolind.2023.110165.

References

Abildtrup, J., Garcia, S., Olsen, S.B., Stenger, A., 2013. Spatial preference heterogeneity in forest recreation. Ecol. Econ. 92, 67–77.
Agimass, F., Lundhede, T., Panduro, T.E., Jacobsen, J.B., 2018. The choice of forest site for recreation: A revealed preference analysis using spatial data. Ecosyst. Serv. 31, 445–454.
Bernetti, I., Chirici, G., Sacchelli, S., 2019. Big data and evaluation of cultural ecosystem services: an analysis based on geotagged photographs from social media in Tuscan forest (Italy). Iforest-Biogeosci. For. 12, 98–105. https://doi.org/10.3832/ifor2821-011.
Bettinger, P., Boston, K., Siry, J.P., Grebner, D.L., 2016. Forest Management and Planning. Academic press.
Brown, G., 2017. A review of sampling effects and response bias in internet participatory mapping (PPGIS/PGIS/VGI). Trans. GIS 21 (1), 39–56.
Cardoso, A.S., Renna, F., Moreno-Llorca, R., Alcaraz-Segura, D., Tabik, S., Ladle, R.J., Vaz, A.S., 2022. Classifying the content of social media images to support cultural ecosystem service assessments using deep learning models. Ecosyst. Serv. 54, 101410.
Carvalho-Ribeiro, S.M., Lovett, A., 2011. Is an attractive forest also considered well managed? Public preferences for forest cover and stand structure across a rural/urban gradient in northern Portugal. Forest Policy Econ. 13 (1), 46–54.
Ciesielski, M., Stereńczak, K., 2021. Using Flickr data and selected environmental characteristics to analyse the temporal and spatial distribution of activities in forest areas. Forest Policy Econ. 129, 102509.
Deng J., Dong W., Socher R., Li L.-J., Li K., Fei-Fei, L. 2009. Imagenet: A large-scale hierarchical image database. 2009 IEEE conference on computer vision and pattern recognition.
Filyushkina, A., Agimass, F., Lundhede, T., Strange, N., Jacobsen, J.B., 2017. Preferences for variation in forest characteristics: Does diversity between stands matter? Ecol. Econ. 140, 22–29.
Ghermandi, A., 2018. Integrating social media analysis and revealed preference methods to value the recreation services of ecologically engineered wetlands. Ecosyst. Serv. 31, 351–357.
Ghermandi, A., Depietri, Y., Sinclair, M., 2022. In the AI of the beholder: A comparative analysis of computer vision-assisted characterizations of human-nature interactions in urban green spaces. Landsc. Urban Plan. 217, 104261.
Ghermandi, A., Sinclair, M., 2019. Passive crowdsourcing of social media in environmental research: A systematic map. Glob. Environ. Chang. 55, 36–47.
Gosal, A.S., Ziv, G., 2020. Landscape aesthetics: Spatial modelling and mapping using social media images and machine learning. Ecol. Ind. 117, 106638.
FAO (2020). The State of the World's Forests (SOFO).
Hausmann, A., Toivonen, T., Slotow, R., Tenkanen, H., Moilanen, A., Heikinheimo, V., Di Minin, E., 2018. Social media data can be used to understand tourists' preferences for nature-based experiences in protected areas. Conserv. Lett. 11 (1), e12343.
Havinga, I., Marcos, D., Bogaart, P.W., Hein, L., Tuia, D., 2021. Social media and deep learning capture the aesthetic quality of the landscape. Sci. Rep. 11 (1), 1–11.
He K., Zhang X., Ren S., Sun J. 2016. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition.
Hościło, A., Tomaszewska, M., 2014. CORINE Land Cover 2012–4th CLC inventory completed in Poland. Geoinformation Issues 6 (1), 49–58.
Howard, J., Gugger, S., 2020. Deep Learning for Coders with fastai and PyTorch. O'Reilly Media.
Kux S., Haider W. 2014. Non-Motorized Outdoor Recreation in British Columbia in 2012: Participation and Economic Contributions.
Lingua, F., Coops, N.C., Lafond, V., Gaston, C., Griess, V.C., Sharma, L.K., 2022. Characterizing, mapping and valuing the demand for forest recreation using crowdsourced social media data. PLoS One 17 (8), e0272406.
Lingua, F., Coops, N.C., Griess, V.C., 2022. Valuing cultural ecosystem services combining deep learning and benefit transfer approach. Ecosystem Services. Ecosystem Services 58, 101487. https://doi.org/10.1016/j.ecoser.2022.101487.
Matasci, G., Hermosilla, T., Wulder, M.A., White, J.C., Coops, N.C., Hobart, G.W., Bater, C.W., 2018. Three decades of forest structural dynamics over Canada's forested ecosystems using Landsat time-series and lidar plots. Remote Sensing of Environment 216, 697–714.
Matasci, G., Hermosilla, T., Wulder, M.A., White, J.C., Coops, N.C., Hobart, G.W., Zald, H.S., 2018. Large-area mapping of Canadian boreal forest cover, height, biomass and other structural attributes using Landsat composites and lidar plots. Remote Sensing of Environment 209, 90–106.
Norton, L.R., Inwood, H., Crowe, A., Baker, A., 2012. Trialling a method to quantify the 'cultural services' of the English landscape using Countryside Survey data. Land Use Policy 29 (2), 449–455.
Paracchini, M.L., Zulian, G., Kopperoinen, L., Maes, J., Schägner, J.P., Termansen, M., Zandersen, M., Perez-Soba, M., Scholefield, P.A., Bidoglio, G., 2014. Mapping cultural ecosystem services: A framework to assess the potential for outdoor recreation across the EU. Ecol. Ind. 45, 371–385.
Pastur, G.M., Peri, P.L., Lencinas, M.V., García-Llorente, M., Martín-López, B., 2016. Spatial patterns of cultural ecosystem services provision in Southern Patagonia. Landsc. Ecol. 31 (2), 383–399.
Payntar, N.D., Hsiao, W.-L., Covey, R.A., Grauman, K., 2021. Learning patterns of tourist movement and photography from geotagged photos at archaeological heritage sites in Cuzco, Peru. Tour. Manage. 82, 104165.
Pleasant, M.M., Gray, S.A., Lepczyk, C., Fernandes, A., Hunter, N., Ford, D., 2014. Managing cultural ecosystem services. Managing cultural ecosystem services. *Ecosyst. Serv.* 8, 141–147.
Plieninger, T., Dijks, S., Oteros-Rozas, E., Bieling, C., 2013. Assessing, mapping, and quantifying cultural ecosystem services at community level. Land Use Policy 33, 118–129.
Potschin, M., Haines-Young, R., 2013. Landscapes, sustainability and the place-based analysis of ecosystem services. Landsc. Ecol. 28 (6), 1053–1065.
Ritchie, H., & Roser, M. (2021). Forests and deforestation. Our World in Data.
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. Ecol. Ind. 53, 187–195.
Richards, D.R., Tunçer, B., 2018. Using image recognition to automate assessment of cultural ecosystem services from social media photographs. Ecosyst. Serv. 31, 318–325.
Ritter, E., Dauksta, D., 2013. Human–forest relationships: ancient values in modern perspectives. Environ. Dev. Sustain. 15 (3), 645–662.
Rodriguez, E., Morris, C.S., Belz, J.E., 2006. A global assessment of the SRTM performance. Photogramm. Eng. Remote Sens. 72 (3), 249–260.
Roovers, P., Hermy, M., Gulinck, H., 2002. Visitor profile, perceptions and expectations in forests from a gradient of increasing urbanisation in central Belgium. Landsc. Urban Plan. 59 (3), 129–145.
Rosenberger R. S., White E. M., Kline J. D., Cvitanovich C. 2017. Recreation economic values for estimating outdoor recreation economic benefits from the National Forest System. *Gen. Tech. Rep. PNW-GTR-957. Portland, OR: US Department of Agriculture, Forest Service, Pacific Northwest Research Station. 33 p., 957*.
Schirpke, U., Meisch, C., Marsoner, T., Tappeiner, U., 2018. Revealing spatial and temporal patterns of outdoor recreation in the European Alps and their surroundings. Ecosyst. Serv. 31, 336–350.
Sinclair, M., Ghermandi, A., Sheela, A.M., 2018. A crowdsourced valuation of recreational ecosystem services using social media data: An application to a tropical wetland in India. Sci. Total Environ. 642, 356–365.
Tachikawa T., Hato M., Kaku M., Iwasaki A. 2011. Characteristics of ASTER GDEM version 2. In: 2011 IEEE international geoscience and remote sensing symposium.
Tenkanen, H., Di Minin, E., 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. Sci. Rep. 7 (1), 1–11.
Termansen, M., McClean, C.J., Jensen, F.S., 2013. Modelling and mapping spatial heterogeneity in forest recreation services. Ecol. Econ. 92, 48–57.
Torrey, L., Shavlik, J., 2010. Transfer learning. In: Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques. IGI global, pp. 242–264.
Weller, P., Elsasser, P., 2018. Preferences for forest structural attributes in Germany-Evidence from a choice experiment. Forest Policy Econ. 93, 1–9.
Wildlife Conservation Society - WCS, Center for International Earth Science Information Network - CIESIN - Columbia University. (2005). *Last of the Wild Project, Version 2, 2005 (LWP-2): Global Human Influence Index (HII) Dataset (Geographic)* NASA Socioeconomic Data and Applications Center (SEDAC).
You, S., Zheng, Q., Chen, B., Xu, Z., Lin, Y., Gan, M., Zhu, C., Deng, J., Wang, K.e., 2022. Identifying the spatiotemporal dynamics of forest ecotourism values with remotely sensed images and social media data: A perspective of public preferences. J. Clean. Prod. 341, 130715.
Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., Torralba, A., 2017. Places: A 10 million image database for scene recognition. IEEE Trans. Pattern Anal. Mach. Intell. 40 (6), 1452–1464.