ElQadi M., Lesiv M., Dyer A.G., Dorin, A., "Computer vision-enhanced selection of geo-tagged photos on social network sites for land cover classification", Environmental Modelling and Software, 128, June 2020, 104696, (doi.org/10.1016/j.envsoft.2020.104696), 7pp.

Computer vision-enhanced selection of geo-tagged photos on social network sites for land cover classification

Moataz Medhat ElQadi [a, b], Myroslava Lesiv [b], Adrian G. Dyer [c], Alan Dorin [a]

[a] Faculty of Information Technology, Monash University, Clayton, VIC 3800, Australia
[b] Ecosystems Services and Management Program, International Institute for Applied Systems Analysis (IIASA), Laxenburg A-2361, Austria
[c] School of Media and Communications, RMIT University, Melbourne, VIC 3000, Australia

Abstract

Land cover maps are key elements for understanding global climate and land use. They are often created by automatically classifying satellite imagery. However, inconsistencies in classification may be introduced inadvertently. Experts can reconcile classification discrepancies by viewing satellite and high-resolution images taken on the ground.

We present and evaluate a framework to filter relevant geo-tagged photos from social network sites for land cover classification tasks. Social network sites offer massive amounts of potentially relevant data, but its quality and fitness for research purposes must be verified.

Our framework uses computer vision to analyse the content of geo-tagged photos on social network sites to generate descriptive tags. These are used to train artificial neural networks to predict a photo's relevance for land cover classification. We apply our models to four African case studies and their neighbours. The framework has been implemented within Geo-Wiki to fetch relevant photos from Flickr.

Keywords

land cover, social network, geo-tagged photos, computer vision, machine learning

1 Introduction

World Wide Web data have been broadly applied to research applications including public opinion measurement (O'Connor, Balasubramanyan, Routledge, & Smith, 2010) and election forecasting (Tumasjan, Sprenger, Sandner, & Welpe, 2010), epidemiology (Culotta, 2010), and even to philosophy (Cheong, 2018). Geo-tagged visual media in particular, as a form of Volunteered Geographic Information (VGI), has seen a strong interest in scientific research, as they allowed scientists to make firsthand observations from remote, sparsely populated places that would be otherwise impractical or expensive for data collection (Barve, 2014; Daume, 2016; ElQadi et al., 2017; Estima & Painho, 2013). This ability to make such omnipresent observations is particularly useful in land cover research to improve land cover maps.

Land cover impacts global climate by changing biogeochemical cycles and consequently the composition of the atmosphere, as well as changing the biogeophysical processes that affect energy absorption at the Earth's surface (Feddema et al., 2005). An understanding of changes in land cover and the associated monitoring of human land use such as agriculture, mining, and urban development, also enables us to better grasp human encroachment on natural ecosystems and habitats. Mapping land cover is therefore essential for understanding and simulating anthropogenic climate change and our impact on Earth's ecosystems (Feddema et al., 2005). Land cover maps are usually created by automatic classification of satellite imagery, a process that results in discrepancies between global land cover products (Fritz et al., 2011; McCallum, Obersteiner, Nilsson, & Shvidenko, 2006). To aid in solving discrepancies, citizen science is a high value resource. For example, the International Institute for Applied Systems Analysis (IIASA) has created Geo-Wiki, an information portal that allows volunteers to improve data on land cover using satellite imagery from Google Earth (Fritz et al., 2017; See et al., 2015). Results show volunteer-contributed data to be generally equivalent to expert data (See et al., 2013).

Satellite imagery interpretation frequently depends on the existence of images taken in-situ, requiring researchers to seek alternative sources of visual data to build a robust model of the environment. Estima and Painho (2013) explored the adequacy of Flickr images to help the quality control of the CORINE land cover (CLC). They concluded there is potential for such use, admitting however that their study has not looked into the content of the images and their adequacy for this purpose. Nevertheless, ElQadi et al. (2017) demonstrated that since Flickr contains such a massive amount of photos, even if only a small fraction are relevant, there is potential to achieve high value outcomes from open data on social network sites (SNS) by careful filtering, which avoids some of the issues associated with raw image data (Barve, 2014).

Social network sites (SNS) geo-tagged photographs were used in previous research to determine land cover classes. For instance, Estima, Fonte, and Painho (2014) compared CORINE Land Cover information obtained from Flickr geo-tagged photos against classification based on satellite imagery. They concluded that the SNS geotagged photos are a valuable supplementary data source. Oba, Hirota, Chbeir, Ishikawa, and Yokoyama (2014) retrieved photos from Flickr using text tags corresponding to land cover classes. They then classified regions to land cover types based on image feature classification using a support vector machine (SVM), as well as photos' titles and tags. Xu, Zhu, Fu, Dong, and Xiao (2017) and Xing, Meng, Wang, Fan, and Hou (2018) used Convolutional Neural Networks (CNN) to classify geotagged photos from the Global Geo-Referenced Field Photo Library and Flickr into land cover classes.

Our work seeks to support remote sensing scientists' decision making on land cover type, rather than to automate the decision process. This distinction is made primarily because our research problem stems from discrepancies in land cover satellite imagery.

Consequently, our work provides a practical framework to solve existing problems in land cover classification in understudied regions, rather than suggesting theory to be validated in well-studied regions as was the focus of previous work.

To achieve our objectives, we developed a framework that uses computer vision to describe image visual content, and an artificial neural network classification model to decide whether or not the image is relevant based on this description.

The main outcome of our study is a reusable framework to find and filter imagery that can help determine land cover types. Therefore, we investigate whether particular countries should have customised models, or whether data from all countries can be used to build one generalised model. To build the suggested framework, human labour is required to label images for training the machine learning algorithm. It is therefore worthwhile testing whether models can be reused in countries other than those from where the training data originates. We developed country-based and generalised models for Africa, and compared the results to establish the most appropriate ways to use them. Our framework was integrated into Geo-wiki.

Although our framework is independent of the social network site data source, we chose Flickr for its favourable Application Programming Interface (API) since it accepts queries simultaneously filtered spatially, temporally, and textually. We used the Python Flickr API (Mignon, 2016).

In the next sections we present our methodology, and discuss the results of the different models we developed. Finally, we present our publicly available API allowing other users to invoke our models.

2 Methodology

2.1 Study area

The chosen study area consists of four African countries with varied geographic, demographic, and economic characteristics: Egypt, Kenya, Zambia, and Côte d'Ivoire (Figure 1). These are all subjects of a practical interest in obtaining land cover photos because although they fall within regions where high resolution data is available, it has previously been noted (Lesiv et al., 2017) that the accuracy of this data is as low as 65%. Also, these four countries offer valuable case studies due to their variations in climate and geography.

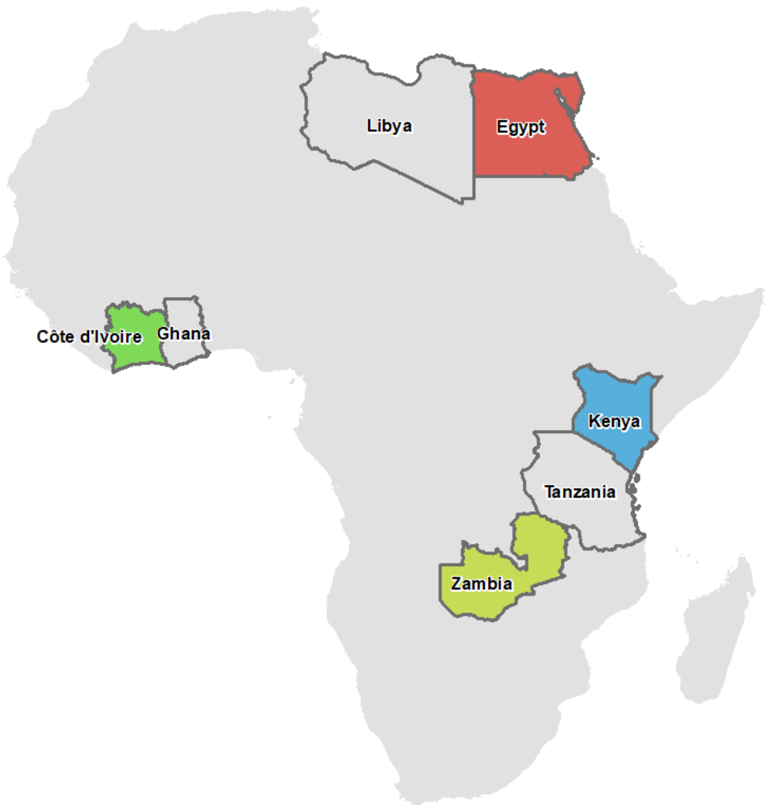

Figure 1 Four African countries with varied climate and geography are used as case studies (Egypt, Kenya, Côte d'Ivoire, and Zambia). Photo classification models from these countries were scored against data from 3 countries neighbouring the original test cases, shown in grey (Libya, Ghana, and Tanzania). Map data from gadm.org

2.2 Process overview

Our process uses cloud-based computer vision services to analyse the visual content of geotagged photos on social network sites and generate descriptive tags for that content. We then use these tags to train an artificial neural network to predict a photo's adequacy for land cover classification. Figure 2 demonstrates the workflow. First, we collect photos in the specified study area. We then filter the collected photos to exclude those in urban regions, since built-up land cover is already well-known and mapped, while we are primarily interested in natural and other non-urban regions.

We then select a random sample of photos from the filtered data set that we subject to automatic computer vision classification programs to generate descriptive text tags for the samples. The sample set of photos is also manually labelled by expert researchers as either relevant or irrelevant based on the photo utility to land cover determination (e.g. outdoor, outside settlement areas, containing trees, shrubs, grassland, rocks, sand, or agricultural areas, etc.). The text tags and associated relevant/irrelevant Boolean labels are used to build a classification model that can predict whether a photo described by a certain set of tags is relevant for land cover classification. The model is later applied to filter photos based on visual content, by users classifying land cover (Figure 3).

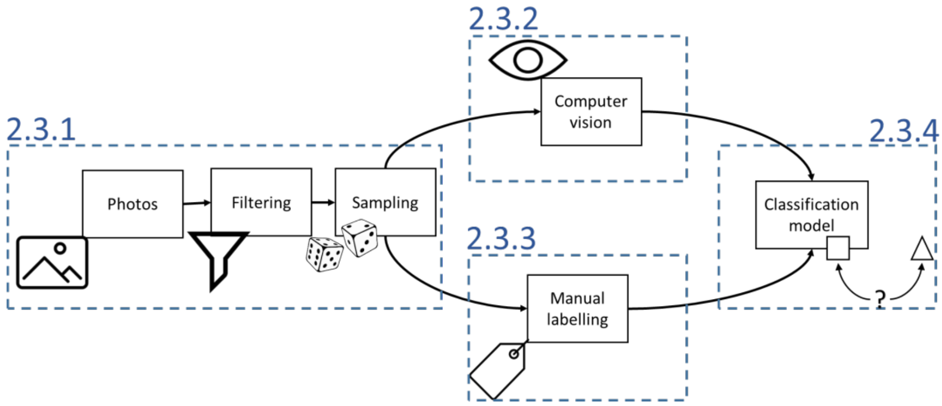

Figure 2 Methodology overview: Building a classification model to predict a photo's relevance to land cover determination. Numbers (e.g. 2.3.1) refer to text sections.

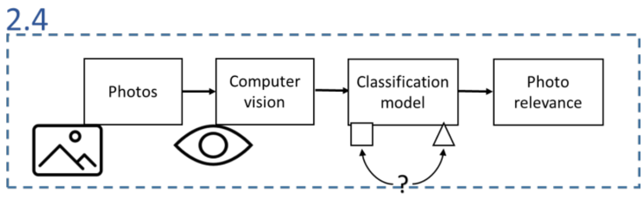

Figure 3 Our reusable classification models.

2.3 Building the classification models

2.3.1 Photo collection

The method we describe can be applied to any social network site that offers an Application Programming Interface (API) to retrieve geotagged photos. In addition to the convenience of Flickr's API already noted, we chose Flickr (www.flickr.com) for its wealth of photos; with at one stage about 2 million new photos being added every day (Michel, 2016).

Flickr can be queried for photos in a particular geographic bounding box. However, for any given query, the API only returns 3600 unique results. In order to collect a larger number of photos, we divided each of our test case countries into a grid, where its cells are then used sequentially as the bounding box for photo queries. To create the grid, the outline of a country is divided into cells of side length of 1 decimal degree. The grid cells intersecting the urban centres are replaced by a smaller grid of a cell side length of 0.2 decimal degrees to allow for high-density retrieval of photos. We then delete the grid cells that are totally contained within urban areas, since we are only interested in land cover in non-urban areas. Grid cells containing smaller urban settlements, i.e. cells that are not totally covered by urban settlements, are used to query Flickr. The full set of collected photos in a country is filtered using an urban settlement mask based on the Copernicus land cover dataset (© European Union, Copernicus Land Monitoring Service 2018, European Environment Agency (EEA)), leaving only photos outside registered urban settlements. A sample of the remaining photos is randomly selected for the following steps of visual tagging and data labelling (Table 3).

Although the Flickr API allows searching by location and keyword simultaneously, we didn't limit our queries by any keywords and only searched by geographic bounding box because the text tags supplied by the original image posters can sometimes be misleading (ElQadi et al., 2017; Xing et al., 2018).

2.3.2 Computer vision tagging

In this step, we use computer vision to assign tags to the photos based on their visual content. To achieve this, we used the computer vision cognitive services from Microsoft Azure. Microsoft Azure (https://azure.microsoft.com) is a set of cloud-based services, which include "cognitive services" (Microsoft, 2018); machine learning models, and artificial intelligence algorithms. Among these, computer vision is one of the available cognitive services.

In the cognitive services API, the "Image" function can be applied for a given image where more than one tag is returned for the image, with varied confidence levels, with unity being the highest confidence level.

We ran the "Analyse Image" service on every image in the selected sample and only chose the tags with confidence levels higher than 0.5. This is an arbitrary value that corresponds to the computer vision system being more confident than not about the generated tag. A lower threshold would have simply given more tags in which the computer vision system had poor confidence. We excluded images that had no tags with enough confidence from the machine learning model training. For each country, we compiled a list of all tag occurrences for all photos in that country. To build the classification models, we first order tags returned from our set of images in order of decreasing frequency. Next, starting with tags that are most frequent, we selected tags for inclusion in our classifier, stopping when we have the set of tags present in 95% of the records. These selected tags served as our classification model features. If we were to consider the minimum set of tags occurring in a larger number of records, say 99%, we would have a much-larger set of tags (Table 1). But many of these tags would only be present in very few photos. This, consequently, would result in a higher number of classification model features, many of which would carry very little information.

Table 1 The minimum set of tags occurring in 99% of records is significantly higher than the set occurring in 95% of the records.

		Percentage of records covered by selected tags	
		95%	99%
Number of text tags selected	Egypt	19	136
	Kenya	10	28
	Zambia	12	33
	Côte d'Ivoire	13	107

Preliminary analysis in this step showed that the frequent tags are different in each country, despite the fact that photos in all cases were outside urban centres. For example, in Egypt where people primarily visit to see monuments and practice water sports, the tags include: outdoor, sky, nature, person, water, building, ground, indoor, etc. (Figure 4). Conversely, in Kenya, where safari is prominent, the tags include outdoor, grass, animal, field, sky, mammal, tree, etc. (Figure 5).

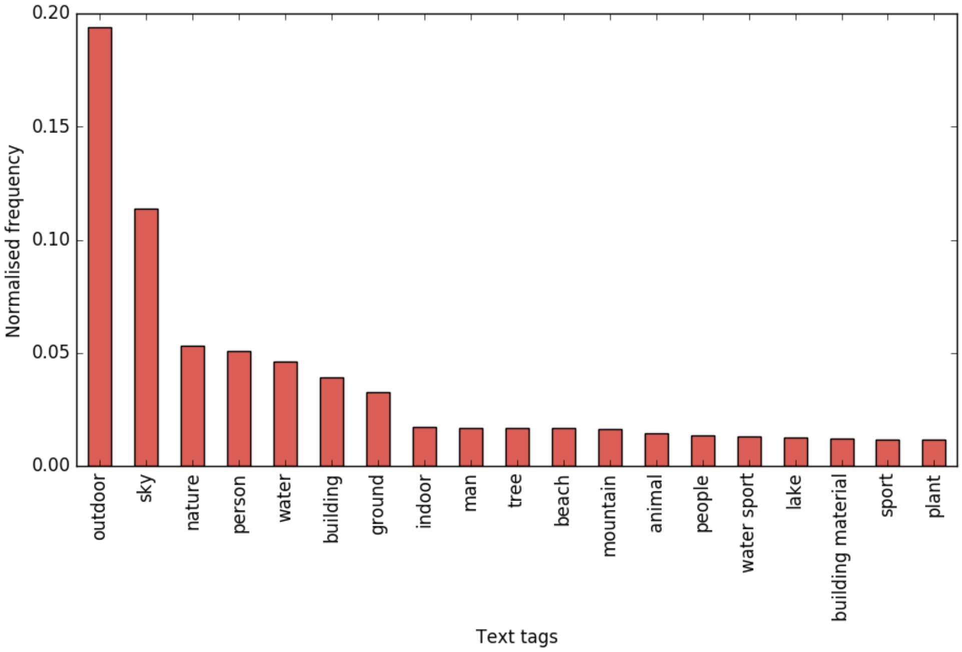

Figure 4 Normalised frequency of text tags associated with a (2%) sample of Flickr photos from Egypt outside urban areas. The tags common among 95% of the data set were selected as classification features

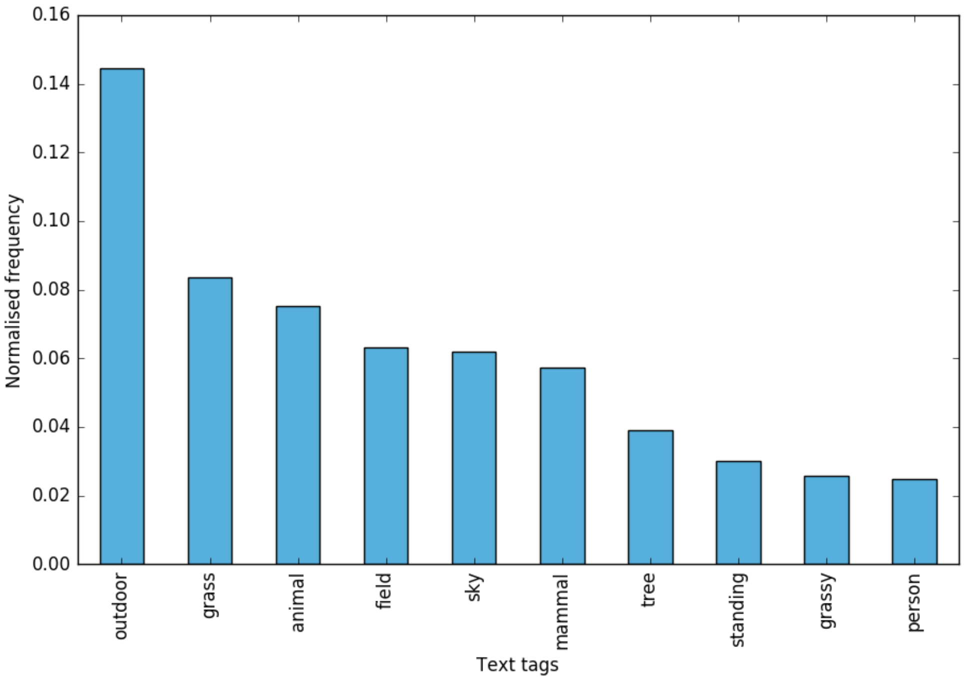

Figure 5 Normalised frequency of text tags associated with a (2%) sample of Flickr photos from Kenya outside urban areas. The tags common among 95% of the data set were selected as classification features.

2.3.3 Manual data labelling

The sample set of photos that have been automatically tagged based on their visual content are then manually labelled by expert researchers from IIASA as relevant if they perceived the image content presented potentially meaningful land cover information. Otherwise, the images are labelled as irrelevant. The relevance ratio is different among countries as shown in Table 4. That these ratios differ among countries may be attributed to the different photographic activity profile of each country as discussed in 2.3.2.

2.3.4 Classification model training

Artificial neural networks (ANNs), were chosen to build the classification model. ANNs are capable of classifying patterns unseen in training data. They are tolerant to noisy data, and appropriate when little is known about the relationships in input data (Han, Pei, & Kamber, 2012). And they are ubiquitous in the remote sensing community (e.g. (Xing et al., 2018))

In order to train an ANN that can predict whether a photo is relevant based on its visual tags, we created a table containing the label we applied (i.e. whether the photo is relevant or not). Next, the tags identified as selection features in 2.3.2 were added to the table as binary columns showing whether each tag was present in each row (i.e. photo). See Table 2.

Table 2 Example rows from the Kenya data set. Every row corresponds to a photo, and represents the presence of the text tags selected in (section 2.3.2) in that photo. The first column is the manual label assigned to indicate whether a photo is relevant to land cover (section 2.3.3). The classification model is trained to predict the target based on the feature columns (text tags).

Model target	Model features (Text tags)									
relevant	standing	outdoor	grassy	person	Tree	sky	field	animal	mammal	grass
false	false	true	false	false	true	true	false	false	false	false
false	false	true	false	true	false	true	false	false	false	false
true	false	true	false	false	false	false	false	True	true	true
…	…	…	…	…	…	…	…	…	…	…

To build our classification models, we used fully connected neural networks with one hidden layer consisting of n neurons, where n $=\frac{Training\ set\ size}{2*(Input\ features + output)}$. The number of neurons in the hidden layer is based on experimentation with the rules of thumb suggested in (Heaton, 2008). Our ANNs had a 0.1 learning rate, 0.1 initial learning weight, and a Min-Max normaliser. We randomly selected 70% of the data set to train the model, and the remaining 30% for scoring. The machine learning experiment was fully implemented in Microsoft's Azure machine learning studio, a browser-based graphical user interface that runs machine learning algorithms and models in the cloud.

2.4 Using the classification models

The driver behind this work is the practical need for geotagged photos to help determine land cover in understudied areas. Geo-wiki users are presented with a web interface. This follows the workflow shown in Figure 3, for a given location viewed on the map, images are retrieved from SNS. Next, tags are generated for all photos using the computer vision API. The tags are then sent to our Machine Learning model that responds with a Boolean value whether the photo is relevant to land cover determination. Only relevant photos are displayed to the user.

It is worth noting that in the image-retrieval step, the user can choose to retrieve images from SNS based on associated text tags, which are proved to be valuable in predicting ecological features (Jeawak, Jones, & Schockaert, 2019). However, our system may be used in under-studied regions, as described in the case studies of this paper. In such regions, no a priori assumptions are made about the coverage, and different languages may be used in text tags associated with the photos. Hence, the user may opt to retrieve images from SNS based on location only, forgoing the text tags, and depend on our filtration tool to find the relevant photos. Consequently, we trained our filtration models based on visual content of images from study regions, without considering associated text tags. The user retains the freedom to query SNS images using text tags. The retrieved images in all cases are subject to our filtration models.

2.5 Experiments

An overview of these experiments is outlined below and in figure 7.

1. For each country in set 1 (Egypt, Kenya, Côte d'Ivoire, and Zambia), we collected images (2.3.1), tagged them (2.3.2), labelled them (2.3.3) and trained a model (2.3.4).
2. A union of data, drawn from across set 1 to ensure a spread of representation from each country, was used to train and test a single generalised model through the process outlined in (2.3.1 to 2.3.4).
3. For each country in set 2 (Libya, Ghana, and Tanzania), which neighbour set 1 countries, we collected images (2.3.1), tagged them (2.3.2), labelled them (2.3.3) but did not train a model.

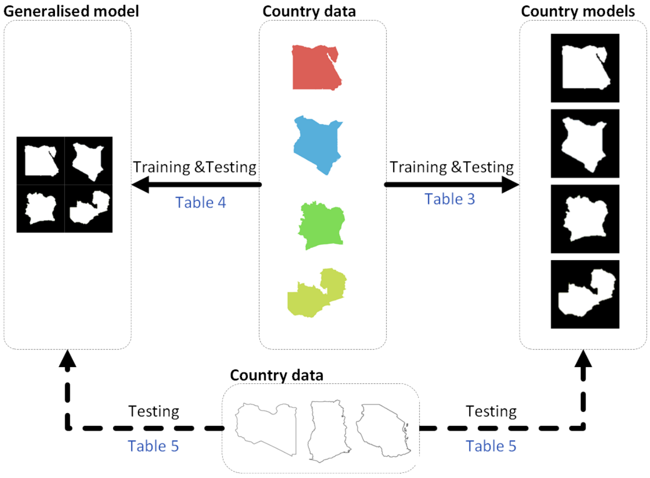

Figure 6 Data from four countries (shown in Figure 1) were used to train and test four corresponding country models. A union of data, drawn from across set 1 to ensure a spread of representation from each of the same four countries, was used to train and test a single generalised model. Data from three neighbouring countries (shown in grey in Figure 1) were tested against the country-specific models and the generalised model.

The first experiment enabled us to test our models in the countries from which the data were obtained. The second experiment tests the performance of a unified, generalised model. The third experiment enabled us to test model reusability. I.e., can a model from one country filter or predict the relevance of data from another?

3 Results and discussion

3.1 Experiment 1: Country-specific models

The results of the data collection process described in 2.3.1 are summarised in Table 3. Here we report "Total number of available photos" which is the count of records reported by the Flickr web interface when searching for geotagged images in a country. "Total collected" is the number of photos collected outside large urban areas using the grid cells we calculated in 2.3.1. The API limits the number of downloaded photos in grid cells with high photos density. These would mostly be urban areas. "Non-urban" is the count of remaining records after masking out urban areas. "Sample" is the count of records randomly selected for automatic visual tagging and manual labelling, the sample size is about 2% in Egypt and Kenya. However, since fewer data are available in Côte d'Ivoire and Zambia, the sample size was taken to be 50% and 10% respectively to allow for a representative sample from these relatively small data sets that was sufficiently large to allow for training and scoring.

Table 3 Summary of data collection activities for the four case study countries.

Flickr photos	**Egypt**	**Kenya**	**Côte d'Ivoire**	**Zambia**
Total number of available photos	580,884	261,382	8,347	38,970
Total collected	103,335	80,617	3,536	21,821
Non-Urban	50,634	69,021	1,098	11,272
Sample	985	1,273	545	1,054

The photos in the sample were tagged and labelled as described in sections 2.3.2 and 2.3.3. Some of the photos couldn't be tagged or labelled with sufficient confidence and so were removed from the analysis; hence the difference between dataset size in Table 4 and the sample size in Table 3.

The ratio between the two classification classes (relevant and irrelevant) are shown in Table 4. The table also reports on the configuration of our ANN classification model (section 2.3.4), and the performance metrics for classifying the scoring dataset.

Table 4 Configuration and performance metrics for the trained neural networks. TP is True Positive, TN is True Negative, FP is False Positive, and FN is False Negative.

		Formula	Egypt	Kenya	Côte d'Ivoire	Zambia
Metadata	Dataset size (tagged & labelled)	-	897	960	481	840
	Relevant: Total records	$\frac{Relevant\ record\ count}{Total\ record\ count}$	0.3	0.7	0.4	0.6
Configuration	No. input features	-	19	10	13	12
	No. neurons	$\frac{Training\ set\ size}{2*(Input\ features\ +output)}$	16	31	12	23
Metrics	Specifity, selectivity	$\frac{TN}{TN+FP}$	0.83	0.72	0.90	0.79
	Accuracy, Recognition rate	$\frac{TP+TN}{TP+TN+FP+FN}$	0.84	0.83	0.84	0.85

	Precision	$\frac{TP}{TP+FP}$	0.67	0.86	0.83	0.89
	Recall, Sensitivity	$\frac{TP}{TP+FN}$	0.86	0.89	0.76	0.88
	F1 score	$2*\frac{Precision\ *\text{Recall}}{Precision+Recall}$	0.75	0.88	0.79	0.89

3.2 Experiment 2: A generalised model

Although the distribution of tags differs by country as we discussed in 2.3.2, we checked whether there is sufficient value in spending resources on creating separate models. We thus collated records from every dataset to train a generalised model with records from each country. The number of columns, i.e. selection features, was the superset of all datasets. 70% of the data was used to train the generalised model. The validation dataset from each country was scored against the generalised model, as well as the original (country-specific) model.

Results shown in Table 5 demonstrate that differences in model accuracy between per-country models and the generalised model are statistically significant when the per-country models perform better than the generalised model (Egypt and Côte d'Ivoire). In cases where the generalised model slightly outperformed the per-country models (Kenya and Zambia), the difference was statistically insignificant.

Table 5 For each country's test dataset: accuracy of individual and generalised models are compared. Also, the dataset results from both models are compared for statistical significance using McNemar's test.

		Egypt	**Kenya**	**Côte d'Ivoire**	**Zambia**
Model Accuracy	Individual	0.84	0.83	0.84	0.85
	Generalised	0.71	0.85	0.71	0.87
McNemar`s test	p-value	0.01	0.13	0.004	0.5
	95% C.I.	Significant	Insignificant	Significant	Insignificant

3.3 Experiment 3: Set 2 countries

Here we address the question of model reusability: could we use a model from one country to filter (predict relevance of) data from another country? And, given that country-specific models are better than the generalised model, how would the generalised model perform against data from countries not used in model training?

To answer these questions, we collected new test data from a second set of countries: Libya, sharing a border with Egypt; Tanzania, sharing borders with both Kenya and Zambia; and Ghana, sharing a border with Côte d'Ivoire. Samples from these data were manually labelled (156 from Libya, 165 from Tanzania, and 162 from Ghana), then scored using all individual country models, and the generalised model. Results are shown in Table 6. The generalised model's performance with data from a new country is the best, or at least as good as, the model from a neighbouring country.

Table 6 Comparing model accuracy for data from countries not used to build the models. Models from countries sharing a border with the data country in bold.

		Model (Set 1 countries)				
		Egypt	Kenya	Côte d'Ivoire	Zambia	Generalised
Data (Set 2 Countries)	Libya	**0.77**	0.74	0.81	0.50	0.78
	Tanzania	0.55	**0.81**	0.74	**0.81**	0.81
	Ghana	0.73	0.79	**0.80**	0.82	0.83

3.4 Related costs and technology

The machine learning logic was implemented in the Azure machine learning cloud service from Microsoft which not only masked many of the low-level details of the neural network that are of little interest to our research, but also helped in the automatic generation of consumable web services allowing Geo-wiki users to benefit from our models as a service. In Geo-wiki, photos are retrieved from social network sources using a search module. Then, every photo is subjected to the process shown in Figure 3. The time taken to process each photo depends on multiple factors such as the image resolution, the speed of the computer vision API, the classification model speed, the throughput of the web application host server and the bandwidth between these components. We empirically measured the mean processing time of computer vision tagging to be 1.4 seconds/image (sd=0.4, n=100), and the mean processing time of our classification model to be 0.24 seconds/image (sd=0.12, n=100). In our experiments, we opted for the highest available resolution of the Flickr photos used. However, the overall processing speed might be enhanced through optimising between resolution and computer vision accuracy for specific user's requirements.

Our framework weeds out the majority of irrelevant photos (accuracy around 0.8) from the massive amounts of SNS photos. Noting that the ratio of relevant photos in samples we checked was between 0.3 and 0.7 (Egypt and Kenya respectively), our framework is saving researchers valuable time they would otherwise spend browsing many irrelevant photos.

We have opted to include text tags in a descending order of frequency (2.3.2), until we had included tags appearing in at least 95% of the records (Table 1). However, the accuracy of our classification filter can be further enhanced by optimising the number of input features (text tags) participating in the model. Users may elect to explore the possibility of better training their classifier to give more accurate assessments, however this will incur a time cost, it requires more data, it may result in issues related to overtraining, and in fact, it may not save the users sufficient amounts of human time to be worthwhile.

Microsoft cognitive services computer vision is implemented using deep learning (Tran et al., 2016). These are often mis-calibrated (i.e. overly confident in their own results) (Guo, Pleiss, Sun, & Weinberger, 2017). Although Microsoft cognitive services computer vision contains a dedicated module for confidence estimation (Tran et al., 2016), we arbitrarily chose to consider tags with confidence>0.5. Empirical calibration of this confidence cut-off may prove useful to the overall framework performance in the future.

4 Conclusions

Social network sites provide a wealth of data for researchers in many disciplines. Data quality is a primary concern in leveraging this rich data source. Photos on SNS are not always fit for research purposes. In land cover mapping, SNS can provide much needed high-resolution geotagged photos which can be manually, or automatically, classified to determine the type of land cover in a photo. However, to retrieve and classify geotagged photos from SNS, filtration methods are needed to remove irrelevant photos from the classification pipeline. In this work, we suggested, developed, and tested a framework to filter SNS photos relevant to land cover classification.

Our framework uses commercially available APIs to favour simple implementation by interdisciplinary researchers who may need to reproduce it. Technology is improving over time, we are sure the methods and frameworks we suggest in this research would be improved further when paired with more capable computer vision and machine learning in the future.

5 Acknowledgements

The project was part of the IIASA Young Scientist Summer Program (YSSP), the participation of the first author in this program was funded by Potsdam Institute for Climate Impact Research. The experiments used APIs and Services from Microsoft Azure, supported by a *Microsoft AI for Earth* grant. The research was also supported by the Australian Research Council grant DP 160100161. The authors would also like to thank Christoph Perger for his help at YSSP with Geo-wiki development.

References

Barve, V. (2014). Discovering and developing primary biodiversity data from social networking sites: A novel approach. *Ecological Informatics, 24*, 194-199.

Cheong, M. (2018). *Experimental investigations of #authenticity online*. Paper presented at the AAP NZAP Conference, Victoria University, Wellington, New Zealand.

Culotta, A. (2010). *Towards detecting influenza epidemics by analyzing Twitter messages.* Paper presented at the Proceedings of the first workshop on social media analytics.

Daume, S. (2016). Mining Twitter to monitor invasive alien species—An analytical framework and sample information topologies. *Ecological Informatics, 31*, 70-82.

ElQadi, M. M., Dorin, A., Dyer, A., Burd, M., Bukovac, Z., & Shrestha, M. (2017). Mapping species distributions with social media geo-tagged images: Case studies of bees and flowering plants in Australia. *Ecological Informatics, 39*, 23-31. doi:http://dx.doi.org/10.1016/j.ecoinf.2017.02.006

Estima, J., Fonte, C. C., & Painho, M. (2014, 3-6 June). *Comparative study of Land Use/Cover classification using Flickr photos, satellite imagery and Corine Land Cover database.* Paper presented at the AGILE'2014 International Conference on Geographic Information Science,, Castellón.

Estima, J., & Painho, M. (2013). Flickr geotagged and publicly available photos: Preliminary study of its adequacy for helping quality control of corine land cover. In *Computational Science and Its Applications–ICCSA 2013* (pp. 205-220): Springer.

Feddema, J. J., Oleson, K. W., Bonan, G. B., Mearns, L. O., Buja, L. E., Meehl, G. A., & Washington, W. M. (2005). The importance of land-cover change in simulating future climates. *Science (New York, N.Y.), 310*(5754), 1674. doi:10.1126/science.1118160

Fritz, S., See, L., McCallum, I., Schill, C., Obersteiner, M., Van der Velde, M., . . . Achard, F. (2011). Highlighting continued uncertainty in global land cover maps for the user community. *Environmental Research Letters, 6*(4), 044005.

Fritz, S., See, L., Perger, C., McCallum, I., Schill, C., Schepaschenko, D., . . . Obersteiner, M. (2017). A global dataset of crowdsourced land cover and land use reference data. *Scientific Data, 4*, 170075. doi:10.1038/sdata.2017.75

Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). *On calibration of modern neural networks.* Paper presented at the Proceedings of the 34th International Conference on Machine Learning-Volume 70.

Han, J., Pei, J., & Kamber, M. (2012). *Data mining: concepts and techniques* (Third ed.): Elsevier.

Heaton, J. (2008). *Introduction to neural networks with Java* (Second ed.): Heaton Research, Inc.

Jeawak, S. S., Jones, C. B., & Schockaert, S. (2019). *Embedding Geographic Locations for Modelling the Natural Environment Using Flickr Tags and Structured Data*, Cham.

Lesiv, M., Fritz, S., McCallum, I., Tsendbazar, N., Herold, M., Pekel, J.-F., . . . Obersteiner, M. (2017). *Evaluation of ESA CCI prototype land cover map at 20m*. Retrieved from IIASA, Laxenburg, Austria: http://pure.iiasa.ac.at/id/eprint/14979/

McCallum, I., Obersteiner, M., Nilsson, S., & Shvidenko, A. (2006). A spatial comparison of four satellite derived 1km global land cover datasets. *International Journal of Applied Earth Observation and Geoinformation, 8*(4), 246-255. doi:https://doi.org/10.1016/j.jag.2005.12.002

Michel, F. (2016). How many public photos are uploaded to Flickr every day, month, year? Retrieved from https://www.flickr.com/photos/franckmichel/6855169886/in/photostream/

Microsoft. (2018). Computer Vision API for Microsoft Cognitive Services. Retrieved from https://docs.microsoft.com/en-us/azure/cognitive-services/

Mignon, A. (2016). python-flickr-api [Python package]: github.com. Retrieved from https://github.com/alexis-mignon/python-flickr-api

O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010). From tweets to polls: Linking text sentiment to public opinion time series. *ICWSM, 11*(122-129), 1-2.

Oba, H., Hirota, M., Chbeir, R., Ishikawa, H., & Yokoyama, S. (2014). *Towards Better Land Cover Classification Using Geo-tagged Photographs.* Paper presented at the Multimedia (ISM), 2014 IEEE International Symposium on.

See, L., Comber, A., Salk, C., Fritz, S., van der Velde, M., Perger, C., . . . Obersteiner, M. (2013). Comparing the Quality of Crowdsourced Data Contributed by Expert and Non-Experts. *PloS one, 8*(7), e69958. doi:10.1371/journal.pone.0069958

See, L., Fritz, S., Perger, C., Schill, C., McCallum, I., Schepaschenko, D., . . . Obersteiner, M. (2015). Harnessing the power of volunteers, the internet and Google Earth to collect and validate global spatial information using Geo-Wiki. *Technological Forecasting and Social Change, 98*(Supplement C), 324-335. doi:https://doi.org/10.1016/j.techfore.2015.03.002

Tran, K., He, X., Zhang, L., Sun, J., Carapcea, C., Thrasher, C., . . . Sienkiewicz, C. (2016). *Rich image captioning in the wild.* Paper presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *ICWSM, 10*, 178-185.

Xing, H., Meng, Y., Wang, Z., Fan, K., & Hou, D. (2018). Exploring geo-tagged photos for land cover validation with deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing, 141*, 237-251. doi:https://doi.org/10.1016/j.isprsjprs.2018.04.025

Xu, G., Zhu, X., Fu, D., Dong, J., & Xiao, X. (2017). Automatic land cover classification of geo-tagged field photos by deep learning. *Environmental Modelling & Software, 91*, 127-134. doi:https://doi.org/10.1016/j.envsoft.2017.02.004