Active Learning Approaches for Deforested Area Classification

Fernanda B.J.R. Dallaqua*, Fabio A. Faria*, and Alvaro L. Fazenda*
*Instituto de Ciência e Tecnologia
Unversidade Federal de São Paulo, São José dos Campos
Email: {fernanda.dallaqua, ffaria, alvaro.fazenda}@unifesp.br

***Abstract*—The conservation of tropical forests is a social and ecological relevant subject because of its important role in the global ecosystem. Forest monitoring is mostly done by extraction and analysis of remote sensing imagery (RSI) information. In the literature many works have been successful in remote sensing image classification through the use of machine learning techniques. Generally, traditional learning algorithms demand a representative and huge training set which can be an expensive procedure, especially in RSI, where the imagery spectrum varies along seasons and forest coverage. A semi-supervised learning paradigm known as active learning (AL) is proposed to solve this problem, as it builds efficient training sets through iterative improvement of the model performance. In the construction process of training sets, unlabeled samples are evaluated by a user-defined heuristic, ranked and then the most relevant samples are labeled by an expert user. In this work two different AL approaches (*Confidence Heuristics* and *Committee*) are presented to classify remote sensing imagery. In the experiments, our AL approaches achieve excellent effectiveness results compared with well-known approaches existing in the literature for two different datasets.**

I. Introduction

Conservation of tropical forests is a social and ecological relevant issue because of its important role in the global ecosystem. Tropical forests have a great diversity of fauna and flora, besides regulating the climate and rainfall, absorbing large quantities of carbon dioxide and being indigenous dwellings. Unfortunately, millions of hectares of tropical forests have been lost and degraded over years [1].

Technology can be a great ally for the preservation of the tropical forest as remote sensing images, data and classification algorithms are used to analyze, identify and quantify changes in the environment, making possible to monitor forest deforestation and degradation [1]. One example of forest monitoring program is PRODES (Amazon Deforestation Monitoring Project) [2], produced by the Brazilian National Institute for Space Research (INPE), which carries annual deforestation survey by a semi-automated process [1].

In image classification, the supervised learning algorithms performance strongly depends on the representativeness of the training set. However, to build a great training set experts and extensive manual analysis are usually mandatory, which can make the procedure to be slow and financially expensive. So, small training sets to obtain high classification accuracy is always desirable.

In machine learning literature this sampling approach is known as AL, which aims to achieve high classification effectiveness rates by using few training samples. This technique consists in iteratively selecting unlabeled instances to be labeled by an oracle (expert user) in such way that only the samples that improve the classification model will be included in the training set [3].

An important step in AL is to choose which samples are the most interesting to be included in the training set. This can be done by several families of heuristics as committee-based, large margin-based, and posterior probability-based [3]. These strategies obtained significantly better results and with far fewer samples compared to random sample selection [4].

In this paper different AL approaches were applied to a remote sensing image obtained by Landsat-8 satellite (launched as the Landsat Data Continuity Mission - LDCM), which collects data covering the entire Earth every 16 days with 30 meters for spatial resolution. The case of study covers a section of Amazon forest located on Rondônia state (north of Brazil) at July 2016, in order to classify the pixels that represent forest or non-forest areas. Classified data from PRODES was used to train the classification models and to compare the results. The experiments presented here validated the AL approaches applied to forest monitoring issue.

II. Background and Related Works

This section presents background and related works about deforestation detection and AL, subjects of this paper.

A. Brazilian Amazon Deforestation detection

The Brazilian Legal Amazon (BLA) is one of the world's best-known rainforests, hosting the highest biodiversity of several forest-dependent species and being the largest continuous rainforest in the world [5]. Thus, its deforestation can bring great consequences, being needed monitoring programs for controlling and prevention of deforestation. These monitoring programs are done by INPE, being PRODES and DETER two of them.

PRODES was created in 1988 to carry annual deforestation surveys in BLA. At first, the classification was only visual and manually performed by experts, analyzing printed remote sensing imagery. In 2003 it started to use a semi-automated procedure to perform TM/Landsat images digital processing, classifying forest areas and clear-cut deforested areas like pastures, savannas, crops, abandoned lands, water bodies and rivers, urban areas and remaining cloud areas. Both data, images and tabular informations are annually posted on the program's website[1] [1].

As PRODES gives annual surveys and quick actions are needed to stop the beginning of a deforestation process, a program for near real-time deforestation detection was created. It is called DETER [5].

DETER uses imagery from MODIS (Moderate Resolution Imaging Spectroradiometer), which has a revisit time of roughly 1.5 days and 250 meters of spatial resolution [1] against 16 days and 30 meters of spatial resolution for Landsat [6]. Although the spatial resolution of MODIS, the high observation frequency enables surveys and issuing alert almost in real-time. The area where a deforestation process was identified is marked and an alert is sent to the Brazilian Institute of Environment and Renewable Natural Resources (IBAMA) [5]. These alerts are also monthly posted in the program's website[2].

Not only INPE has monitoring programs for BLA. A non-profit research institution called Imazon (in portuguese, Instituto do Homem e Meio Ambiente da Amazônia) created a near-real time deforestation monitoring system named SAD (in portuguese, Sistema de Alerta de Desmatamento) and provides monthly surveys that are available in the institute's website[3]. The SAD system also uses MODIS images, where deforestation change detection is performed based on Normalized Difference Fraction Index (NDFI). If the pixels present NDFI below 125 they are classified as deforestation, while pixels with NDFI values between 125 and 165 are forest degradation [7].

Another monitoring programs are ForestWatchers and GLAD (Global Land Analysis & Discovery) Alerts. ForestWatchers is a Citizen Science [8] [9] project created in 2012 aiming to monitor the tropical forest deforestation by volunteers analyzing and classifying remote sensing images from MODIS [1] [10]. GLAD Alerts is supported by Global Forest Watch (GFW)[4], an online platform that provides data and tools to allow near real-time information about forest monitoring. It uses Landsat-7/8 with 30 meters of spatial resolution to alert tree-cover loss not only in BLA but in different countries in the Amazon, Congo Basin, and Southeast Asia. Metrics from previous Landsat imagery and the latest image are run through decision trees to calculate a median probability of forest disturbance, where pixels with the probability higher than 50% are considered as tree cover loss and an alert is issued [11].

B. Active Learning

One of the biggest challenges in traditional machine learning techniques is to create a training set that represents the real data behavior using as few samples as possible of the dataset. This fact might be caused by high intra-classes or low inter-classes variance that make a learned model to fail whether trained with an inefficient training set.

In literature AL is used to sort out this problem of sample selection of datasets. It aims to build an efficient training set through iterative performance improvement of models using sampling [3].

Usually, AL approaches begin with a small number of instances in the training set and iteratively try to build a great training set that minimizes the classification error [12]. Therefore, a user-defined heuristic is used to sort all the unlabeled instances (candidates) and select those instances that are more valuable for the learning model improvement. Finally, an expert classification is given for each most interesting sample selected by the heuristic. This process repeats iteratively until a stop criteria is satisfied [3]. Algorithm 1 shows a general procedure of AL.

Algorithm 1 GENERAL PROCEDURE FOR ACTIVE LEARNING

```
Inputs : Initial training set X
         Pool of training samples candidates U
         Number of samples q to add at each iteration
repeat
    Train a model with current training set X.
    for each candidate in U do
        Evaluate a user-defined heuristic
    end
    Rank the candidates in U according to the score of the
      heuristic.
    Select the q most interesting samples.
    The user assigns a label to the selected samples.
    Add these samples to the training set X.
    Remove the samples from the pool of candidates U.
until stop criteria is reached;
```

AL approaches have been used in the literature to solve problems in several application domains such as medicine [13], biology [14], chemistry [15], biometric [16], and remote sensing.

Recently, in remote sensing imagery works, AL approaches have been proposed for a different type of tasks such as classification [12], [17]–[19], segmentation [20], [21], retrieval [22], and detection of Land-Cover transitions [23].

In image classification tasks, Tuia et al. [12] proposed two AL algorithms based on margin sampling strategies using support vector machines (SVM). Persello and Bruzzone [17] adopted domain adaptation paradigm to exploit labeled samples of source domain and to minimize the number of target domain samples to be labeled with the definition of the final training set. Li et al. [18] proposed a framework for spectral-spatial classification, which exploits marginal probability distributions in hyperspectral data. Stumpf et al. [19] developed region-based query strategies to select spatial batches with high sample uncertainty and diversity through the use of a tree ensemble classifier (Random Forest - RF).

[1] http://www.obt.inpe.br/prodes/
[2] http://www.obt.inpe.br/deter/
[3] http://imazon.org.br/en/
[4] https://www.globalforestwatch.org/

In image segmentation tasks, Li et al. [20] introduced a Bayesian approach for hyperspectral image segmentation, which uses a multinomial logistic regression (MLR) model to learn the class posterior probability distributions and then to segment the hyperspectral image. Mitra et al. [21] explored a support vector machines technique to minimize the number of labeled data that is used to train a learning model.

For retrieval, Demir and Buzzone [22] developed a novel AL to improve a content-based image retrieval (CBIR) by relevance feedback from large archives of remote sensing images. The proposed AL method defines a small as possible set of relevant and irrelevant images. The same author joined a new researcher [23] and also worked to detect Land-Cover transitions using AL. The new developed technique detects land-cover transitions in a pair of remote sensing images acquired in the same area at different times through a multi-temporal training set. The dataset consists of unlabeled pixels aligned at the same location in two available images, which present maximum uncertainty by joint-entropy criteria.

III. Proposed Active Learning Approaches

In this work, two AL approaches have been proposed. The first, *confidence heuristic* approach, take into account the sample classification confidence of a single classifier. Second, *committee* approach use the disagreement of a classifier set.

1) Confidence Heuristics: In this approach, a ranking aggregating candidate instances was created following a specific criteria according to the classification's confidence. The three different *Confidence Heuristics* used to define the instance ranking are defined next:

- **Low Confidence**: the rank value for each sample is evaluated by ascending order over the difference in module between the sample probability for forest and non-forest ($|Prob_F - Prob_{NF}|$, remembering that $Prob_F + Prob_{NF} = 1.0$), i.e., instances where the probability difference depicted was close to zero represent image samples with low confidence because the probability is similar to be a forest or non-forest pixel.
- **High Confidence**: the sample ranking is generated in the opposite way to Low Confidence, i.e., considering a descending order for $|Prob_F - Prob_{NF}|$. The pixels ranking in this approach are classified with high confidence by the classifier.
- **Hybrid Confidence**: half of the inserted samples are low confidence samples and the rest are high confidence samples, trying to reach a more stable behavior by also adding samples that have high confidence classification.

2) Committee: In this approach, a committee is created by joining classifiers k-Nearest Neighbors, Linear Discriminant Analysis and Multi-layer Perceptron. These classifiers were chosen because they achieved good result and they aren't based on ensemble.

The samples of the pool of candidates are classified and the ones to be inserted in the training set are defined by the disagreement between the classifier set. The samples that have the maximum disagreement are the ones to be inserted in the training set. The test set is classified by majority vote according to the classifier committee used in the task.

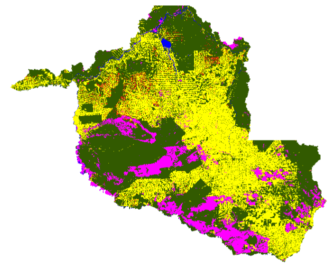


Fig. 1. Rondônia state classified by PRODES (2016).

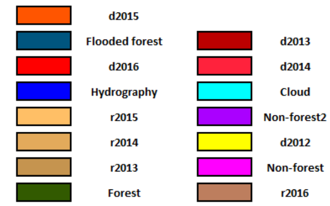


Fig. 2. Color code for PRODES image.

IV. Experimental Methodology

This section defines the Datasets used, the Classifiers and Baseline methods evaluated and the Experimental Setup adopted in this work.

A. Datasets

For this work, we use a remote sensing image from Landsat-8 imagery, freely available in EarthExplorer[5], of Rondônia state (north of Brazil) at July 2016. This image is composed of 7 bands (ultra blue, blue, green, red, near infrared, shortwave infrared 1 and shortwave infrared 2) with 30 meters of spatial resolution. A labeled image was obtained by PRODES, explained in II-A.This labeled image is built with 60 meters of spatial resolution.

Figures 1 and 2 show the mosaic of the state of Rondônia classified by PRODES for 2016 and its legend. The designated areas of non-forest (Non-forest and Non-forest2) are vegetation areas not considered as Amazon Forest; d2012 represents the agglomerated deforestation until 2012; d2013,...,d2016 represent the deforestation of each year; r2013,...,r2016 are areas where deforestation was detected in that year but that already existed in the past and wasn't mapped by the specialist, making not possible to assert the correct year of this deforestation. PRODES does not reclassify the deforested areas, that is, even if forest regeneration occurs, the area will continue to be classified as deforested. Thus, noise data can be found in the groundtruth.

In our application we just consider two classes: forest - corresponding to the label with the same name in PRODES - and non-forest - corresponding to all the other classes that are different than the forest in PRODES.

In our experiments, we used two different remote sensing images that correspond to small areas belonging to the state of Rondônia, Figure 3a and 4a. Both images were resized from 30 meters of spatial resolution to 60 meters and they are shown with RGB composition in here but for the study 7 bands were considered. Figure 3a corresponds to an area of approximately $115km^2$ and it was used on a cross-validation protocol [24], which is composed of 32,096 non-redundant pixels, where

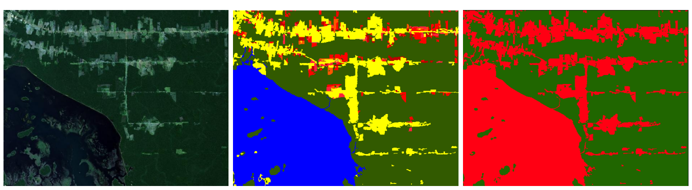
[5] https://earthexplorer.usgs.gov/

16,390 are forest and 15,706 are non-forest, divided by 5 folds. Figure 4a corresponds to an area of approximately 1,102km^2 and it was used for a cross-dataset experiment, which includes more diverse PRODES non-forest classes that can occur in a real application, when acquiring new remote sensing images for different areas, dates, and acquisition devices. This image has 306,114 pixels, where 168,999 are forest and 137,115 are non-forest.

As we consider a binary problem, the PRODES image is transformed in an image composed by green (forest) and red (non-forest) pixels, according to Figures 3c and 4c, corresponding to training/cross-validation and cross-dataset image, respectively.

(a) Original Image. **(b) PRODES Image.** **(c) Binary Image.**

Fig. 3. Image used in cross-validation experiments with AL approaches.



(a) Original Image. **(b) PRODES Image.** **(c) Binary Image.**

Fig. 4. Image used in cross-dataset experiments with the best AL approaches.

B. *Classifiers*

We have used eight different classifiers: AdaBoost (ADA) with 50, 100 and 200 decision trees as estimators, Gradient Boosting Classifier (GBC) with 50 regression trees as estimators, k-Nearest Neighbors (kNN) with $k = \{1, 3, 5, 7, 9\}$, Multi-Layer Perceptron (MLP) with one hidden layer composed by 3 nodes, Gaussian Naïve Bayes (GNB), Linear Discriminant Analysis (LDA) and Random Forest (RF) with 10, 100 and 500 estimators. Remembering that the *Committee* approach uses kNN with $k = \{5, 7, 9\}$, LDA and MLP. The proposed approaches have been implemented using Python 2.7.5 and Scikit-Learn[6] [25].

These classifiers include a method called *predict_proba(X)* which returns the probability estimates for the test data X. These probability estimates that will be used to calculate the confidence of each candidate sample for the *Confidence Heuristics* approach.

[6]http://scikit-learn.org/stable/ (As of January 2018)

C. *Baseline Approaches*

Baseline approaches used the *MATLAB Active Learning Toolbox for Remote Sensing* [3], which uses Support Vector Machine (SVM) as the classifier. The heuristics *Normalized Entropy Query-by-Bagging* (nEQB) [26] and *Margin Sampling* (MS) [21] [27] [28] were used to define the samples to be included in the training set. The parameters used in these methods were the default parameters provided by the toolbox. A random sampling was also used to compare the approaches used in this work.

D. *Experimental Setup*

For the 5-fold cross-validation experiment, one fold is used as the test set and the remaining folds as training and candidate sets. The initial training set was composed by ten labeled pixels (five per class) and, for each AL iteration, the samples are ranked by the approaches detailed in III and the six most interesting pixels/samples are introduced in the training set. This process is repeated up to 500 iterations and compared with the traditional supervised learning.

Taking the first experiment results, the best single classifier with the two best *Confidence Heuristics*, *Committee* approaches and baselines are used to classify a different image in a cross-dataset experiment. Six iteration cut-points (10, 20, 30, 40, 50, and 100) are defined and the training sets built in those cut-points are used in the AL approaches depicted, in order to perform the cross-dataset classification. A supervised classification is also evaluated and compared with the results.

For the analysis of the results, the classifications for all experiments are compared to groundtruth (PRODES) and Cohen's Kappa [29] and Overall Accuracy (OA) are computed.

V. RESULTS

In this work, three different analysis has been performed. First, a comparative analysis among our proposed sampling selection approaches (*Confidence Heuristics* and *Committee*) V-A. Second, a comparative study among the best *Confidence Heuristics* approaches (RF-low and RF-hybrid), the *Committee* approach and two well-known baseline approaches (MS and nEQB) existing in the literature V-B. Finally, aiming at a more realistic problem, a cross-dataset experiment is performed and the visual results are showed V-C.

A. *Effectiveness Analysis among Active Learning Approaches*

In Figure 5 six different results with our proposed AL approaches are presented. As *Confidence Heuristics* the five best results presented corresponding to ADA with 100 estimators, kNN using $k = \{7, 9\}$ (kNN7 and kNN9), MLP, and RF with 500 estimators. Furthermore, we present our *Committee* (Figure 5f) using five different classifiers (kNN5, kNN7, kNN9, LDA and MLP). It is named *Committee* 5CB.

As we can observe, in Figure 5(a – e), almost all of the *Confidence Heuristics* approaches based on low (blue line) and hybrid (green line) confidences have achieved the best effectiveness results than approaches based on random (pink line) and high (beige line) confidence. In addition, our

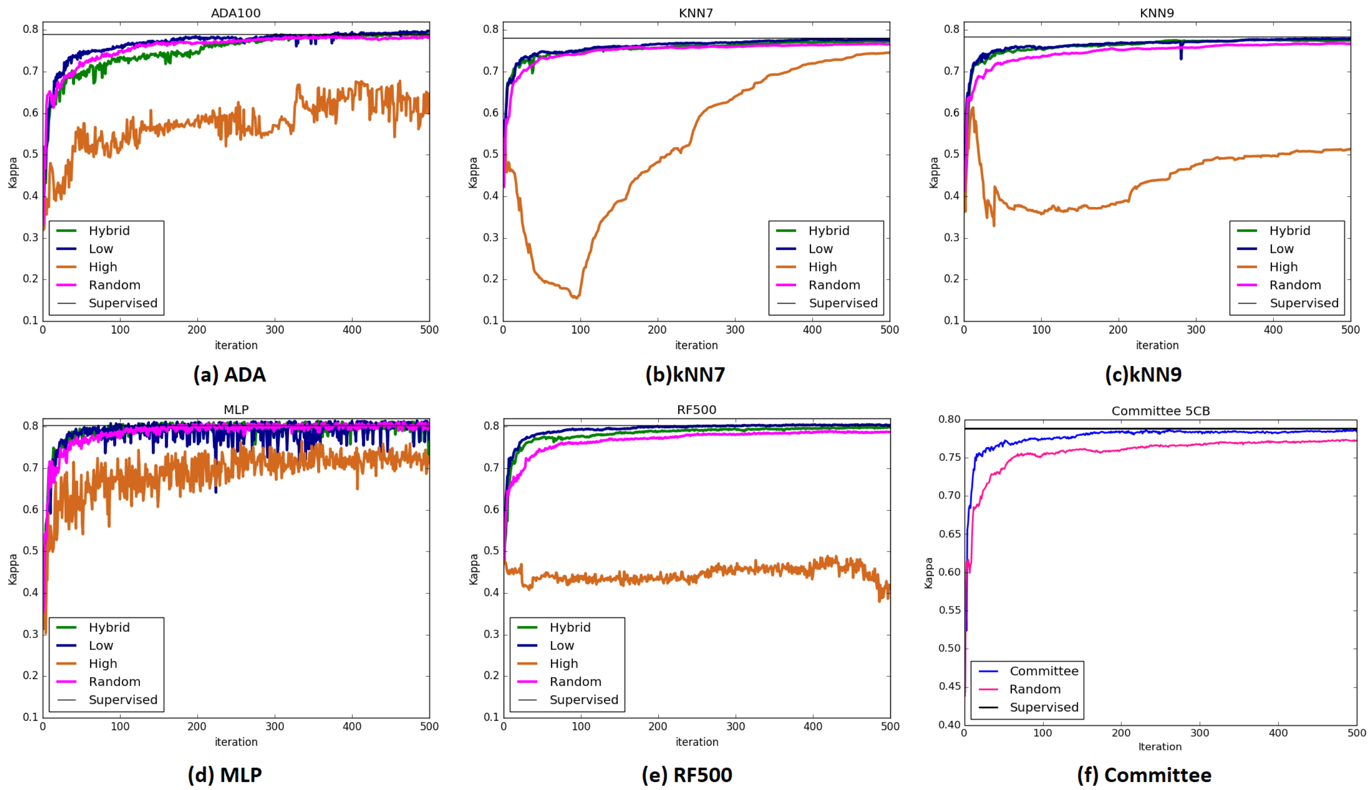


Fig. 5. Effectiveness results of each sampling selection approach proposed in this work. In (a)–(e) are five *Confidence Heuristics* approaches. In (f) is the *Committee* approach (5CB).

approaches based on low and hybrid achieved similar results to the supervised classifier, however, they used much fewer samples to train. Our approaches used around 1200 samples against 32.096 samples used by the supervised classifier. The best single classifier was RF as it achieves higher Kappa, with stability through the iterations and with both low (RF-low) and hybrid (RF-hybrid) confidences behaving better than random.

In Figure 5(f), it is possible to note that our *Committee* approach (*Committee* 5CB) have been better than random approach (pink line) and also achieved similar results to supervised classifiers using approximately 1200 samples.

B. Comparison among the Best Approaches

In this analysis, we compare our best *Confidence Heuristics* approaches (RF-Low and RF-hybrid), our *Committee* approach (*Committee* 5CB) and two well-known baselines (MS and nEQB) existing in the literature.

As we can observe, in Figure 6, the baselines MS (green line) and nEQB (orange line) have been better than our proposed approaches. From our approaches RF-low presented better result and *Committee* 5CB and RF-hybrid start with equivalent performance but after some iterations RF-hybrid had similar result than RF500-Low, being *Committee* 5CB the worst of the approaches.

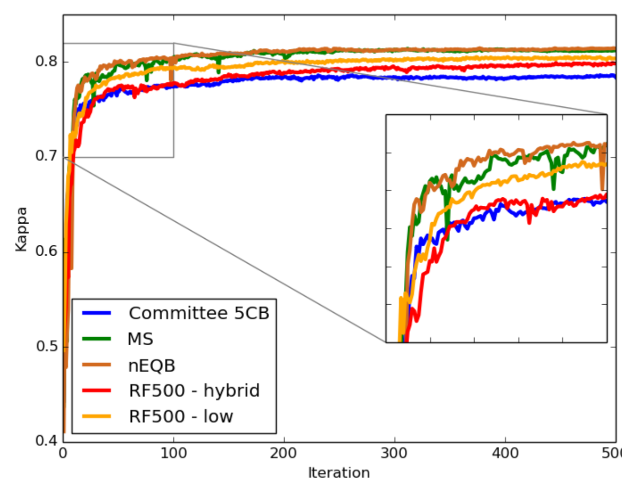


Fig. 6. Effectiveness results among our proposed approaches and two well-known baseline approaches.

C. Cross-Dataset Scenario

In this analysis, we compare all of the best AL approaches for a cross-dataset scenario, which each trained approach in the previous experiment with an image predicts another image (Figure 4) in different iteration cut-points $(10, 20, 30, 40, 50, 100)$.

In order to verify the robustness of the approaches, the mean Kappa Index (arithmetic average for five training sets), the mean OA and theirs confidence interval (CI) were computed and presented in Tables I and II. These Tables show the *Committee* and RF-Low approaches achieving the best results for most of iteration cut-points. The baseline approaches, specially nEQB, showed smaller robustness comparing to *Committee* 5CB and the best *Confidence Heuristic*- RF-low and RF-hybrid - since the average Kappa Index and OA presented low values with high CI especially after cut-point 20, where can be noticed that the CI is higher than the average values.

All evaluated approaches present better results for the initial iteration cut-points, possibly due to better generalization with less training samples, which promotes robustness. The baseline approaches presented unstable results and decreased as more samples were used in the training set. This behavior could be explained by the grid-search tuning method to set SVM parameters, making it less robust, but it requires more investigation to confirm. A supervised learning was also evaluated. RF for both low and hybrid confidence showed similar or better results than supervised for all cut-points; *Committee* 5CB at the cut-point 20 also achieved similar result than supervised with much fewer samples; SVM as classifier presented the worst results.

Table III shows the averages of Kappa Index and OA for all cut-points presented by Tables I and II. It is possible to notice that the proposed strategies (RF and *Committee*) achieved better Kappa Index and OA than the baseline. Can be highlighted that the classifiers present in the Committee as well as RF are totally based in free software while the baseline uses proprietary software. Also, the processing time of the supervised tests for the classifiers SVM, RF and *Committee* (kNN, LDA and MLP) were measured. All executions were made in a Intel(R) Xeon(R) CPU E5-2660 v4 @ 2.00GHz machine and RF performed in approximately 11,8% of the total processing time of SVM. On the other hand, the *Committee* performed in 35,3% of SVM's time, even including 5 classifiers being executed sequentially. For parallel execution the *Committee* time was even smaller, also performing in 14,1%.

Figure 7 shows pixel classification results for the cross-dataset image using the best active learning approaches presented in this work. The images are taken for iteration cut-point 20, which present best results for almost all of the approaches. These images were generated by doing majority vote among all of the classification results of the five different training sets. The image classifications for MS and *Committee* 5CB approaches present similar OA and Kappa Index when evaluated in this cut-point. However, *Committee* 5CB showed lower value of confidence interval, i.e., less variation among further iterations. Furthermore, *Committee* 5CB showed to be a more robust approach in the following cut-point analysis in comparison with the baseline approaches (MS and nEQB).

VI. Conclusion

In this work, we presented two different active learning approaches for Deforested Area Classification (*Confidence Heuristics* and *Committee*). *Confidence Heuristics* approaches use low, high, and hybrid confidences of single classifiers for selecting samples, at each iteration, that will compose the training sets. *Committee* approach uses the disagreement in a set of classifiers to decide which samples should be selected for the next iteration.

In our experiments, three analysis has been performed. Firstly, an effectiveness analysis among our active learning approaches showed us that the *Confidence Heuristics* approaches based on low and hybrid confidences have achieved better results than sampling random and high confidence approaches. Random Forest (RF-Low and RF-Hybrid) was the best technique among all of the classifiers used in this work. Also, the *Committee* approach also have achieved excellent results when compared with sampling random approaches.

Secondly, a comparative study of our best approaches (RF-low, RF-hybrid and *Committee* 5CB) and two well-known baseline approaches (MS and nEQB) have been performed. In this analysis, both baselines have been better than our active learning approaches.

Finally, in order to verify the behavior of each approach in a real application problem, we have performed an analysis of a cross-dataset scenario. In this analysis, each approach trained with samples from an image and labeled samples from another different image. Thus, we could see that our *Committee* 5CB and RF-low approaches have achieved excellent results for most of iteration cut-points selected in this analysis. This fact showed us that our approaches proved to be robust solutions of active learning for deforested area classification task in the cross-dataset scenario.

According to the analyzes carried out, it is concluded that the proposed methods have advantages over the free availability of Scikit-Learn and over the processing time of the baseline while having similar results for cross-validation (section V-B) and superior in the cross-dataset case (section V-C).

VII. Future work

For future work can be highlighted new cross-dataset experiments with more images and a big amount of pixels (more than 40 million) in order to better attest the robustness of the methods and to better evaluate their behaviors. Also, it is planned to study the noise data that can be found as PRODES does not reclassify deforested areas even in case of regeneration and the consequences of using these in the training set.

Another study that it's in the beginning stages is about instead of using specialist's classification as groundtruth, use volunteers to classify remote sensing images. The use of lay volunteers to gather or classify data in scientific research is known as Citizen Science. It's a subject area that keeps growing due to the amount of data that can be processed in a cheaper, faster and reliable way. With volunteers instead of

TABLE I
EFFECTIVENESS RESULTS AMONG THE BEST AL APPROACHES FOR A CROSS-DATASET SCENARIO. AVERAGE KAPPA INDEX MEANS THE ARITHMETIC MEAN OF THE KAPPA INDEX FOR THE FIVE TRAINING SETS.

Technique	Iteration Cut-Points (Average Kappa Index ± CI 95%)						Supervised
	10	**20**	**30**	**40**	**50**	**100**	
Committee 5CB	0,39 ±0,20	**0,60±0,24**	**0,46±0,22**	0,33±0,03	0,32±0,03	**0,37±0,16**	**0,68±0,10**
MS [3], [21], [27], [28]	**0,49±0,38**	0,57±0,30	0,16±0,17	0,26±0,28	0,03±0,12	0,12±0,34	0,03±0,21
nEQB [3], [26]	0,35±0,30	0,11±0,36	0,17±0,37	-0,06±0,26	-0,22±0,18	-0,18±0,21	
RF - hybrid	0,22±0,14	0,33±0,02	0,27±0,05	0,29±0,07	0,29±0,07	0,31±0,03	0,30±0,12
RF - low	0,48±0,26	0,33±0,05	0,36±0,02	**0,36±0,02**	**0,34±0,02**	0,35±0,02	

TABLE II
EFFECTIVENESS RESULTS AMONG THE BEST AL APPROACHES FOR A CROSS-DATASET SCENARIO. AVERAGE OA MEANS THE ARITHMETIC MEAN OF THE OA FOR THE FIVE TRAINING SETS.

Technique	Iteration Cut-Points (Average OA ± CI 95%)						Supervised
	10	**20**	**30**	**40**	**50**	**100**	
Committee 5CB	0,70±0,09	**0,80 ±0,12**	**0,74±0,10**	0,68±0,02	0,68±0,01	0,70 ±0,07	**0,85±0,05**
MS [3], [21], [27], [28]	0,73±0,22	0,79±0,13	0,54±0,10	0,60±0,16	0,53±0,05	0,58±0,16	0,50±0,10
nEQB [3], [26]	0,68±0,15	0,55±0,20	0,57±0,20	0,43±0,15	0,35±0,08	0,38±0,12	
RF - hybrid	0,62±0,07	0,68±0,01	0,65±0,03	0,66±0,04	0,66±0,03	0,67±0,02	0,66±0,05
RF - low	**0,75±0,12**	0,68±0,02	0,70±0,01	**0,70±0,01**	**0,69±0,01**	**0,70±0,01**	

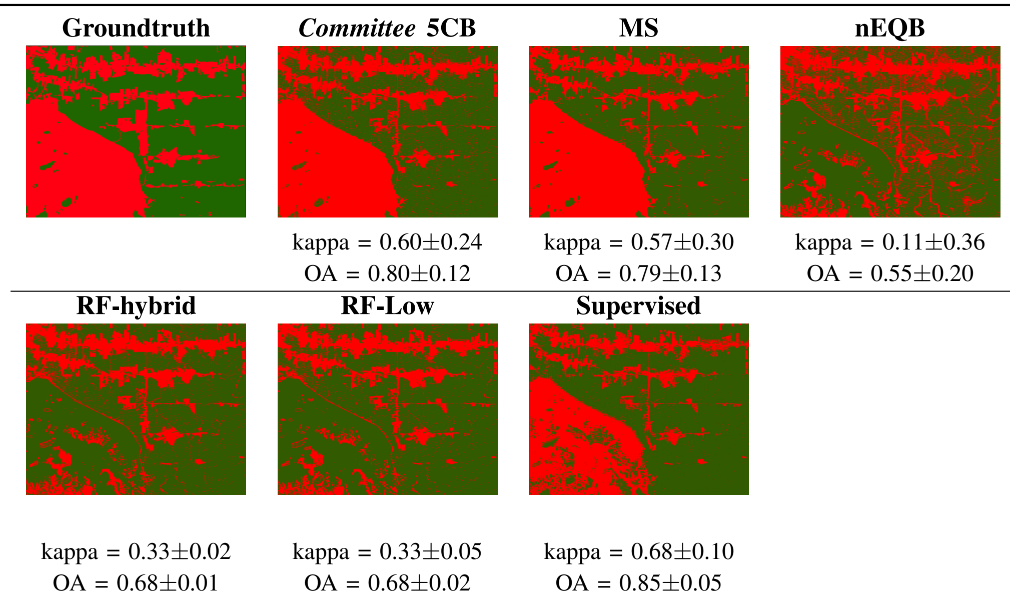


Fig. 7. Pixel classification results of each compared approach using iteration cut-point 20 for the cross-dataset experiment.

TABLE III
CROSS-DATASET EXPERIMENT'S AVERAGES OF KAPPA INDEX AND OA.

Technique	Average Kappa Index	Average OA
Committee 5CB	0,41±0,10	0,72±0,05
MS	0,27±0,21	0,63±0,11
nEQB	0,03±0,22	0,49±0,13
RF - hybrid	0,28±0,04	0,66±0,02
RF - low	0,37±0,06	0,70±0,02

specialists for classification of deforested areas allied to AL, we hope to achieve a cheap and fast procedure to track deforestation in places where there aren't deforestation monitoring programs and to raise more awareness to the general public.

As deep learning [30], [31] is achieving great results in many subject areas [32]–[34] we intend to study it joined with semantic segmentation [35]–[37] to classify remote sensing images, also allying it with AL and Citizen Science, which is the major goal of our research.

VIII. ACKNOWLEDGMENT

This research is part of the INCT of the Future Internet for Smart Cities funded by CNPq, proc. 465446/2014-0, CAPES proc. 88887.136422/2017-00, and FAPESP, proc. 2014/50937-1.

F. A. Faria thanks the support of scientific funding agency CNPq through the Universal Project (grant #408919/2016-7).

Landsat-8 image courtesy of the U.S. Geological Survey (https://usgs.gov).

REFERENCES

[1] Eduardo F. P. Luz et al. "The ForestWatchers: A Citizen Cyberscience Project for Deforestation Monitoring in the Tropics". *Human Computation*, vol. 1, pp. 137-145, 2014.

[2] Matthew C. Hansen, et al. "Comparing annual MODIS and PRODES forest cover change data for advancing monitoring of Brazilian forest cover". *Remote Sensing of Environment*, vol. 112, n. 10, pp. 3784-3793, 2008.

[3] Devis Tuia et al. "A survey of active learning algorithms for supervised remote sensing image classification". *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, n. 3, pp. 606-617, 2011.

[4] Jordi Muñoz-Marí, Devis Tuia, and Gustavo Camp-Valls. "Semisupervised Classification of Remote Sensing Images With Active Queries". *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, n. 10, pp. 3751-3763, 2012.

[5] Cesar Guerreiro Diniz et al. "DETER-B: The new Amazon near real-time deforestation detection system". *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, n. 7, pp. 3619-3628, 2015.

[6] Donald T. Lauer, Stanley A. Morain and Vincent V. Salomonson. "The Landsat program: Its origins, evolution, and impacts". *Photogrammetric Engineering and Remote Sensing*, vol. 63, n. 7, pp. 831-838, 1997.

[7] Carlos M. de Souza Jr., Sanae Hayashi and Adalberto Verssimo. "Near real-time deforestation detection for enforcement of forest reserves in Mato Grosso", in *Proceedings of Land Governance in Support of the Millennium Development Goals: Responding to New Challenges*, World Bank Conference, Washington, DC, 2009.

[8] Jonathan Silvertown. "A new dawn for citizen science". *Trends in Ecology & Evolution*, vol. 24, n. 9, pp. 467-471, 2009.

[9] François Grey. "Viewpoint: The age of citizen cyberscience". *CERN Courier*. Internet: https://cerncourier.com/viewpoint-the-age-of-citizen-cyberscience/, April 29, 2009 [Sep. 03, 2018].

[10] Jeferson S. Arcanjo et al. "Methods for evaluating volunteers contributions in a deforestation detection citizen science project". *Future Generation Computer Systems*, vol. 56, pp. 550-557, 2016.

[11] Matthew C. Hansen et al. "Humid tropical forest disturbance alerts using Landsat data". *Environmental Research Letters*, vol. 11, n. 3, p. 034008, 2016.

[12] Devis Tuia et al. "Active learning methods for remote sensing image classification". *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, n. 7, p. 2218, 2009.

[13] Steven. C. H. Hoi et al. "Batch mode active learning and its application to medical image classification", in *Proceedings of the 23rd International Conference on Machine Learning*, ACM, pp. 417-424, 2006.

[14] Priscila T. M. Saito et al. "Robust active learning for the diagnosis of parasites". *Pattern Recognition*, vol. 48, n. 11, pp. 3572-3583, 2015.

[15] Manfred K. Warmuth et al. "Active Learning with Support Vector Machines in the Drug Discovery Process". *Journal of chemical information and computer sciences*, vol. 43, n. 2, pp. 667-673, 2003.

[16] Robin Hewitt and Serge Belongie. "Active Learning in Face Recognition: Using Tracking to Build a Face Model". *Computer Vision and Pattern Recognition Workshop*, p. 157, 2006.

[17] Claudio Persello and Lorenzo Bruzzone. "Active Learning for Domain Adaptation in the Supervised Classification of Remote Sensing Images". *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, n. 11, pp. 4468-4483, 2012.

[18] Jun Li, José. M. Bioucas-Dias and Antonio Plaza. "Spectral-Spatial Classification of Hyperspectral Data Using Loopy Belief Propagation and Active Learning". *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, n. 2, pp. 844-856, 2013.

[19] André Stumpf et al. "Active Learning in the Spatial Domain for Remote Sensing Image Classification". *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, n. 5, pp. 2492-2507, 2014.

[20] Jun Li, José. M. Bioucas-Dias and Antonio Plaza. "Hyperspectral Image Segmentation Using a New Bayesian Approach With Active Learning". *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, n. 10, pp. 3947-3960, 2011.

[21] Pabitra Mitra, B. Uma Shankar and Sankar K. Pal. "Segmentation of multispectral remote sensing images using active support vector machines". *Pattern Recognition Letters*, vol. 25, n. 9, pp. 1067-1074, 2004.

[22] Begüm Demir and Lorenzo Bruzzone. "A Novel Active Learning Method in Relevance Feedback for Content-Based Remote Sensing Image Retrieval". *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, n. 5, pp. 2323-2334, 2015.

[23] Begüm Demir, Francesca Bovolo and Lorenzo Bruzzone. "Detection of Land-Cover Transitions in Multitemporal Remote Sensing Images With Active-Learning-Based Compound Classification". *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, n. 5, pp. 1930-1941, 2012.

[24] Ron Kohavi et al. "A study of cross-validation and bootstrap for accuracy estimation and model selection". *International Joint Conference on Artificial Intelligence*, vol. 14, n. 2, pp. 1137-1145, 1995.

[25] Fabian Pedregosa et al. "Scikit-learn: Machine Learning in Python". *Journal of Machine Learning Research*, vol. 12, pp. 2825-2830, Oct. 2011.

[26] Loris Copa et al. "Unbiased query-by-bagging active learning for VHR image classification". *Image and Signal Processing for Remote Sensing XVI, International Society for Optics and Photonics*, vol. 7830, 2010.

[27] Colin Campbell et al. "Query learning with large margin classifiers", in *Proceedings of the Seventeenth International Conference on Machine Learning*, 2000, pp. 111-118.

[28] Greg Schohn and David Cohn. "Less is more: Active learning with support vector machines", in *Proceedings of the Seventeenth International Conference on Machine Learning*, 2000, pp. 839-846.

[29] Robert L. Brennan and Dale J. Prediger. "Coefficient kappa: Some uses, misuses, and alternatives". *Educational and Psychological Measurement*, vol. 41, n. 3, pp. 687-699, 1981.

[30] Yann Lecun, Yoshua Bengio and Geoffrey Hinton. "Deep learning". *Nature*, vol. 521, n. 7553, p. 436, 2015.

[31] Jürgen Schmidhuber. "Deep learning in neural networks: An overview". *Neural Networks*, vol. 61, pp. 85-117, 2015.

[32] Alex Krizhevsky, Ilya Sutskever and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks". *Advances in Neural Information Processing Systems*, pp. 1097-1105, 2012.

[33] Geoffrey Hinton et al. "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups". *IEEE Signal Processing Magazine*, vol. 29, n. 6, pp. 82-97, 2012.

[34] Xavier Glorot, Antoine Bordes and Yoshua Bengio. "Domain adaptation for large-scale sentiment classification: A deep learning approach", in *Proceedings of the 28th International Conference on Machine Learning*, 2011, pp. 513-520.

[35] Liang-Chieh Chen et al. "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, n. 4, pp. 834-848, 2018.

[36] Vijay Badrinarayanan, Alex Kendall and Roberto Cipolla. "Segnet: A deep convolutional encoder-decoder architecture for image segmentation". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, n. 12, pp. 2481-2495, 2017.

[37] Jonathan Long, Evan Shelhamer and Trevor Darrell. "Fully convolutional networks for semantic segmentation", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431-3440.