Integrating active learning and crowdsourcing into large-scale supervised landcover mapping algorithms

Stephanie R. Debats[a,*], Lyndon D. Estes[a], David R. Thompson[b], Kelly K. Caylor[a,c,d]

[a]*Department of Civil & Environmental Engineering, Princeton University, Princeton, NJ, USA*
[b]*NASA Jet Propulsion Laboratory, Pasadena, CA, USA*
[c]*Department of Geography, University of California, Santa Barbara, CA, USA*
[d]*Bren School of Environmental Science & Management, University of California, Santa Barbara, CA, USA*

Abstract

Sub-Saharan Africa and other developing regions of the world are dominated by smallholder farms, which are characterized by small, heterogeneous, and often indistinct field patterns. In previous work, we developed an algorithm for mapping both smallholder and commercial agricultural fields that includes efficient extraction of a vast set of simple, highly correlated, and interdependent features, followed by a random forest classifier. In this paper, we demonstrated how active learning can be incorporated in the algorithm to create smaller, more efficient training data sets, which reduced computational resources, minimized the need for humans to hand-label data, and boosted performance. We designed a patch-based uncertainty metric to drive the active learning framework, based on the regular grid of a crowdsourcing platform, and demonstrated how subject matter experts can be replaced with fleets of crowdsourcing workers. Our active learning algorithm achieved similar performance as an algorithm trained with randomly selected data, but with 62% less data samples.

Keywords: land cover, agriculture, Sub-Saharan Africa, computer vision, machine learning, active learning

---

*Corresponding author
Email address: stephaniedebats@gmail.com (Stephanie R. Debats)

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3004v1 | CC BY 4.0 Open Access | rec: 6 Jun 2017, publ: 6 Jun 2017

1. Introduction

Supervised machine learning algorithms enable fast, repeatable landcover classification that can generalize to large geographic areas (Debats et al., 2016; Khatami et al., 2016; Lu & Weng, 2007). Algorithm performance depends on the selection and labeling of training data by human experts from which a supervised classifier can learn (Li & Sethi, 2006). Efficient training data sets of only the most informative samples limit the expense and time of training data preparation, whether through visual interpretation of satellite imagery or gathering groundtruth samples in field campaigns (Chi & Bruzzone, 2005; Li et al., 2010; Li & Sethi, 2006; Tuia et al., 2011a,b). Compact data sets also increase computational efficiency (Shi et al., 2016) and generalization performance (Campbell et al., 2000; Cohn et al., 1994; Crawford et al., 2013). This paper explores active learning and crowdsourcing, which are two methods that have been individually employed to optimize training data sets, though their joint contribution to supervised classification is just beginning to be explored.

Determining the number of required samples and how these samples are selected are two main considerations for supervised classification. Often, the number of samples may simply be limited by availability of data or manpower for labeling. Alternatively, simple heuristics, such as $30p$ samples, where p is the number of multi-spectral bands, aim to quantify the number of training samples required based on the dimensionality of remote sensing data (Foody et al., 2006; Van Niel et al., 2005). For selecting the specified number of samples, random sampling is most commonly used, but this method is sub-optimal when classes are imbalanced or rare, which is frequently the case in complex natural landscapes (Lewis & Catlett, 1994; Lewis & Gale, 1994). In high-resolution satellite imagery, landcover classes exhibit lower inter-class and higher intra-class spectral variability, suggesting random selection of training samples may not fully describe a class (Lu & Weng, 2007; Tokarczyk et al., 2013, 2015). In a recent study, human experts attempted to improve upon random selection by instead identifying a small subset of informative samples, based on their own judgement and knowledge of ancillary data, like soil type (Foody & Mathur, 2004; Foody

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3004v1 | CC BY 4.0 Open Access | rec: 6 Jun 2017, publ: 6 Jun 2017

et al., 2006).

Regardless of the selection mechanism, all of these methods are examples of passive learning, in which supervised classifiers passively receive training data from human experts (Li & Sethi, 2006). For remote sensing applications, active learning is showing promise as an alternative approach, in which an algorithm iteratively guides the selection of samples to produce efficient training data sets, creating a two-way flow of data between human experts and the algorithm (Crawford et al., 2013; Tuia et al., 2011a,b). An algorithm identifies the most informative samples in each iteration and queries a human expert for labels, which are then added to the training data set to retrain the algorithm, until the desired accuracy is achieved (Angluin, 1988; Baum, 1991; Cohn et al., 1994, 1996; Lewis & Catlett, 1994; Lewis & Gale, 1994; Li & Sethi, 2006; Plutowski & White, 1993). Active learning is based on the idea that a classifier trained on a set of carefully chosen examples will outperform one trained on a larger randomly-selected set, both in terms of accuracy and computational efficiency (Cohn et al., 1994, 1996; MacKay, 1992; Shi et al., 2016).

Using the quintuple notation of Li & Sethi (2006), an algorithm for active learning can be defined as follows:

- C: classifier
- L: labeled training data set
- U: pool of unlabeled samples
- Q: query function to select samples from unlabeled pool
- S: human supervisor who is capable of labeling samples

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3004v1 | CC BY 4.0 Open Access | rec: 6 Jun 2017, publ: 6 Jun 2017

Algorithm 1 Active learning, based on Li & Sethi (2006)

```
Randomly select small set of samples from U
Query S to label selected samples
Initialize L with labeled samples
Train C
while stopping criterion not satisfied do
    Select sample from U based on Q
    Query S to label selected sample
    Add new sample to L
    Retrain C
end while
```

The query function that an algorithm uses to select samples for labeling by a human expert distinguishes different active learning methods. Queries can leverage knowledge of how a particular classifier functions, such as weighting samples by their distance from a support vector machine's decision boundary (Campbell et al., 2000; Cheng & Shih, 2007). Alternatively, the output of probabilistic classifiers, such as the random forest, can be directly used in a query function as a measure of the classifier's confidence in assigning a label (Lewis & Catlett, 1994; Lewis & Gale, 1994; Li & Sethi, 2006).

Active learning generally assumes humans with expert knowledge of the problem will serve as supervisors. However, in this age of citizen science and volunteered geographic information, expert individuals are giving way to fleets of crowdsourcing workers (Goodchild, 2007). Humans excel at pattern recognition, even with occlusion or noise (Biederman, 1987), and are preferable to machines for certain tasks (i.e. CAPTCHA (Von Ahn et al., 2003)). In remote sensing, crowdsourcing has been used to assess disasters (Xie et al., 2016), create landcover maps (Estes et al., 2016; Salk et al., 2016; See et al., 2013b, 2015), and most recently, provide training data for supervised machine learning algorithms (Ofli et al., 2016).

Most crowdsourcing workers are non-experts, participating either for compensation or

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3004v1 | CC BY 4.0 Open Access | rec: 6 Jun 2017, publ: 6 Jun 2017

due to personal interest in the issue, who receive basic training for the required task (Estes et al., 2016; Salk et al., 2016; See et al., 2016). Tasks range from classification, which requires assigning a label to an image, to the more complex task of digitization, which involves creating a digital representation (i.e. delineation) of an identified object (Albuquerque et al., 2016). See et al. (2013a) found that experts and non-experts differed minimally in their ability to identify human impacts and landcover types, while Salk et al. (2016) demonstrated that contributors often improved with experience. Comber et al. (2015, 2016) found larger differences between cultural groups, who vary in their conceptualization and visual interpretation of landcover, than between experts and non-experts.

Quality control measures are critical for turning crowdsourcing results into accurate landcover maps. Typically, expert-validated data sets are used to judge the quality of crowdsourcing workers' results (Estes et al., 2016; Salk et al., 2016). Though multiple workers mapping the same area increases time and expense, worker agreement has been shown to be highly correlated with correct classification (Albuquerque et al., 2016) and multiple workers' digitizations can be combined to increase overall map accuracy (Estes et al., 2016). To facilitate serving images to workers, repeated mappings, and comparisons to the expert data sets, crowdsourcing systems typically divide the area of interest into various image patches using a regular survey grid (Estes et al., 2016; Jacobson et al., 2015). The gridded structure and output of crowdsourcing platforms create opportunities to improve active learning query functions. Samples in an active learning framework typically equate to individual pixels, yet point-wise labeling wastes the ability of humans to perceive high-order objects (Biederman, 1987; Henderson & Hollingworth, 1999) and leads to repeated pixel queries in the same geographic region (Stumpf et al., 2014).

Thus, crowdsourcing produces high-quality landcover classifications suitable for use as training data for supervised classifiers, while active learning guides the selection of training data samples. In this paper, we present an integrated framework using both crowdsourcing and active learning to train a supervised classifier, imposed on a regular grid of a typical

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3004v1 | CC BY 4.0 Open Access | rec: 6 Jun 2017, publ: 6 Jun 2017

crowdsourcing platform. Within this framework, we developed a patch-based uncertainty criterion for the active learning query function to interact with crowdsourcing workers. Finally, we present the results of a case study of agricultural field digitization in high-resolution, multi-spectral satellite imagery of South Africa, comparing the performance of our active learning system to a model using traditional random sampling.

2. Integrated framework

2.1. Supervised classifier

An active learning framework requires an algorithm that is fast to train and computationally inexpensive to enable repeated iterations (Lewis & Catlett, 1994). We utilized a random forest supervised pixel-wise classification algorithm, presented in Debats et al. (2016). To summarize, random forests (Breiman, 2001) are used extensively by the remote sensing community for their classification accuracy and speed, as well as their handling of nonlinear interactions and high-dimensional data sets (Belgiu & Drăguţ, 2016; Khatami et al., 2016; Pelletier et al., 2016). Random forests have also been shown to be particularly well-suited to agricultural mapping (Li et al., 2016). Though random forests sometimes struggle with low generalization performance when transferred to regions far from training areas (Belgiu & Drăguţ, 2016; Crawford et al., 2013; Juel et al., 2015; Pelletier et al., 2016; Vetrivel et al., 2015), the addition of spatial/textural features has been shown to ameliorate this issue as well as increase overall accuracy (Debats et al., 2016; Du et al., 2015; Khatami et al., 2016; Ursani et al., 2012). Our algorithm learned from hand-labeled field boundaries during training, extracting several thousand simple, highly correlated, and interdependent features, using an expanded version of Randomized Quasi-Exhaustive (RQE) features (Tokarczyk et al., 2013, 2015), which, in aggregate, are able to capture the subtle textural changes denoting the boundaries of smallholder fields (Debats et al., 2016).

The original algorithm was re-implemented for this paper as an open-source algorithm, using Python and Apache Spark. Apache Spark is an open-source framework for distributed

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3004v1 | CC BY 4.0 Open Access | rec: 6 Jun 2017, publ: 6 Jun 2017

computing with data parallelism and fault tolerance for large-scale data processing (Zaharia et al., 2010). Through its use of in-memory processing, Apache Spark is able to outperform other distributed computing frameworks, like Hadoop MapReduce, which continually reads and writes to disk. In a Spark application, a driver program coordinates all processes, connects to a cluster manager to allocate resources, and sends tasks to executors on the worker nodes. The Spark implementation of our algorithm enables faster training and easier scalability, which are necessary for the rapid, repeated iterations of active learning.

2.2. Crowdsourcing platform

To facilitate the rapid creation of groundtruth labels for training data, we based our framework on DIYlandcover, a crowdsourcing platform. DIYlandcover was built with open-source software to support the Mapping Africa project (`mappingafrica.princeton.edu`), which draws on workers to map the boundaries of agricultural fields in high-resolution satellite imagery. A full description of the DIYlandcover platform is included in Estes et al. (2016).

From a worker's perspective, they are connected to the DIYlandcover platform via a crowdsourcing marketplace, like Amazon Mechanical Turk. After passing a training module, a worker is presented with an image patch and utilizes tools in a mapping interface to draw polygons around agricultural fields (Figure 1). Upon completion of the task, the worker receives a small payment.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3004v1 | CC BY 4.0 Open Access | rec: 6 Jun 2017, publ: 6 Jun 2017

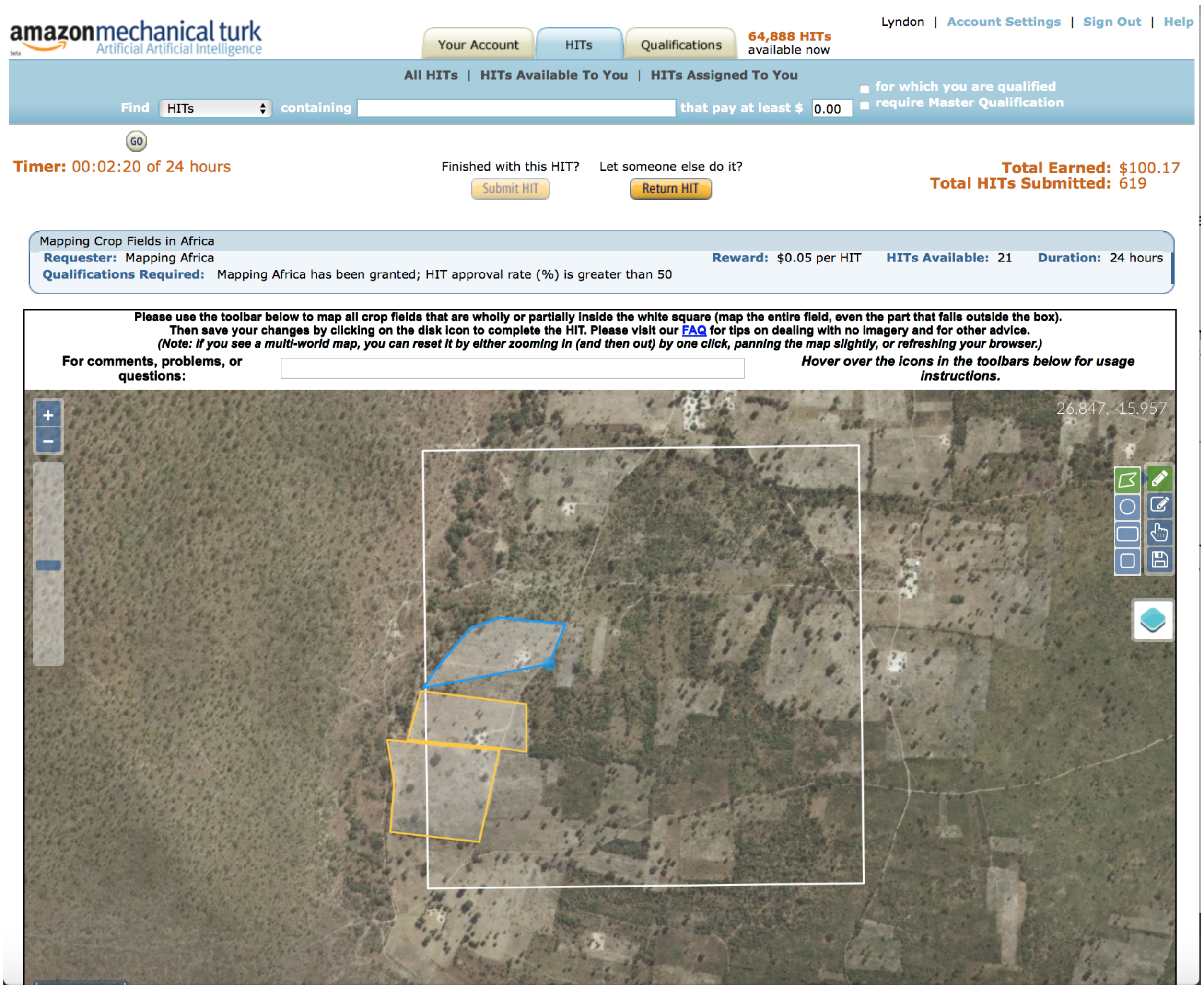

Figure 1: DIYlandcover interface for crowdsourcing agricultural field boundaries. A worker is served one image patch and uses the point-and-click tools provided to outline the boundaries of individual agricultural fields.

In the background, the DIYlandcover platform continually monitors worker mapping skill. This information is used to accept or reject mapping results from specific workers, pay out performance-based bonuses, and estimate overall map accuracy. DIYlandcover includes a main server hosting the platform's database, a Map API from which imagery is served, and a crowdsourcing marketplace where jobs are assigned to workers. The platform uses a regular survey grid to define image patches to be served to workers, including quality control sites, in order to frequently assess overall accuracy, compute worker-specific confidence scores, and enable repeated mappings of areas. In a recent production run, DIYlandcover achieved 91%

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3004v1 | CC BY 4.0 Open Access | rec: 6 Jun 2017, publ: 6 Jun 2017

accuracy in the digitization of agricultural fields in South Africa using novice workers. Based on this trial, it is estimated that 500 workers, each working 1 hour per day, could map the entire continent of Africa in 1.9 years at a cost of about $2 million (Estes et al., 2016).

2.3. Active learning

Introducing active learning into the DIYlandcover crowdsourcing platform would expand the definition of human supervisors from a handful of experts to a global pool of crowdsourcing workers. Active learning integrates with the crowdsourcing platform by replacing the current method of randomly sampling image patches weighted by the probability of landcover presence. Instead, selection of training samples is based on an uncertainty criterion calculated from the posterior probabilities produced by the current algorithm (Figure 2). DIYlandcover's accuracy assessments, including periodic checks of worker's performance on quality control sites, would run simultaneously to ensure the creation of high-quality training data for the algorithm to learn from. The global, on-demand nature of crowdsourcing works well with the iterative nature of active learning with intermittent periods of algorithm retraining and querying workers.

An uncertainty criterion, which is used in a query function to direct the selection of image patches from the unlabeled pool, is defined on a regular grid, in order to integrate with the crowdsourcing platform's survey grid and work simultaneously with worker quality control assessments. By basing the uncertainty criterion on a regular grid, areas of high uncertainty are prioritized, avoiding repeated queries in the same area as seen with pixel-based active learning queries (Stumpf et al., 2014). In addition, a query consisting of a single patch samples the higher intra-class variability present in high-resolution imagery and provides more information than a query of a single point, which is critical for large geographic areas of interest.

This criterion directly uses the probabilistic output of the random forest classifier to increasingly penalize pixels with more ambiguous classifications, and sums these penalties over image patches defined by the crowdsourcing platform's regular grid. Thus, the uncertainty

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3004v1 | CC BY 4.0 Open Access | rec: 6 Jun 2017, publ: 6 Jun 2017

criterion, Q, for an image patch, I, compares each pixel's posterior probability, $p(x,y)$, of belonging to a field, as determined by the current iteration's algorithm, to a value of 0.5, which denotes maximum uncertainty in the algorithm output, as follows:

$$Q(I) = 1 - \sum_{I(x,y)\in I} (p(x,y) - 0.5)^2 \tag{1}$$

At each iteration, the image patch with the maximum value of the uncertainty criterion is selected to be labeled and added to the training data set. The selected image is deemed to add the most additional information and increase the diversity of the training data set the most.

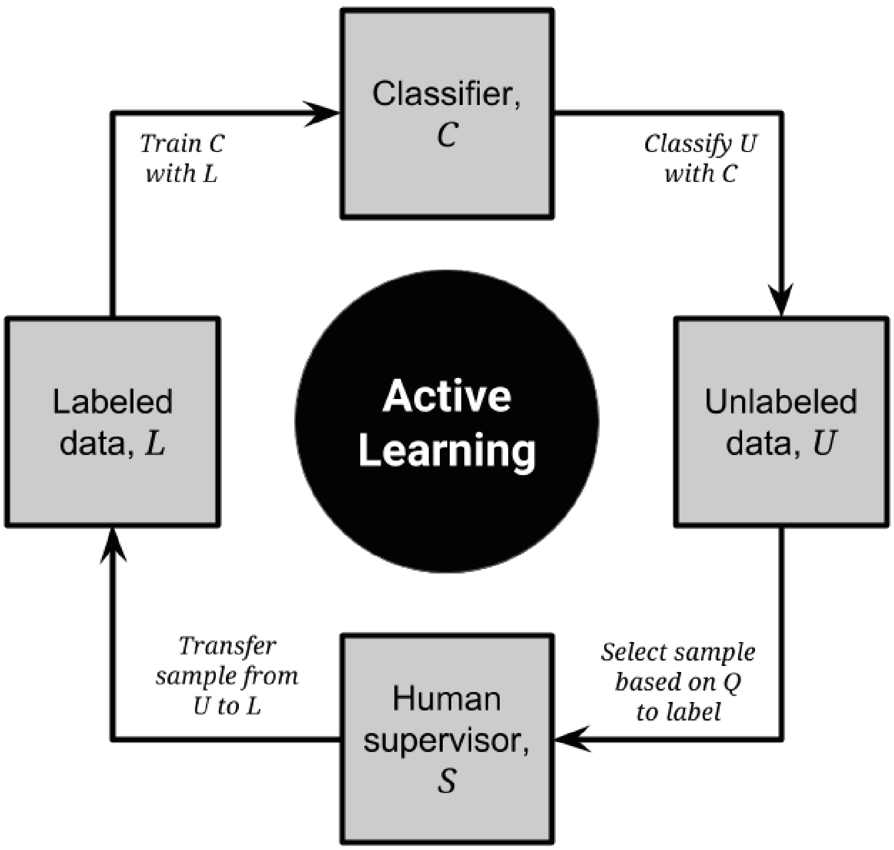

Figure 2: Active learning framework using quintuple terminology, adapted from (Li & Sethi, 2006) and (Wang & Zhai, 2016). During an iteration, an uncertainty criterion is calculated for each sample in the unlabeled data pool. The sample which the algorithm had the most difficulty in classifying (i.e. the most uncertain sample) is selected for labeling by a human supervisor and transferred to the labeled data pool. At this point, another iteration begins and the algorithm is retrained with the new, expanded labeled data pool.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3004v1 | CC BY 4.0 Open Access | rec: 6 Jun 2017, publ: 6 Jun 2017

3. Case study

We explored the use of the proposed active learning crowdsourcing platform in a case study of mapping agricultural fields in South Africa. The goal was to assess performance improvements as active learning was used to add new images to the training data set. To facilitate comprehensive accuracy assessments of the algorithm and repeated experiments, the entire area of interest was mapped offline, as opposed to in an on-demand fashion as the active learning algorithm ran.

The code for the case study was written in Python and Apache Spark and deployed on the Princeton University BigData cluster. The cluster is a SGI Hadoop Linux cluster, consisting of 6 data nodes and 4 service nodes, utilizing 2.80GHz Intel Xeon CPU E5-2680 v2 processors. The BigData cluster features a Hadoop Distributed File System (HDFS), which is a fault-tolerant file system that enables distributed storage of large files and rapid data transfer between compute nodes.

3.1. Study area & satellite imagery

This case study builds upon the work of Debats et al. (2016), including the use of the following satellite imagery data set. 8 study sites in South Africa capture a range of agricultural types, including commercial center pivot irrigated, commercial rainfed, and smallholder rainfed subsistence (Figure 3). Maize is the predominant crop across these sites, which is representative of Sub-Saharan Africa overall (Jones & Thornton, 2003). Each site is covered by a pair of DigitalGlobe Worldview-2 images, including a growing season image (December - April) and an off season image (July - November) within the same year or one year apart. The images are each 25 km^2, orthorectified, and aggregated using mean values to 2 m resolution. Spectral resolution spans one panchromatic band and 8 multi-spectral bands.

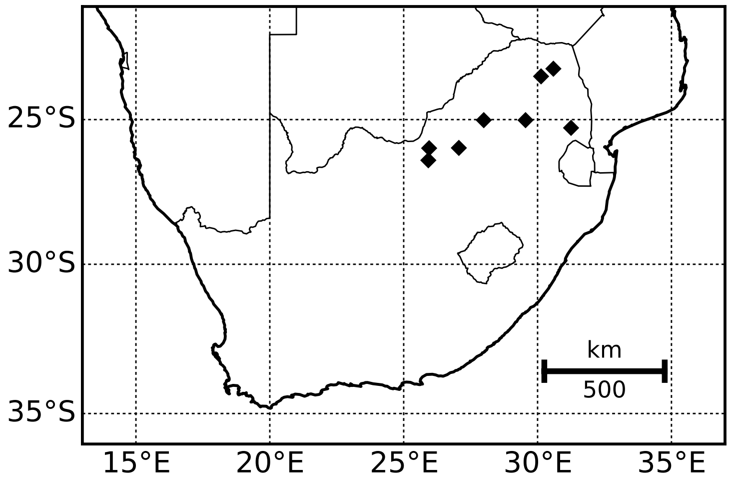

Figure 3: Worldview-2 imagery sites in South Africa (n=8). Figure reproduced from (Debats et al., 2016).

3.2. Methodology

Using an 8 x 8 grid, the satellite images of South Africa (8 images of 2400 x 2400 pixels at 2 m resolution) were each divided into image patches of 300 x 300 pixels, resulting in 512 image patches of sufficient size for workers to identify fields in a crowdsourcing platform. The 512 image patches was divided into the following pools:

- Unlabeled pool: 384 image patches were randomly selected for the unlabeled pool. At each iteration, one image was selected for labeling and transferred to the labeled pool.
- Labeled pool: The algorithm learned from the labeled images in this pool. At each iteration, one image was transferred to the labeled pool.
- Holdout pool: The remaining 128 image patches and their corresponding labels were completely separate from the training process. This pool was used to assess the algorithm's performance at each iteration on an independent data set.

Using these three pools, cross-validation was employed in the experiments to assess generalization performance. A 4-fold cross-validation scheme was selected to ensure a sufficient variety of images in each fold, given the data set size. Each fold's 128 image patches became the holdout pool and the remaining 384 image patches were assigned to the unlabeled pool.

At each iteration, one image patch was selected from the unlabeled pool. This image patch was matched with its labeled data and transferred to the labeled pool to be included in the training of the next iteration. Accuracy metrics were calculated at each iteration on the holdout set and averaged across folds to provide insights on the algorithm's generalization performance.

Our new implementation of the supervised pixel-wise classification algorithm was utilized for this case study. Initially, the classifier was trained with one randomly selected image patch. At each subsequent iteration, an additional image patch was selected to be labeled and added to the training data set, either based on (1) random selection, or (2) the highest scoring image patch according to the uncertainty criterion of the active learning framework. For each fold, the active learning experiment was run once, while the random selection experiment was run four times and averaged to account for the varying amounts of additional information in a randomly selected sample.

At each iteration, a performance metric, specifically the true skill statistic (TSS), was calculated on the holdout pool of images for both the random selection and active learning experiments. Unlike the more common kappa statistic, TSS is independent of prevalence in presence-absence studies (Allouche et al., 2006). The use of TSS is appropriate for this study, where non-field areas are more common than fields. For a binary classification, the true skill statistic is defined as:

$$\begin{aligned} TSS &= \text{sensitivity} + \text{specificity} - 1 \\ &= \frac{TP}{TP + FN} + \frac{TN}{FP + TN} - 1, \end{aligned} \tag{2}$$

where TP is true positives, TN is true negatives, FP is false positives, and FN is false negatives.

Learning curves were constructed for both the active learning and random selection experiments to assess the number of training samples required to achieve a desired performance.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3004v1 | CC BY 4.0 Open Access | rec: 6 Jun 2017, publ: 6 Jun 2017

In a learning curve, the performance metric (in this case, the TSS of the holdout set) was calculated at each iteration and plotted against the number of image patches currently in the training data set.

3.3. Results

Figure 4 compares the learning curves, using the TSS metric, of active learning using the patch-based based uncertainty criterion and the traditional approach of passive learning using random selection of training samples without regard to algorithm performance. At the end of the experiment when 45 image patches had been added to the training data set, active learning had achieved a TSS of 0.69, while random selection had achieved a TSS of 0.65. More importantly, if 0.65 is taken as a baseline, active learning required only 17 training samples to achieve the required accuracy, compared to random selection's 45 training samples, a 62% reduction.

Figure 5 provides a visual comparison between active learning and random selection for two sample images over 45 iterations. For both samples, active learning approaches a suitable classification with fewer iterations than random selection (15 versus 45). In addition, active learning has fewer false positives along roadways and more distinct boundaries between agricultural fields.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3004v1 | CC BY 4.0 Open Access | rec: 6 Jun 2017, publ: 6 Jun 2017

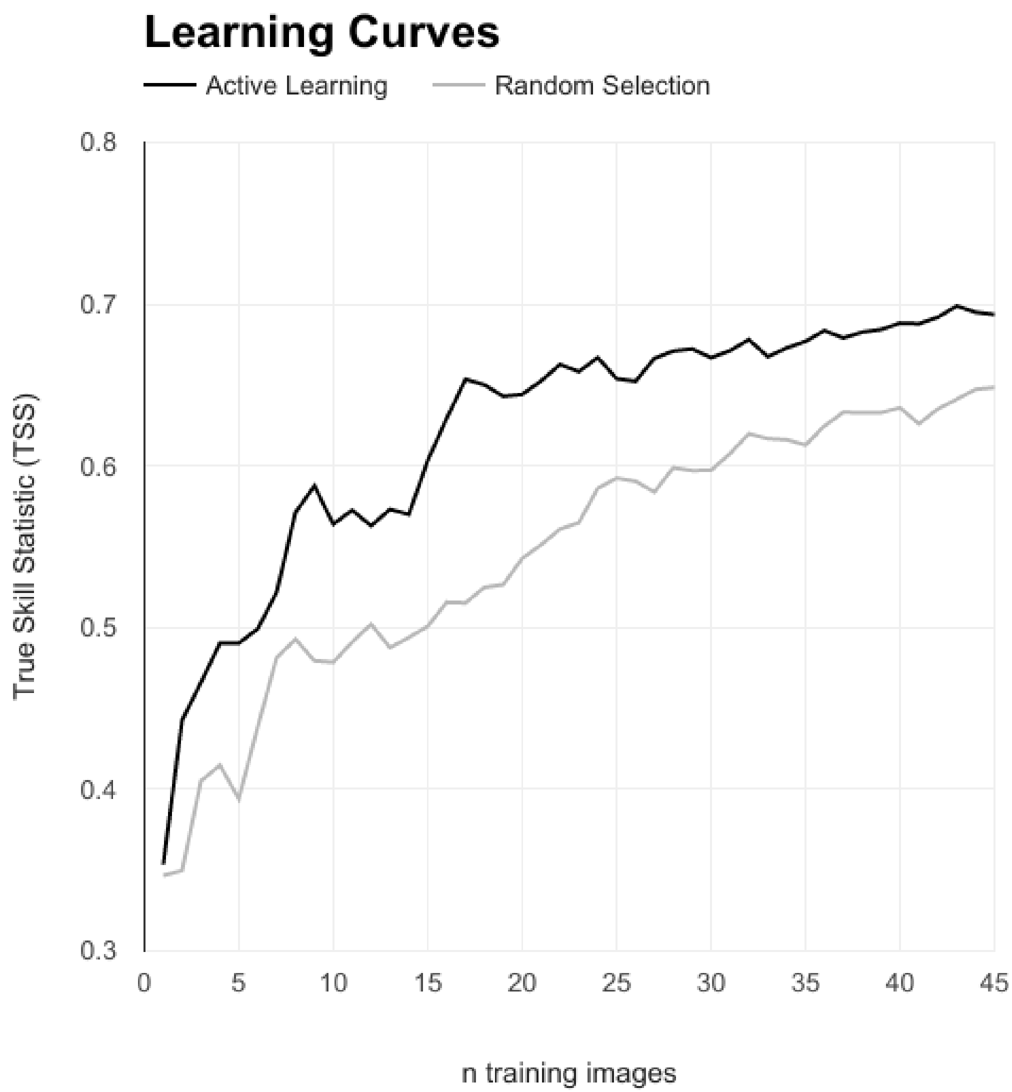

Figure 4: Learning curves constructed for active learning and random selection. Within 45 iterations, the algorithm trained with randomly selected data achieved a peak True Skill Statistic (TSS) of 0.65, while the algorithm trained through active learning reached a TSS of 0.69. The learning curves highlight that active learning was able to achieve the same level of performance as random selection, but with 17 training samples instead of 45.

4. Discussion and conclusions

In this paper, we presented an integrated framework that joins crowdsourcing and active learning with a supervised classification algorithm for large-scale landcover mapping. Crowdsourcing is increasingly recognized as a legitimate means of collecting high-quality

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3004v1 | CC BY 4.0 Open Access | rec: 6 Jun 2017, publ: 6 Jun 2017

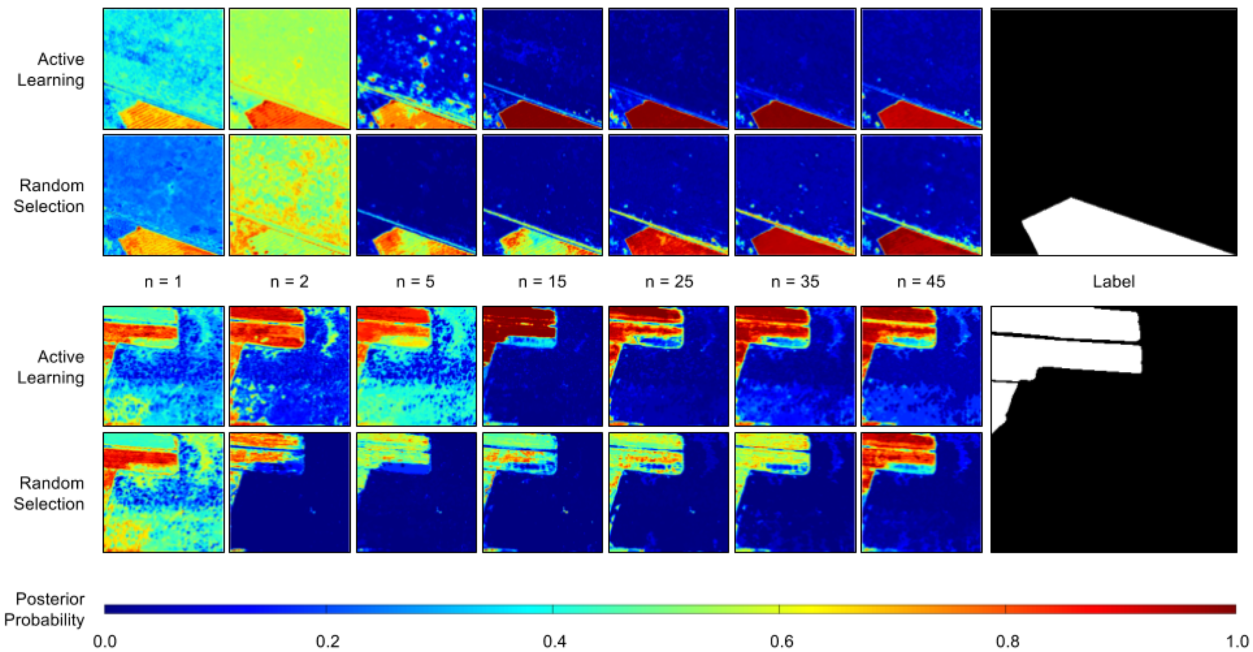

Figure 5: The progression of the algorithm output through 45 iterations of active learning and random selection for two sample images. These images demonstrate how the active learning algorithm converged to a better mapping of agricultural fields with fewer training samples than an algorithm trained with randomly selected samples. The active learning algorithm showed more distinct field boundaries and fewer false positives along roadways.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3004v1 | CC BY 4.0 Open Access | rec: 6 Jun 2017, publ: 6 Jun 2017

landcover data, given proper platform design and appropriate worker assessments, providing new opportunities for creating training data and reducing the dependence on subject matter experts.

In our case study of digitizing agricultural field boundaries in high-resolution satellite imagery of South Africa, the number of samples needed to achieve a desired level of accuracy was reduced by 62% with active learning over typical random selection. Furthermore, based on qualitative analysis of the algorithm output (Figure 5), active learning resulted in more distinct field boundaries and fewer false positives along roadways. In operation, this reduction in training samples would be reflected in the training time and required computational resources, as well as the worker hours and distributed payments in the crowdsourcing platform.

When scaling the case study to much larger areas, it is important to consider other trade-offs between active learning, which requires less labeled training data but more training iterations, and random selection, which requires much more data but does not require iterative training. Iteratively training the algorithm can be scaled indefinitely across compute nodes, using on-demand cloud computing resources, like Amazon EC2 and Google Cloud Compute. Crowdsourcing more data can also scale indefinitely, but is more limited by the human element: the number of workers active at a given time and whether they have been qualified to participate by passing a training module. Given these considerations, we believe the trade-offs favor an active learning approach in large-scale applications.

Furthermore, by operating within a regular grid using the proposed patch-based uncertainty criterion, the active learning algorithm inherits the crowdsourcing platform's robust quality control measures for filtering workers based on the accuracy of their labeling to minimize error and uncertainty in the training data provided to the algorithm. The patch-based approach also prioritizes areas of high uncertainty with a single query per iteration. However, in large-scale applications, batch-mode active learning would be needed to iteratively improve performance in a reasonable amount of time and take advantage of fleets of crowd-

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3004v1 | CC BY 4.0 Open Access | rec: 6 Jun 2017, publ: 6 Jun 2017

sourcing workers functioning in parallel. To avoid adding a batch of image patches with redundant information, future work will focus on including diversity measures and spatial information in the patch-based uncertainty criterion, building on relevant approaches for pixel-based uncertainty criteria (Brinker, 2003; Fu et al., 2012; Gao et al., 2016; Huo & Tang, 2014; Liu et al., 2009; Pasolli et al., 2011; Persello et al., 2014; Shi et al., 2016).

From previous work, it was estimated that mapping the entire continent of Africa with crowdsourcing alone would take 1.9 years at a cost of $2 million (Estes et al., 2016). Using a supervised classification algorithm for mapping agricultural fields, we could conservatively designate a random sample of 50% of Africa's landmass as training data, using a simple rule-of-thumb from (Hastie et al., 2001). It would take crowdsourcing workers almost 1 year to produce this much training data at a cost of $1 million. Simply scaling our current findings, we may estimate that an active learning approach would require 62% less training data to achieve similar results, or only 19% of Africa's landmass. This scenario represents just over 4 months of mapping and $380,000 in crowdsourcing worker payments. The benefits of active learning would likely be even greater in large-scale applications, where similar visual patterns across landscapes would further reduce the size of an efficient training data set. By joining crowdsourcing with active learning and a classification algorithm, the problem of mapping agricultural fields across the entire continent of Africa becomes more feasible.

Acknowledgements

This work was supported by funds from the Princeton Environmental Institute, including the Walbridge Fund, the Mary and Randall Hack '69 Research Fund, the Program in Science, Technology, and Environmental Policy (PEI-STEP) Fellowship, and the PEI Grand Challenges program; the NASA Jet Propulsion Laboratory Strategic University Partnerships (JPL SURP) Graduate Research Program (1524338); the National Science Foundation (SES-1360463, SES-1534544, BCS-1026776); and the NASA New Investigator Program (NNX15AC64G).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3004v1 | CC BY 4.0 Open Access | rec: 6 Jun 2017, publ: 6 Jun 2017

References

Albuquerque, J., Herfort, B., & Eckle, M. (2016). The Tasks of the Crowd: A Typology of Tasks in Geographic Information Crowdsourcing and a Case Study in Humanitarian Mapping. *Remote Sensing*, *8*, 859.

Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, *43*, 1223–1232.

Angluin, D. (1988). Queries and concept learning. *Machine Learning*, *2*, 319–342.

Baum, E. B. (1991). Neural net algorithms that learn in polynomial time from examples and queries. *IEEE Transactions on Neural Networks*, *2*, 5–19.

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *114*, 24–31.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review*, *94*, 115–147.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.

Brinker, K. (2003). Incorporating diversity in active learning with support vector machines. In *Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003)*. Washington DC.

Campbell, C., Cristianini, N., & Smola, A. (2000). Query learning with large margin classifiers. In *Proceedings of Seventeenth International Conference on Machine Learning*.

Cheng, S., & Shih, F. Y. (2007). An improved incremental training algorithm for support vector machines using active query. *Pattern Recognition*, *40*, 964–971.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3004v1 | CC BY 4.0 Open Access | rec: 6 Jun 2017, publ: 6 Jun 2017

Chi, M., & Bruzzone, L. (2005). A Semilabeled-Sample-Driven Bagging Technique for Ill-Posed Classification Problems. *IEEE Geoscience and Remote Sensing Letters*, *2*, 69–73.

Cohn, D., Atlas, L., & Ladner, R. (1994). Improving generalization with active learning. *Machine Learning*, .

Cohn, D. A., Ghahramani, Z., & Jordan, M. I. (1996). Active learning with statistical models. *Journal of Artificial Intelligence Research*, *4*, 129–145.

Comber, A., Mooney, P., & Purves, R. S. (2015). Comparing national differences in what people perceive to be there: mapping variations in crowd sourced land cover. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *40*, 71–75.

Comber, A., Mooney, P., Purves, R. S., Rocchini, D., & Walz, A. (2016). Crowdsourcing: It Matters Who the Crowd Are. The Impacts of between Group Variations in Recording Land Cover. *PLoS ONE*, *11*, e0158329.

Crawford, M. M., Tuia, D., & Yang, H. L. (2013). Active learning: Any value for classification of remotely sensed data? In *Proceedings of the IEEE*.

Debats, S. R., Luo, D., Estes, L. D., Fuchs, T. J., & Caylor, K. K. (2016). A generalized computer vision approach to mapping crop fields in heterogeneous agricultural landscapes. *Remote Sensing of Environment*, *179*, 210–221.

Du, P., Samat, A., Waske, B., Liu, S., & Li, Z. (2015). Random Forest and Rotation Forest for fully polarized SAR image classification using polarimetric and spatial features. *ISPRS Journal of Photogrammetry and Remote Sensing*, *105*, 38–53.

Estes, L. D., McRitchie, D., Choi, J., Debats, S., Evans, T., Guthe, W., Luo, D., Ragazzo, G., Zempleni, R., & Caylor, K. K. (2016). A platform for crowdsourcing the creation of representative, accurate landcover maps. *Environmental Modelling & Software*, *80*, 41–53.

Foody, G. M., & Mathur, A. (2004). Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification. *Remote Sensing of Environment*, *93*, 107–117.

Foody, G. M., Mathur, A., Sanchez-Hernandez, C., & Boyd, D. S. (2006). Training set size requirements for the classification of a specific class. *Remote Sensing of Environment*, *104*, 1–14.

Fu, Y., Zhu, X., & Li, B. (2012). A survey on instance selection for active learning. *Knowledge and Information Systems*, *35*, 249–283.

Gao, F., Lv, W., Zhang, Y., Sun, J., & Wang, J. (2016). A novel semisupervised support vector machine classifier based on active learning and context information. *Multidimensional Systems and Signal Processing*, *27*, 969–988.

Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, *69*, 211–221.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The Elements of Statistical Learning*. Springer Series in Statistics.

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual review of psychology*, *50*, 243–271.

Huo, L.-Z., & Tang, P. (2014). A Batch-Mode Active Learning Algorithm Using Region-Partitioning Diversity for SVM Classifier. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*, 1036–1046.

Jacobson, A., Dhanota, J., Godfrey, J., Jacobson, H., Rossman, Z., Stanish, A., Walker, H., & Riggio, J. (2015). Environmental Modelling & Software. *Environmental Modelling & Software*, *72*, 1–9.

Jones, P., & Thornton, P. (2003). The potential impacts of climate change on maize production in Africa and Latin America in 2055. *Global Environmental Change*, *13*, 51–59.

Juel, A., Groom, G. B., Svenning, J.-C., & Ejrnæs, R. (2015). Spatial application of Random Forest models for fine-scale coastal vegetation classification using object based analysis of aerial orthophoto and DEM data. *International Journal of Applied Earth Observation and Geoinformation*, *42*, 106–114.

Khatami, R., Mountrakis, G., & Stehman, S. V. (2016). A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research. *Remote Sensing of Environment*, *177*, 89–100.

Lewis, D. D., & Catlett, J. (1994). Heterogeneous uncertainty sampling for supervised learning. In *Proceedings of ICML-94, 11th International Conference on Machine Learning* (pp. 148–156).

Lewis, D. D., & Gale, W. A. (1994). A sequential algorithm for training text classifiers. In *Proceedings of 17th Annual International ACM SIGIR conference on Research and Development in Information Retrieval*.

Li, J., Bioucas-Dias, J. M., & Plaza, A. (2010). Semisupervised Hyperspectral Image Segmentation Using Multinomial Logistic Regression With Active Learning. *IEEE Transactions on Geoscience and Remote Sensing*, *48*, 4085–4098.

Li, M., Ma, L., Blaschke, T., Cheng, L., & Tiede, D. (2016). International Journal of Applied Earth Observation and Geoinformation. *International Journal of Applied Earth Observations and Geoinformation*, *49*, 87–98.

Li, M., & Sethi, I. K. (2006). Confidence-based active learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*, 1251–1261.

Liu, A., Jun, G., & Ghosh, J. (2009). Spatially Cost-sensitive Active Learning. In *SIAM International Conference on Data Mining (SDM)* (pp. 814–825). Sparks, NV, USA: Society for Industrial and Applied Mathematics.

Lu, D., & Weng, Q. (2007). A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, *28*, 823–870.

MacKay, D. (1992). Information-based objective functions for active data selection. *Neural computation*, *4*, 590–604.

Ofli, F., Meier, P., Imran, M., Castillo, C., Tuia, D., Rey, N., Briant, J., Millet, P., Reinhard, F., Parkan, M., & Joost, S. (2016). Combining Human Computing and Machine Learning to Make Sense of Big (Aerial) Data for Disaster Response. *Big Data*, *4*, 47–59.

Pasolli, E., Melgani, F., Tuia, D., Pacifici, F., & Emery, W. J. (2011). Improving active learning methods using spatial information. In *2011 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2011* (pp. 3923–3926). IEEE.

Pelletier, C., Valero, S., Inglada, J., Champion, N., & Dedieu, G. (2016). Remote Sensing of Environment. *Remote Sensing of Environment*, *187*, 156–168.

Persello, C., Boularias, A., Dalponte, M., Gobakken, T., Naesset, E., & Schoelkopf, B. (2014). Cost-Sensitive Active Learning With Lookahead: Optimizing Field Surveys for Remote Sensing Data Classification. *IEEE Transactions on Geoscience and Remote Sensing*, *52*, 6652–6664.

Plutowski, M., & White, H. (1993). Selecting concise training sets from clean data. *IEEE Transactions on Neural Networks*, *4*, 305–318.

Salk, C. F., Sturn, T., See, L., Fritz, S., & Perger, C. (2016). Assessing quality of volunteer

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3004v1 | CC BY 4.0 Open Access | rec: 6 Jun 2017, publ: 6 Jun 2017

crowdsourcing contributions: lessons from the Cropland Capture game. *International Journal of Digital Earth*, *9*, 410–426.

See, L., Comber, A., Salk, C., Fritz, S., van der Velde, M., Perger, C., Schill, C., McCallum, I., Kraxner, F., & Obersteiner, M. (2013a). Comparing the Quality of Crowdsourced Data Contributed by Expert and Non-Experts. *PLoS ONE*, *8*, e69958.

See, L., McCallum, I., Fritz, S., Perger, C., Kraxner, F., Obersteiner, M., Baruah, U. D., Mili, N., & Kalita, N. R. (2013b). Mapping cropland in Ethiopia using crowdsourcing. *International Journal of Geosciences*, *4*, 6–13.

See, L., Mooney, P., Foody, G., Bastin, L., Comber, A., Estima, J., Fritz, S., Kerle, N., Jiang, B., Laakso, M., Liu, H.-Y., Milčinski, G., Niksič, M., Painho, M., Pődör, A., Olteanu-Raimond, A.-M., & Rutzinger, M. (2016). Crowdsourcing, Citizen Science or Volunteered Geographic Information? The Current State of Crowdsourced Geographic Information. *ISPRS International Journal of Geo-Information*, *5*, 55.

See, L., Schepaschenko, D., Lesiv, M., McCallum, I., Fritz, S., Comber, A., Perger, C., Schill, C., Zhao, Y., Maus, V., Siraj, M. A., Albrecht, F., Cipriani, A., Vakolyuk, M., Garcia, A., Rabia, A. H., Singha, K., Marcarini, A. A., Kattenborn, T., Hazarika, R., Schepaschenko, M., van der Velde, M., Kraxner, F., & Obersteiner, M. (2015). Building a hybrid land cover map with crowdsourcing and geographically weighted regression. *ISPRS Journal of Photogrammetry and Remote Sensing*, *103*, 48–56.

Shi, Q., Huang, X., Li, J., & Zhang, L. (2016). Active learning approach for remote sensing imagery classification using spatial information. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS* (pp. 1520–1523). IEEE.

Stumpf, A., Lachiche, N., Malet, J.-P., Kerle, N., & Puissant, A. (2014). Active Learning in the Spatial Domain for Remote Sensing Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, *52*, 2492–2507.

Tokarczyk, P., Wegner, J. D., Walk, S., & Schindler, K. (2013). Beyond hand-crafted features in remote sensing. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *1*, 35–40.

Tokarczyk, P., Wegner, J. D., Walk, S., & Schindler, K. (2015). Features, Color Spaces, and Boosting: New Insights on Semantic Classification of Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, *53*, 280–295.

Tuia, D., Pasolli, E., & Emery, W. J. (2011a). Using Active Learning to Adapt Remote SensingImage Classifiers. *Remote Sensing of Environment*, *115*, 2232–2242.

Tuia, D., Volpi, M., Copa, L., Kanevski, M., & Munoz-Mari, J. (2011b). A Survey of Active Learning Algorithms for Supervised Remote Sensing Image Classification. *IEEE Journal of Selected Topics in Signal Processing*, *5*, 606–617.

Ursani, A. A., Kpalma, K., Lelong, C. C. D., & Ronsin, J. (2012). Fusion of Textural and Spectral Information for Tree Crop and Other Agricultural Cover Mapping With Very-High Resolution Satellite Images. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, *5*, 225–235.

Van Niel, T. G., McVicar, T. R., & Datt, B. (2005). On the relationship between training sample size and data dimensionality: Monte Carlo analysis of broadband multi-temporal classification. *Remote Sensing of Environment*, *98*, 468–480.

Vetrivel, A., Gerke, M., Kerle, N., & Vosselman, G. (2015). Identification of damage in buildings based on gaps in 3D point clouds from very high resolution oblique airborne images. *ISPRS Journal of Photogrammetry and Remote Sensing*, *105*, 61–78.

Von Ahn, L., Blum, M., Hopper, N. J., & Langford, J. (2003). CAPTCHA: Using Hard AI Problems for Security. In *"Advances in Cryptology - EUROCRYPT 2003: International Conference on the Theory and Applications of Cryptographic Techniques Proceedings* (pp. 294–311).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3004v1 | CC BY 4.0 Open Access | rec: 6 Jun 2017, publ: 6 Jun 2017

Wang, X., & Zhai, J. (2016). *Learning with Uncertainty*. Boca Raton: CRC Press.

Xie, S., Duan, J., Liu, S., Dai, Q., Liu, W., Ma, Y., Guo, R., & Ma, C. (2016). Crowdsourcing Rapid Assessment of Collapsed Buildings Early after the Earthquake Based on Aerial Remote Sensing Image: A Case Study of Yushu . . . . *Remote Sensing*, .

Zaharia, M., Chowdhury, M., Franklin, M. J., Shenker, S., & Stoica, I. (2010). Spark: cluster computing with working sets. In *HotCloud*.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3004v1 | CC BY 4.0 Open Access | rec: 6 Jun 2017, publ: 6 Jun 2017