ForestEyes Project: Conception, Enhancements, and Challenges[*]

Fernanda B. J. R. Dallaqua, Álvaro Luiz Fazenda, Fabio A. Faria

[1]Instituto de Ciência e Tecnologia – Universidade Federal de São Paulo (ICT-UNIFESP)
Avenida Cesare Mansueto Giulio Lattes, nº 1201 – Eugênio de Mello – SP – Brasil

{fernanda.dallaqua,alvaro.fazenda,ffaria}@unifesp.br

Abstract. *Rainforests play an important role in the global ecosystem. However, significant regions of them are facing deforestation and degradation due to several reasons. Diverse government and private initiatives were created to monitor and alert for deforestation increases from remote sensing images, using different ways to deal with the notable amount of generated data. Citizen Science projects can also be used to reach the same goal. Citizen Science consists of scientific research involving nonprofessional volunteers for analyzing, collecting data, and using their computational resources to outcome advancements in science and to increase the public's understanding of problems in specific knowledge areas such as astronomy, chemistry, mathematics, and physics. In this sense, this work presents a Citizen Science project called ForestEyes, which uses volunteer's answers through the analysis and classification of remote sensing images to monitor deforestation regions in rainforests. To evaluate the quality of those answers, different campaigns/workflows were launched using remote sensing images from Brazilian Legal Amazon and their results were compared to an official groundtruth from the Amazon Deforestation Monitoring Project PRODES. In this work, the first two workflows that enclose the State of Rondônia in the years 2013 and 2016 received more than $35,000$ answers from 383 volunteers in the $2,050$ created tasks in only two and a half weeks after their launch. For the other four workflows, even enclosing the same area (Rondônia) and different setups (e.g., image segmentation method, image resolution, and detection target), they received $51,035$ volunteers' answers gathered from 281 volunteers in $3,358$ tasks. In the performed experiments, it was possible to observe that the volunteers achieved satisfactory overall accuracy in the classification of forestation and non-forestation areas using the ForestEyes project. Furthermore, considering an efficient segmentation and a better image resolution, they can achieve outstanding effectiveness results in the classification task of recent deforestation images. Therefore, these results show that Citizen Science might be a powerful tool in monitoring deforestation regions in rainforests as well as in obtaining high-quality labeled data.*

1. Introduction

Rainforests have an important role in the global ecosystem once they have a great diversity of fauna and flora, regulate the climate and rainfall, absorb large amounts of carbon dioxide, and are indigenous housing [M]. Unfortunately, thousands of square kilometers of rainforests are deforested or degraded daily due to livestock, agriculture, urban area expansion, wood extraction, and forest fires [L].

According to one of the most well-known and successful rainforests monitoring programs - PRODES [L, H, S] - in the period between August/2018 and July/2019 the Brazilian Legal Amazon's deforestation reached $10,129km^2$, an increase of 34% in comparison with the previous period (August/2017 to July/2018) [I]. These high rates of deforestation can have irreversible and catastrophic consequences as biodiversity loss, climate change, desertification, water shortage, increase of diseases, and even emergence of pandemics [M, L, C, A].

As the conservation of rainforests is urgently needed, monitoring programs were created by government agencies and non-profit institutions. These programs use remote sensing images, image processing, machine learning techniques, and specialists' photointerpretation to analyze, identify, and quantify forest cover changes [L].

The shortage of skilled labor and the high amount of data to be analyzed is a significant challenge for information and communication technology (ICT) [S]. A possible solution to overcome this problem is to use Citizen Science wherein non-specialized volunteers collect, analyze, and classify data to solve several technical and scientific problems [G, S].

CS projects of a wide variety of research fields have drawn the attention of well-known scientific magazines such as Nature [I, G] and Science [G]. Recently, the U.S. [1] and Australian [2] governments, and the European Union [3], have launched official programs to catalog and support CS projects [B].

CS can be a valuable source of data for the Earth Observation field, which includes deforestation monitoring [F]. For CS projects ForestWatchers [L], EarthWatchers [S], and Geo-Wiki [F] the volunteers analyze and classify remote sensing images. These classifications generate deforestation maps or alerts. For the ForestWatcher project, the volunteers collect data *in situ* to confirm deforestation alerts issued by Global Forest Watch [P].

In April 2019 the ForestEyes Project [4] was launched [D]. This CS project is hosted by the Zooniverse.org platform [S] and aims to ally CS with machine learning techniques to monitor rainforests' deforestation. The volunteers analyze and classify remote sensing images. These classifications will be used to train machine learning techniques that will classify new remote sensing images. These new data can provide complementary information to official monitoring programs such as PRODES or even generate information for areas where such programs don't exist.

The feasibility of the ForestEyes project depends on the quality of the data gener-

---

[1] https://www.citizenscience.gov/
[2] https://citizenscience.org.au/
[3] https://www.ecsite.eu/activities-and-services/projects/eu-citizenscience
[4] https://www.zooniverse.org/projects/dallaqua/foresteyes

ated by the volunteers. The volunteers analyzed and classified Landsat-8's [L] remote sensing images from an area of BLA, specifically of the State of Rondônia, wherein PRODES-based data was used as validation of the volunteers' contributions. It is worth mentioning that to ensure this validation, the volunteers analyzed the same Landsat-8 images used by PRODES.

This paper describes in detail all of the steps taken to validate the generated data, including the creation of the tasks in the Zooniverse.org platform, the remote sensing image processing techniques, the launch of the CS's campaigns, the user layout continuous enhancements, and the evaluation of the volunteers' contributions. The remainder of this work is organized as follows. Section 2 presents background about deforestation monitoring and CS. Section 3 presents the proposed ForestEyes Project. Section 4 shows the methodologies and some results for the CS's campaigns, whereas Section 5 presents the comparison between the campaigns and analyses of the volunteers' contributions. Section 6 presents the conclusions, and Section 7 defines the future steps for ForestEyes Project.

2. Background

This section presents background about deforestation detection/monitoring and CS projects.

2.1. Deforestation Detection and Monitoring

The monitoring of rainforests is necessary because of their great biodiversity. Besides the well-known Amazonian Forest, other rainforests also suffer from indiscriminate and progressive deforestation, where over 74 million hectares have been lost or heavily degraded from 2000 to 2012 [L, H]. For BLA, there are monitoring programs - governmental and from non-profit institutions - for deforestation measuring and preventing as PRODES and DETER [D] from INPE, MapBiomas [S] (multi-institutional initiative), SAD from Imazon [d] and GLAD from Global Forest Watch [H].

2.1.1. PRODES

Amazon Deforestation Monitoring Project (PRODES) [L, S] was developed by INPE in 1988 to carry annual deforestation surveys in BLA. Trained specialists classify remote sensing images through photointerpretation. These remote sensing images are usually from Landsat satellites, but other satellites as SENTINEL-2 and CBERS-4 can also be used. By 2013, Landsat-8 was launched, which has a revisit time of 16 days, 30 meters of spatial resolution, and 11 spectral bands.

PRODES methodology is divided into three steps: (1) an image selection technique is performed, where cloud-free images collected near August 1st are enhanced to highlight the areas with complete removal of native forest (clear-cutting); (2) the specialists identify and delimit the deforestation polygons; and (3) the annual deforestation rate is calculated [S]. Both thematic images and quantitative data are publicly available. Figure 1 presents the PRODES' mosaic for the State of Rondônia for the year 2016.

As PRODES quantifies and spatially locates annual BLA deforestation increments, once an area is classified as deforestation, it is included in an exclusion mask

that will be used in subsequent years. Therefore PRODES data don't infer on regenerated areas [W]. This exclusion mask also has regions of different vegetation (which PRODES classifies as non-forest) and hydrography areas.
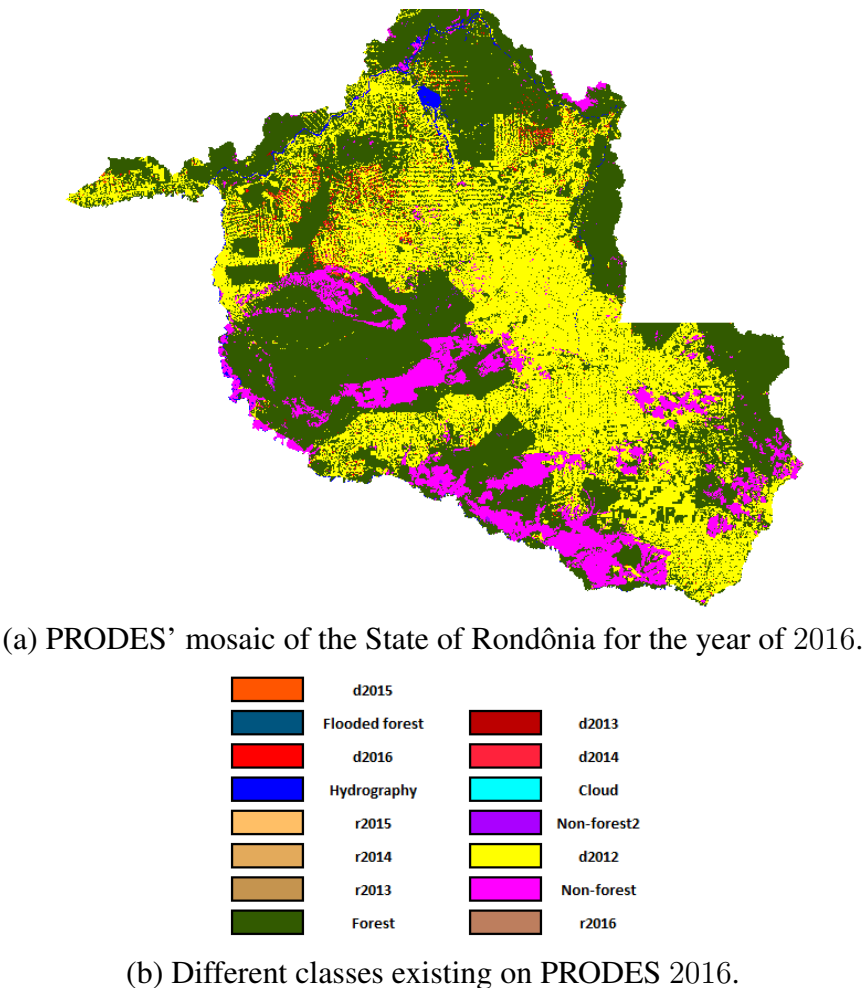

(a) PRODES' mosaic of the State of Rondônia for the year of 2016.

(b) Different classes existing on PRODES 2016.

Figure 1. Rondônia's thematic image for the PRODES year of 2016**. Extracted from TerraBrasilis [d]. In (a) Thematic image of Rondônia in the year** 2016 **and in (b) Legend for the thematic image where "d" means deforestation of the given year and "r" is the residue. The residue is deforestation that was not detected but occurred in previous years [M].**

2.1.2. DETER

The monitoring program DETER was created to detect near-real-time deforestation as quick actions are needed to stop the beginning of a deforestation process. It began by using imagery from MODIS (Moderate Resolution Imaging Spectroradiometer) [J] with a spatial resolution of 250 meters and a revisit time of roughly 1.5 days. An alert is sent to authorities when a deforestation process is identified. These alerts are also posted every month on the program's website. Nowadays, DETER also uses images from other satellites with different spatial resolutions to improve deforestation detection [D].

2.1.3. TerraBrasilis

Recently INPE deployed a new platform, called TerraBrasilis [d], which provides an interactive dashboard in which users can access and use spatial data generated by PRODES and DETER. Also, the data generated by INPE's monitoring programs are used by other programs as TerraClass [A, N] and MapBiomas (Brazilian Annual Land Use and Land Cover Mapping Project) [S], for instance. The project TerraClass was launched in 2011 through a partnership between INPE and EMBRAPA (Brazilian Agricultural Research Corporation) to produce maps showing land use and the cover of the deforested areas mapped by PRODES [A].

2.1.4. MapBiomas

MapBiomas project began in 2015, and it is a multi-institutional initiative to generate annual land use and land cover time series for all Brazilian territory, covering all six official Biomes [S]. Recently MapBiomas has launched an alert system from information generated by DETER, SAD, and GLAD to validate and refine data from daily high-resolution images with a spatial resolution of $3m$, creating surveys to be sent to different authorities.

2.1.5. SAD and GLAD

The monitoring programs SAD (Deforestation Alert System) and GLAD (Global Land Analysis & Discovery) are held by the non-profit institutions Imazon (Institute for humanity and amazon environment) and GFW (Global Forest Watch), respectively. The first analyzes MODIS images with Normalized Difference Fraction Index (NDFI) and provides surveys every month on its website [d]. The second program uses Landsat images and decision trees to alert tree-cover loss not only in BLA but also in other areas with rainforests as Congo Basin and Southeast Asia [H].

2.2. Citizen Science (CS)

The collection, analysis, and classification of data for scientific research through volunteers' contributions (most of them are non-specialist) is known as CS [G, S, I]. The remotest idea of CS goes back more than 2000 years, where people helped to monitor migratory locusts outbreaks destroying harvest in ancient China [I].

The pioneer project called Christmas Bird Count [S] created by ornithologist Frank Chapman in 1900 proposes carrying out a census of the bird population as an alternative to hunting birds on Christmas. Currently, the project is held every year by the National Audubon Society. Besides ecology, CS nowadays is being used in a great variety of research fields as astronomy [F, W], chemistry [C], mathematics [C], and physics [A].

Advances and worldwide growing popularity for ICT resources remotely distributed (e.g., computers, tablets, and mobile phones) helped to increase the dissemination of CS projects [S]. CS can also be referred to as Citizen Cyberscience, which can be divided into three subcategories: Volunteered Computing, Volunteered Thinking, and Participatory Sensing [H].

- **Volunteered Computing**: a distributed processing power is provided by several volunteers donating their computational resources. The first and best-known example of Volunteered Computing project is SETI@home [A], launched in May 1999 by the UCLA/Berkeley. To participate, the volunteers execute a free program that downloads and analyzes radio telescope data, searching for extraterrestrial intelligence. LHC@home [L] and climateprediction.net [S] are other examples of projects.
- **Volunteered Thinking**: the volunteers use their cognitive skills to analyze data or to perform different kinds of tasks. Galaxy Zoo, launched in 2007, is a well-known example of this kind of CS, requiring, for more than a decade, volunteers to classify images of galaxies according to patterns available [L]. Within 24 hours of launch, they received almost $70,000$ classifications in one hour. More than 50 million classifications were received during its first year by more than $150,000$ people. With the success of Galaxy Zoo, it evolved into Zooniverse.org, an online platform that hosts a wide variety of Volunteered Thinking projects. Until 2013, it collected more than 300 million analyses from 1 million volunteers [S]. FoldIt [C], Stardust@home [W], and Planet Hunters [F] are other examples of Volunteered Thinking projects.
- **Participatory Sensing**: the volunteers are responsible for gathering data or information, usually *in situ*. The eBird project, released in 2002 by Cornell Lab of Ornithology and National Audubon Society, aims to engage citizen scientists to report bird observations through standardized protocols, documenting bird distribution, habitat, and patterns [S, S]. Other examples of participatory sensing projects are CoralWatch [M] and NoiseTube [M].

The motivation to volunteer contributions to a CS project was addressed by several authors. Altruism, for example, was cited as a major factor for CS projects related to nature conservation [C, B, K]. For projects on different purposes, like Galaxy Zoo and FoldIt, the main reasons were research contribution and the interest in science [R, C]. Other important factors are the importance of online communities [R, H] and competitiveness [H] through a volunteers' ranking. For the eBird project, for example, a substantial increase in data gathered, collecting more observations in one-month than in two years' project, appears after some changes that included competition between volunteers [H].

Many CS projects have shown that citizen scientists can produce high-quality data, being as efficient as specialists [C, A, G, K, D]. However, statistical data analyses and validation heuristics or algorithms are also important to discard bad or non-reliable data received from volunteers, ensuring good data quality. To send redundant tasks to multiple volunteers, allowing to extract the information that constitutes the majority or consensus between the volunteers, is widely used as a good data quality mechanism.

Calibration tasks containing already classified samples by specialists could also be used as a quality mechanism since they allow a volunteers and specialists comparison. In Galaxy Zoo, for example, only registered volunteers can participate, assigning weights to individual users according to the volunteer's skill [R, L, A].

To overcome the data quality challenge in CS projects, l [L] proposes a method using volunteers' diversity to improve data accuracy by a Bayesian inference algorithm. A survey, gathering information about the volunteer's education level and the motiva-

tion to CS, is used to group the volunteers into distinct classes. This method shows an improvement in data accuracy compared to majority voting.

2.3. ForestWatchers Project

The ForestWatchers was a CS project to monitor the rainforests through the volunteer's collaboration connected to the Internet [A, L]. It was launched in 2012 by researchers of the Institute of Science and Technology of Federal University of São Paulo (ICT/UNIFESP), the Laboratory for Computing and Applied Mathematics at the National Institute for Space Research (LAC/INPE), and the Citizen Cyberscience Centre (CCC).

The project was composed of three experimental applications: (1) Best-Tile; (2) Deforestation Classification; and (3) Correct Classification. The Correct Classification application used PRODES classification and remote sensing images from MODIS sensor to train an artificial neural network (ANN) [H]. New remote sensing images were analyzed by this ANN, creating segmented images in Forest (represented by green pixels) and Non-Forest (displayed as red pixels). Tiles with size 3×3 pixels - corresponding to an area of approximately 56 hectares - with confidence below a given threshold were sent for volunteers' inspection where they agreed or disagreed with the ANN's classification [L].

Two campaigns were performed, each one with an area of BLA: one from the State of Rondônia for the year of 2011, with 72 tasks; and the other was from Awá-Guajá Indigenous Reserve in 2014, with 26 tasks.

3. The Proposed ForestEyes Project

The ForestEyes Project [D], a CS project inspired by ForestWatchers Project and launched in April/2019, aims to monitor deforestation areas, where volunteers classify remote sensing image's segments that can be used as training sets of classification algorithms. These classification algorithms are used to label new segments, providing complementary information to official monitoring programs or even generating data for areas where do not exist this kind of programs.

The project can be divided into five modules: (1) Pre-processing; (2) Citizen Science; (3) Organization and Selection; (4) Machine Learning; (5) and Post-processing. A schematic representation of these modules can be seen in Figure 2 where the colored items are the ones discussed in this paper.
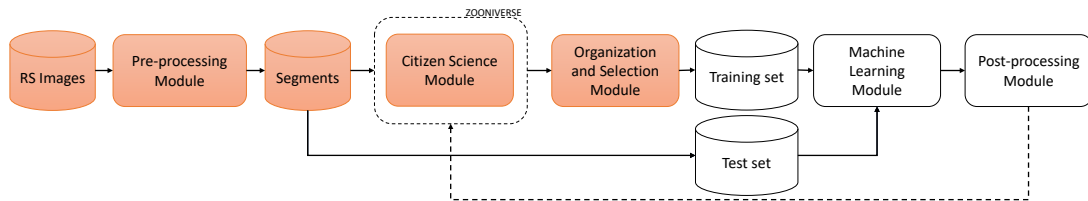

Figure 2. ForestEyes Project's schematic representation. The colored items are the modules discussed in this paper.

3.1. Pre-processing Module

The Pre-processing Module is responsible for the acquisition, processing, and segmentation of the remote sensing images. Figure 3 shows the steps (a) to (g) of the Pre-processing
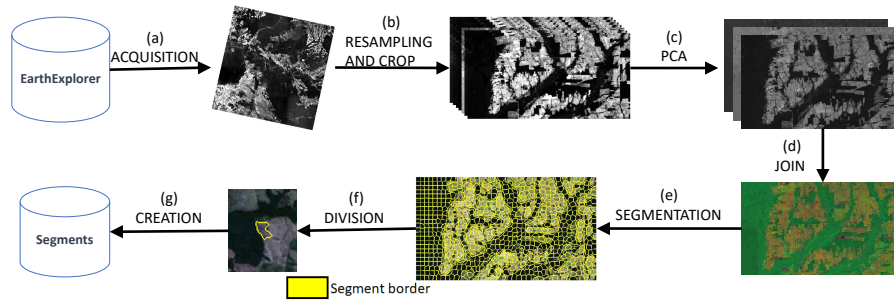
Figure 3. Schematic representation of the Pre-processing Module steps in the ForestEyes Project.

Module. Firstly, in (a), the acquisition of the remote sensing images is performed. Currently, the images used by the project are from Landsat-8, the same ones that are analyzed by PRODES. The Landsat-8 scenes are freely collected at EarthExplorer platform[5]. Landsat-8 has 11 spectral bands but only 7 bands - coastal aerosol or ultra-blue, blue, green, red, near-infrared or NIR, shortwave infrared 1 or SWIR 1, and shortwave infrared 2 or SWIR 2 - are acquired. In (b), the region of interest is cropped and resampled to standardize the remote sensing image according to the PRODES' mosaic.

Generally, some segmentation algorithms have a 3-band image (RGB) as input, and the acquired image is composed of 7 bands. In (c), a dimensionality reduction technique (Principal Component Analysis - PCA) is applied to generate 3 grey color images that are joined into one new 3-band image, simulating RGB image in (d). This new image is used as input to a segmentation algorithm in (e). In this step, 3 segmentation algorithms were used: (1) Simple Linear Iterative Clustering (SLIC) [A]; (2) Image Foresting Transform-SLIC (IFT-SLIC) [A]; and (3) MaskSLIC [I]. In (f), each segment becomes a sample included in the database in (g). Some of the samples are sent for the Citizen Science Module, and the remaining samples compose the test set of the Machine Learning Module.

3.2. Citizen Science Module

At the Citizen Science Module is performed the interaction with the volunteer, and The Zooniverse.org platform was chosen to host this module once it has a large and consolidated community. This platform allows the creation of one or more workflows, defined as the sequence of tasks to be handled by the volunteers. On the ForestEyes Project, the volunteers analyze and classify the segments into three classes (Forest, Non-Forest, or Undefined), according to the proportion of pixels that belong to each class. If 70% or more pixels inside the segment belong to Forest or Non-Forest, the volunteer needs to assign this majority class. Otherwise, the volunteer needs to assign as an Undefined class. The segments are presented with different color-compositions, and a tutorial is available to volunteers.

After 15 or more responses received for each segment from distinct volunteers, the workflow is considered finished, and the contributions are downloaded from the Zooniverse.org platform. These contributions are sent for analysis at the Organization and Selection Module.

---

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

3.3. Organization and Selection Module

At the Organization and Selection Module, the class of each segment is defined by the majority vote of the volunteers' answers, and analyses are performed such as the tasks' difficulty level (Section 3.3.3), consensus convergence (Section 3.3.4), and the volunteers' hit rates and score (Section 3.3.5).

The Organization and Selection Module is also responsible for the comparison between volunteers' classifications with a given groundtruth when available. For now, PRODES data were used to build different groundtruth that will be explained in Section 3.3.2. For the groundtruth, it was needed the definition of a metric called Homogeneity Ratio (HoR), presented in Section 3.3.1.

With the classification of the segments, it is possible to create the training set of the Machine Learning Module. Only the segments with classes Forest and Non-Forest are considered, discarding the Undefined segments.

3.3.1. Homogeneity Ratio (HoR)

The ForestEyes' volunteers choose between three classes: Forest, Non-Forest, and Undefined. However, a pixel-based groundtruth is built as a binary classification (Forest and Non-Forest). This Non-Forest class encompasses everything that is not Forest, such as deforestation, hydrography, clouds, and others.

To evaluate the quality of the segments, a Homogeneity Ratio (HoR) is defined as the percentage of pixels present inside each segment of the majority class in a binary classification task (Forest and Non-Forest). According to equation (1), NFP represents the number of Forest pixels, NNP the Non-Forest pixels' number, and NP the total pixels for a segment ($NP = NFP + NNP$).

$$HoR = \frac{\max(NFP, NNP)}{NP} \tag{1}$$

The HoR metric is partitioned in five different ranges to better represent their distribution: $0.5 \leq HoR < 0.6$, $0.6 \leq HoR < 0.7$, $0.7 \leq HoR < 0.8$, $0.8 \leq HoR < 0.9$, and $HoR = 1.0$. Therefore, if a segment has $0.5 \leq HoR < 0.6$ to Forest class means that it has $0.5 \geq HoR$ to Non-Forest class.

3.3.2. Groundtruth

To compare the classification of the volunteers with PRODES data, two strategies were used to define the groundtruth. The first one is pixel-based. Given an image from PRODES mosaic (Figure 4(a)), all PRODES' classes different of Forest (deforestation, residue, hydrography, clouds, non-forest vegetation, and others) are defined as Non-Forest, resulting in a binary image (Figure 4(b)) that will be referred as GT-PRODES.

The second strategy of groundtruth is segment-based, where two new GTs are created: GT with Undefined (GT-U) and GT with the majority (GT-M). These two GTs are created by considering HoR for each segment.
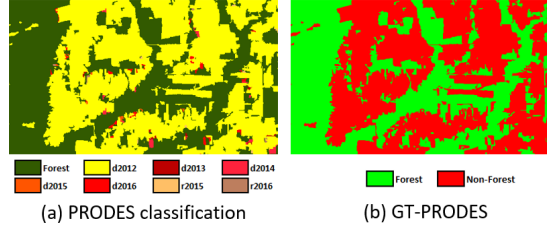
Figure 4. Example of PRODES binarization where the resulting binary image becomes the GT-PRODES. The image presents an area in the State of Rondônia of approximately $64,000$ **hectares.**

With GT-U, the segments with $HoR < 0.7$ are classified as Undefined, otherwise ($HoR \geq 0.7$) receive the respective Forest or Non-Forest label. As for GT-M the segment is classified considering its majority class [D]. Figure 5 presents GT-U and GT-M for the same region depicted in Figure 4.
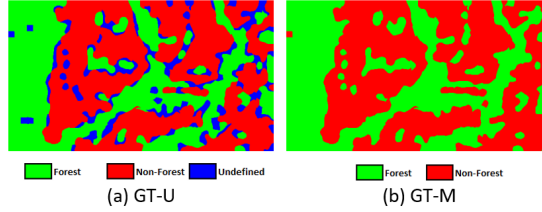

Figure 5. Examples of GT-U and GT-M for the same region.

The volunteer classification accuracy is defined here by equation (2). Let NSC be the number of samples correctly classified and TNS the total number of samples. For GT-PRODES, the samples are just the pixels, and for GT-U and GT-M, the samples are the segments.

$$Accuracy = 100 \times \frac{NSC}{TNS} \tag{2}$$

3.3.3. Task's Difficulty Level

According to c [A], the task's difficulty level can be calculated using the volunteers' answers and Shannon's entropy (HE) [L], which is given by equation (3). The higher the entropy value, the greater the difficulty of the task.

$$HE = -\sum_{i=1}^{N} p_i \times \log_2 p_i \tag{3}$$

Where p_i is the probability of the class i be chosen, calculated by the ratio between the number of votes given to class i and the total of votes for the task, and n is the number of possible classes in the task.

The tasks' entropy values are normalized to the range $(0, 1)$ and empirically separated into three difficulty levels: Easy ($HE \leq 0.33$); Medium ($0.33 < HE \leq 0.66$); and

Hard ($HE > 0.66$).

The average time to the task accomplishment by the volunteers is also taken to be used to measure the difficulty level. This metric could be correlated to entropy, accuracy, or HoR, showing different task hardness applied to different volunteer groups and CS campaigns.

3.3.4. Consensus Convergence

The analysis of the consensus convergence is used to evaluate the number of answers needed in each task to define its final classification. This analysis is performed by checking if the classification given with the predefined number of answers remains the same with fewer answers. For the ForestEyes Project, the first 15 answers define the final classification of each task. Therefore the consensus convergence analysis was considering the achieved classifications for 5, 7, 9, 11, and 13 answers.

3.3.5. Volunteers' Hit Rates and Score

As in c [A], it is assumed that the answer given by volunteers' consensus is the correct one. The number of hits of each volunteer is incremented by 1 when the volunteer agrees with the canonical answer and remains unchanged if it doesn't agree. The volunteer's hit rate (HR) and score (VS) are calculated through equations (4) and (5), where hits mean the number of answers the volunteer agreed with the consensus and total_answers is the number of answers that the volunteer gave.

$$HR = 100 \times \frac{hits}{total\_answers} \tag{4}$$

$$VS = (0.3 \times total\_answers) + (0.7 \times hits) \tag{5}$$

Besides evaluating the volunteers' answers in comparison with the consensus, (4) can also be used to evaluate the volunteers' answers with the GT bases GT-U and GT-M.

3.4. Machine Learning Module

In the Machine Learning Module, a classification algorithm is trained with the set built from the Organization and Selection Module. Then, this algorithm is applied to a test set that consists of segments that were not sent to the Citizen Science Module.

3.5. Post-processing Module

The Post-processing Module will evaluate the obtained results from the Machine Learning Module, creating deforestation alerts, for example. Also, it is planned a feedback cycle where test segments will be sent to the Citizen Science Module to improve the training set and consequently improve the classification algorithm.

4. Citizen Science's campaigns

In this section are shown the methodologies and some results of the CS's campaigns. Each subsection describes one single campaign, also known as a workflow in the Zooniverse.org platform, detailing the image used, segmentation algorithms, and user interface.

4.1. Beta-Review

To launch a CS project as a Zooniverse's project, a Beta-Review is needed. The interface is reviewed by volunteers who evaluate the project and provide improvement suggestions. Based on this evaluation, the platform's team defines if the project is ready for an official launch [D].

The ForestEyes' Beta-Review consisted of a workflow with the same tasks as the ForestWatchers' Correct Classification plus six more tasks from Awá-Guajá as it is needed a minimum of 100 tasks for a project to be sent for Beta-Review. The volunteers needed to analyze tiles highlighted by red squares with size 3×3 pixels from remote sensing images obtained by MODIS sensor, which has a spatial resolution of $250m$. Figure S1 shows the user interface for this Beta-Review.

Contrary to the ForestWatchers Project, which showed the original remote sensing image and the image classified by an automatic procedure, this Beta-Review only showed images from the MODIS sensor. Furthermore, a new answer was created (Undefined class), aiming to identify tasks where there is too much mixture of Forest and Non-Forest pixels (border regions), usually resulting in a low HoR.

The Beta-Review received $2,275$ answers being $1,604$ answers from 77 registered volunteers and 671 answers from 125 anonymous volunteers. For Awá-Guajá 2014, 31 tasks were classified as Forest, and just one as Non-Forest whereas for Rondônia 2011, 65 tasks were classified as Non-Forest, 4 as Forest and in 3 tasks there were ties. Table 1 presents the frequency of tasks over difficulty levels based on entropy ranges.

Table 1. The frequency of the tasks with difficulty level for the Beta-Review workflow in the ForestEyes Project.

Difficulty Level	Easy (E)	Medium (M)	Hard (H)
Entropy (*HE*)	$HE \leq 0.33$	$0.33 < HE \leq 0.66$	$HE > 0.66$
Awá-Guajá 2014	18 (56.25%)	12 (37.5%)	2 (6.25%)
Rondônia 2011	4 (5.56%)	26 (36.11%)	42 (58.33%)

Notice that the volunteers had high consensus with Awá-Guajá tasks, since it was necessary only to compute 5 answers to achieve the same results as computing the first 15, and 93.75% of the tasks were considered easy or medium difficulties. As for Rondônia, the majority of tasks (58.33%) were considered hard, demanding 9 or more answers to achieve a consensus convergence higher than 90%.

Although the Rondônia tasks could be considered a challenging task for the volunteers, their classification results were 88.9% when compared with PRODES classification (GT-PRODES). As feedback from the volunteers, two problems have been reported: (1) image resolutions, which can be observed in Table 1 with a high number of hard tasks;

and (2) how tasks were presented to volunteers. So new workflows were created taking into account their feedback and will be explained in the following.

4.2. First Official Workflows

In the first official workflow for the ForestEyes Project at the Zooniverse.org platform, the remote sensing images from the Landsat-8 satellite were used, which allow a better resolution than MODIS sensors. This image covers an area in Rondônia in the year of 2016, and was acquired at EarthExplorer containing 7 out of 11 bands. However, this image was resampled from resolution $30m$ to $60m$, matching with the classified image by PRODES [D]. The size of the generated image is 558×318 pixels corresponding to an area of approximately $64,000$ hectares.

Besides using an image with better resolution, a new interface to display the tasks was presented as required by volunteers. So, instead of using fixed square tiles, it was decided to use segments generated by the SLIC algorithm. Therefore, the steps of the Pre-processing Module were developed as explained in Section 3.

For the user interface, presented in Figure S2, two different color composition images were available as in Figure 6. In (a) is the original Landsat-8's RGB, and in (b) is a false-color image composed of the bands 7 (shortwave infrared 2), 5 (near-infrared), and 3 (green) from Landsat-8. This new composition will be referred to as 753 false-color composition.
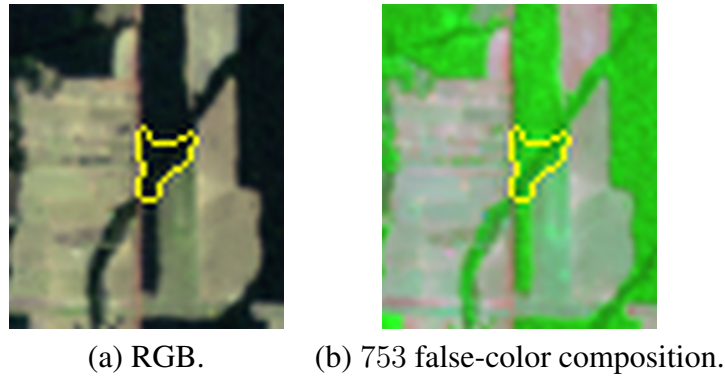

(a) RGB.   (b) 753 false-color composition.

Figure 6. Examples of color compositions available to volunteers.

With this new workflow, the ForestEyes Project was officially launched in April 18, 2019, where just one week later, all $1,022$ tasks have been completed, receiving $19,803$ answers where $17,239$ answers were from 151 registered volunteers and the remaining from 76 anonymous volunteers.

The volunteers classified $1,022$ tasks being 467 Forest, 467 Non-Forest, 73 Undefined, and 15 tasks tied that were randomly chosen, resulting in a GT-PRODES accuracy of 84.15%, GT-U and GT-M accuracies were of 84.25% and 86.50%, respectively.

The second workflow was created with an image of the same area acquired in another year (2013), where its user interface can be seen in Figure S3. This workflow was also completed in almost a week, reaching $15,610$ answers for $1,028$ tasks, with 137 registered volunteers giving $14,578$ answers and 41 anonymous providing 852 contributions.

These $1,028$ tasks were classified as 478 being Forest, 470 Non-Forest, 74 Undefined, and ties happened in 6 tasks, resulting in a GT-PRODES accuracy of 83.76%, GT-U and GT-M accuracies of 84.34% and 88.33%, respectively.

Performed analyses over the results indicated that registered volunteers contributed more than the anonymous. Also, they achieved better accuracies either considering the volunteers' consensus and the groundtruth hit rates (GT HRs), providing more homogeneous answers than the anonymous volunteers. Furthermore, it is possible to observe that the consensus and GT HRs increase as volunteers answered more tasks, stabilizing with a reasonable value [D].

Different from Rondônia 2011, where the majority of tasks were hard, in these new workflows with a better image resolution, around 65% of the tasks were considered easy, with just 13.8% being hard, corresponding to 283 tasks. Also, the consensus convergence for both workflows (2016 and 2013) achieved more than 90% with only 5 answers.

The comparison of the volunteers' classifications for 2016 and 2013 could show the areas where occurred deforestation between those years. Therefore the difference between them was taken and compared with the difference between GT-PRODES 2016 and GT-PRODES 2013.

The difference between the two GT-PRODES showed that $2,184$ pixels correspond to deforestation between the years of 2014 and 2016. However, from the difference between the two workflows, only 570 were correctly classified as Non-Forest by the volunteers, 302 were labeled as Undefined, in 176 tasks happened ties, and 115 were wrongfully classified as Forest. These results may have happened due to errors in the image segmentation process, failures in the tasks display, variability in the spectral satellite image characteristics, or volunteers' error.

Intending to get results even closer to GT-PRODES, especially for the detection of areas with recent deforestation, k-means clustering was performed on the tasks considered hard (high entropy), Undefined, and the ties, resulting in a new workflow that is explained next.

4.3. Refining Segmentation Process through Clustering

From the results obtained by l [D], and depicted in Section 4.2, the Undefined tasks, the tasks with ties, and the hard segments ($HE > 0.66$), received a new segmentation process aiming to refine the polygons/regions. This segmentation turns each original task into two or more samples, which were organized in a new workflow at the Zooniverse.org platform.

The k-means [L] algorithm was used to refine the depicted segments, using the PCA in the pre-processed image, and the elbow method [K] to define the optimal number of clusters. An example of the new segmentation process can be viewed in Figure 7.

Since the k-means algorithm does not take into account the spatial pixel distribution, distant pixels located outside a new segment are also considered an integrated part of the same cluster. Therefore, the number of created news segments could be greater than k. Some of them could also be too small to visualize.

To make the interface and image analysis more friendly, the segments with less than 9 pixels were discarded, reaching $1,285$ new tasks sent to volunteers where 637

ORIGINAL TASK

NEW TASKS

Figure 7. Examples of new tasks for the new workflow. The original task was divided into four new segments that were sent to volunteers.

are related to 2013 and 648 from 2016. The segments had an average size of 35 pixels, representing a minor area than the previous workflows, which had 173 pixels on average.

The user interface, presented in Figure S4, showed two zoomed images side-by-side representing the available color compositions, 753 false-color and RGB. Also, due to possible problems with the segments' size, a new answer option was available to volunteers ("Segment too small"), making possible a more detailed quality evaluation of the new tasks.

From the $1,285$ tasks and considering the first 15 answers, 598 received Forest label, 621 received Non-Forest, 29 were classified as Undefined, 7 was considered too small, and draws happened in 30 tasks. The workflow received $19,496$ answers, being $18,789$ from 142 registered volunteers, and 707 from 40 anonymous. For better comprehension, the results were divided into the ones regarding 2016 (defined here as refined2016) and the ones from 2013 (defined here as refined2013).

For refined2013, 312 segments were labeled as Forest, 283 as Non-Forest, 20 as Undefined, 4 as Small, and in 18 happened ties, resulting in accuracy of 67.50% for GT-PRODES, 59.34% for GT-U, and 67.19% for GT-M. Whereas for refined2016, 286 tasks were classified as Forest, 338 as Non-Forest, 9 as Undefined, 3 as Small, and in 12 happened ties. For the GT-PRODES accuracy was 71.91%, GT-U reaches 62.65% and GT-M 72.68%.

Although the results didn't achieve satisfying values, being very bellow compared to the other workflows, when integrating these new classifications to the first official workflows described in Section 4.2, the previous results from the first workflows regarding GT-PRODES are increased by more than 4%. It also increased the monitoring of recent deforestation by approximately 85.9%. However, improvements were still required since, even with these optimizations, just 48% of GT-PRODES pixels with recent deforestation

were found.

4.4. New Image Resolution and Segmentation Algorithm

By 2017, the PRODES' mosaics started to have a resolution of $30m$, enabling the use of the Landsat-8 images without needing resampling. The use of images with better resolution could improve the user visualization and consequently the volunteers' classification.

A Landsat-8 image for a similar region defined for the previous workflows was collected, as close as possible. This image has 1058×625 pixels, corresponding to an area of approximately $59,500$ hectares, and is the same used to build PRODES mosaic of 2017.

To improve the segmentation, another algorithm (IFT-SLIC) was tested. This algorithm is an extension of SLIC in an IFT framework [A]. Both SLIC and IFT-SLIC were applied to the new remote sensing image, and two new workflows were created at the Zooniverse.org platform: SLIC 2017 and IFT-SLIC 2017. Both had the same user interface that can be seen in Figure S5.

By analyzing the HoR's distribution ranges of each algorithm, SLIC presented higher percentage of segments with $HoR = 1.0$ (43.6%) and higher percentage of segments with $HoR \geq 0.7$ (90.2%) compared to IFT-SLIC that had 39.1% and 87.1%, respectively.

For SLIC 2017, $14,339$ answers were received from 64 registered volunteers and 359 answers from 19 anonymous volunteers, resulting in $14,698$ answers for the 973 tasks, where 417 were classified as Forest, 450 as Non-Forest, 104 as Undefined, and in 2 tasks happened ties. The accuracy were 83.63% for GT-PRODES, 87.87% for GT-U, and 87.05% for GT-M. With IFT-SLIC 2017, $15,360$ answers were received, where $14,977$ from 79 registered volunteers, and 383 from 20 anonymous users. The $1,008$ tasks were classified with 448 as being Forest, 443 as Non-Forest, and 117 Undefined, resulting in a GT-PRODES accuracy of 81.39%, GT-U of 86.31%, and GT-M of 85.62%.

SLIC 2017 presented better results regarding all GTs compared with IFT-SLIC 2017. This could be explained by SLIC's higher percentage of homogeneous segments ($HoR \geq 0.7$ and $HoR = 1.0$), which may have helped the volunteers in the classification task.

4.5. Improving the Detection of Areas with Recent Deforestation

As the project proposes to monitor rainforests' deforestation, detecting areas with recent deforestation is the most important challenge to overcome. As seen in the refined workflow (Section 4.3), only 48% of recent deforestation was found, requiring an enhancement process.

Firstly, to exclusively monitor areas with recent deforestation, it is needed to segment images only in non-consolidated deforestation areas detected by PRODES or deforested areas undetected by previous campaigns.

In this new workflow, a segmentation algorithm called MaskSLIC [I], which allows to segment regions outside a mask over the image, was adopted. A mask from the PRODES 2017 image was created, taking into account only a

period of deforestation before August/2016. These pixels with recent deforestation (encompassing the deforestation between August/2016 and July/2017) were the targets to be detected in this new campaign.

The number of segments was defined as the number of pixels outside the mask divided by 70 pixels, which approximately corresponds to the minimum size area of deforestation monitored by PRODES (6.25 hectares) [G]. So, the final number of segments was $4,466$.

As sending all these segments to the volunteers would be impractical, an analysis using the groundtruth and HoR values was performed to build this new workflow. Hence, all the Non-Forest and Forest segments with $HoR < 1.0$, and some Forest segments with $HoR = 1.0$ were selected, resulting in 92 tasks.

Remaining vegetation are commonly found in the Amazon deforestation process, which hinders the recent deforestation detection [O] not only for automatic algorithms but also for people. To help the volunteers' classification task, an NDVI [G] image was also displayed in the interface. This vegetation index is calculated as a ratio between bands red and NIR, given by Equation (6). It results in a grayscale image, where the darkest pixels can be water, clouds, rock, or sand; moderate pixels can be grassland or shrub, and the brightest pixels (closest to white) represent dense green leaves. Figure 8 shows an example of images displayed to volunteers in this new ForestEyes' workflow with the MaskSLIC algorithm, representing an area with recent deforestation, and Figure S6 presents the user interface of this workflow.

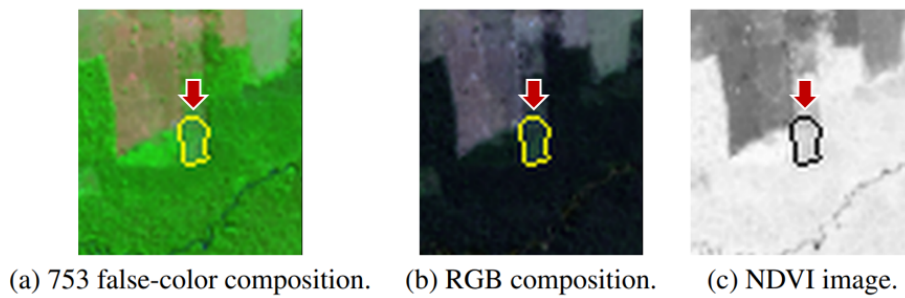$$NDVI = \frac{NIR - Red}{NIR + Red} \qquad (6)$$

(a) 753 false-color composition.   (b) RGB composition.   (c) NDVI image.

Figure 8. Example of task for the new ForestEyes's workflow with MaskSLIC. The images show a recent deforested area.

This new workflow received $1,481$ answers from 28 volunteers where $1,476$ answers from 26 registered volunteers and 5 answers from 2 anonymous. In a total of 92 tasks, 53 were classified as Forest, 23 as Non-Forest, and 16 as Undefined. The resulting accuracies were 75.65% for GT-PRODES, 85.87% for GT-U, and 80.43% for GT-M.

Considering just recent deforested areas, the scheme presented in this subsection shows an improvement in the detection of areas with recent deforestation achieving 65.95% for GT-PRODES accuracy wherein the previous workflows reached only 48%. However, the images taken for both workflows are not the same, differing by years, for almost the same coverage area.

5. Comparison Between Workflows and Discussions

This section shows different comparative analyses between workflows. Firstly, the general information is presented regarding the number of tasks, volunteers, and answers for each workflow. Next, accounting for the workflows tasks' contributions, the consensus convergence, and CS's accuracy for the different GTs are evaluated and analyzed. The tasks' difficulty levels by entropy and answers' average time are also calculated and correlated. Finally, the volunteers' ranking is built according to the different scores computed from all workflows.

5.1. General Information

Table 2 shows the general information for the ForestEyes' official workflows, where the registered volunteers had significant participation, even reaching 295 times greater than the anonymous answers for the MaskSLIC workflow.

Table 2. Summary of the ForestEyes' workflows.

Workflow	2013	2016	refined 2013 & 2016	2017 SLIC	2017 IFT-SLIC	2017 MaskSLIC
Number of Tasks	1,028	1,022	1,285	973	1,008	92
Registered volunteers	137	151	102	64	79	26
Anonymous volunteers	41	76	40	19	20	2
Total of volunteers	178	227	142	83	99	28
Registered Answers	14,758	17,239	18,789	14,339	14,977	1,476
Anonymous Answers	852	2,564	707	359	383	5
Total of Answers	15,610	19,803	19,496	14,698	15,360	1,481

For the next results, the refined workflow 2013&2016 was divided into the tasks from 2013 (refined2013) and the tasks from 2016 (refined2016).

5.2. Consensus Convergence

The consensus convergence by the number of redundant contributions, defined in Section 3.3.4, was taken for the ForestEyes' official workflows, resulting in Table 3. For all workflows, except refined2013, the convergence for just 5 answers is higher than 90%. For SLIC and IFT-SLIC, almost 97% of the tasks have the same classification with 5 and 15 answers. Furthermore, only MaskSLIC presents a decrease when the considered number of answers reaches 13, but the convergence's percentage remains high, with almost 95%.

For future workflows is planned to use fewer answers than the predefined 15. The first attempt will be with 11 answers.

5.3. Citizen Science's Accuracy

With the 3 GTs (pixel-based GT-PRODES, segment-based GT-U and GT-M) defined in Section 3.3.2, the classifications performed by ForestEyes' volunteers achieved the results presented in Table 4.

Table 3. Consensus Convergence for ForestEyes' workflows.

Workflows	Number of Answers				
	5	7	9	11	13
2013	94.75%	95.72%	96.00%	97.47%	98.54%
2016	90.50%	93.44%	95.30%	96.80%	98.14%
refined2013	88.23%	90.11%	90.11%	93.41%	96.08%
refined2016	91.36%	94.44%	95.68%	96.60%	97.84%
SLIC	96.81%	97.74%	98.15%	99.08%	99.08%
IFT-SLIC	97.32%	98.21%	98.51%	99.21%	99.60%
MaskSLIC	91.30%	93.48%	95.65%	96.74%	94.56%

Table 4. Overall accuracies for each workflow and GT in the ForestEyes Project.

Resolution	Workflows	GT-PRODES	GT-U	GT-M
		(Pixel-based)	(Segment-based)	
60m	2013	83.76%	84.34%	88.33%
	2016	84.15%	84.25%	86.50%
	refined2013	67.50%	59.34%	67.19%
	refined2016	71.91%	62.65%	72.68%
30m	SLIC	83.63%	87.87%	87.05%
	IFT-SLIC	81.39%	86.31%	85.62%
	MaskSLIC	75.65%	85.87%	80.43%

The workflows with $60m$ resolution images (2013, 2016 and the refined workflow) have GT-M accuracy better than GT-U, which is explained by their poor accuracy for Undefined segments, reaching only 27.94% for the 2016 workflow. Maybe the volunteers just looked the simple majority class into the segment.

The Undefined class presented better accuracy for the workflows from the 2017 remote sensing image ($30m$ resolution), especially to MaskSLIC workflow, where it achieved a reasonable value of 61.54%. However, compared to the other classes (Forest or Non-Forest), its accuracy is still low. The complex segments needing refinement or adjustments usually demand an undefined classification, but this situation could not be perceived and handled. Investigations to identify Undefined segments before they are sent to volunteers are ongoing, enhancing the image segmentation, which will create better tasks to be sent to volunteers.

The Forest class had the best accuracy for all workflows and GTs, being above 88% for most cases. For GT-PRODES, although MaskSLIC had the worst accuracy compared to the 2013, 2016, SLIC, and IFT-SLIC workflows, it achieved 65.95% accuracy for Non-Forest class which in this case represent pixels for recent deforestation. The areas with recent deforestation detected by previous workflows reached only 26% for the difference between the original 2016 and 2013. By using refined segments (Section 4.3) it increased to 48%. So the accuracy was significantly improved applying the last work-

flow's methodology, although the years and image resolution are different. For future workflows is advisable to use MaskSLIC instead of getting the difference between two images segmented by SLIC. However, a better segmentation segment method will continue to be sought.

5.4. Task's Difficulty Level

With the definition of entropy (HE) and its classification, given in Section 3.3.3, Table 5 shows the difficulty level frequency for each workflow. Although IFT-SLIC seems to be easier for volunteers, SLIC was the one that achieved the best results for all three GTs (see in Table 4).

The workflows with resolution images of $30m$ show a desirable decrease in the proportion of hard tasks. However, MaskSLIC didn't present the same magnitude for decreasing as the other two workflows. Also, it presented low GT accuracy when compared to the former workflows, showing how challenging can be the detection of recent deforestation, even with NDVI's image.

Another difference between MaskSLIC and the other 2017 workflows methodology refers to the segments' size. In MaskSLIC, the segments had an average size of 70 pixels, which could have affected the visualization and the pattern recognition by the volunteers. Remembering a similar issue occurred with the refined segments (Section 4.3).

Table 5. Tasks' difficulty level frequency for ForestEyes' workflows.

Difficulty Level Entropy (*HE*)	Easy (E) $HE \leq 0.33$	Medium (M) $0.33 < HE \leq 0.66$	Hard (H) $HE > 0.66$
2013	681 (66.25%)	205 (19.94%)	142 (13.81%)
2016	648 (63.40%)	233 (22.80%)	141 (13.80%)
refined2013	365 (57.30%)	161 (25.27%)	111 (17.43%)
refined2016	457 (70.52%)	99 (15.28%)	92 (14.20%)
SLIC	665 (68.35%)	267 (27.44%)	41 (4.21%)
IFT-SLIC	734 (72.82%)	241 (23.91%)	33 (3.27%)
MaskSLIC	59 (64.13%)	23 (25.00%)	10 (10.87%)

Table 6 shows the difficulty level by each class. It is possible to verify that Undefined tasks were the hardest ones, followed by the Non-Forest tasks, being Forest segments the easiest ones. For workflows with $60m$ resolution, the vast majority of Undefined segments were difficult. However, for workflows with $30m$ resolution, the great majority of Undefined tasks were easy or of medium difficulty, which can explain the improvement of Undefined and GT-U accuracies.

For MaskSLIC, most Non-Forest segments were hard or of medium difficulty, contrary to the findings with SLIC and IFT-SLIC workflows. This reinforces how challenging can be the detection of recent deforestation.

Table 6. Tasks' difficulty level frequency for each class for ForestEyes' workflows. The Forest class is defined as F, Non-Forest as NF, Undefined as U, and Small as S.

Difficulty Level	Easy (E)	Medium (M)	Hard (H)
Entropy (*HE*)	$HE \leq 0.33$	$0.33 < HE \leq 0.66$	$HE > 0.66$
2013	84.10%F/59.36%NF	9.83%F/30.43%NF/20.27%U	6.07%F/10.21%NF/79.73%U
2016	81.58%F/56.74%NF/2.74%U	10.92%F/35.12%NF/24.66%U	7.50%F/8.14%NF/72.60%U
refined2013	68.59%F/53.36%NF	19.23%F/33.21%NF/20.00%U/50.00%S	12.28%F/13.43%NF/80.00%U/50.00%S
refined2016	81.47%F/66.27%NF	10.49%F/19.53%NF	8.04%F/14.20%NF/100%U/100%S
SLIC	82.49%F/65.11%NF/26.92%U	16.07%F/31.33%NF/56.73%U	1.44%F/3.56%NF/16.35%U
IFT-SLIC	82.14%F/74.94%NF/29.06%U	15.85%F/23.25%NF/57.26%U	2.01%F/1.81%NF/13.68%U
MaskSLIC	81.13%F/43.48%NF/37.50%U	13.21%F/39.13%NF/43.75%U	5.66%F/17.39%NF/18.75%U

5.5. Tasks' Time

The Zooniverse.org platform registers logs for each answer, which define the date and time for the initial and final steps. Therefore, the time spent on each answer was evaluated and filtered to eliminate some outliers. The average times were calculated regarding all workflows' answers. Measurements were taken for different categories, including the various tasks classes (Forest as F, Non-Forest as NF, Undefined as U and Small as S), the volunteers' type (registered as R and anonymous as A), and the difficulty levels (easy as E, medium as M and hard as H). All these results are presented in Table 7.

Table 7. Average time (in seconds) regarding different categories. The classes Forest, Non-Forest, Undefined, and Small are defined as F, NF, U, and S. The registered and anonymous volunteers are defined as R and A. The difficulty levels Easy, Medium, and Hard are defined as E, M, and H.

Workflows	Average Time (s)			
	General	Classification	Volunteers	Difficulty Level
2013	5.83	5.15F/5.85NF/9.06U	5.76R/7.19A	5.06E/6.80M/8.19H
2016	6.96	6.08F/6.94NF/10.48U	6.56R/9.75A	5.80E/8.33M/10.07H
refined2013	3.84	3.57F/3.42NF/6.44U/5.79S	3.75R/6.00A	3.13E/4.45M/5.29H
refined2016	3.49	3.34F/3.07NF/6.25U/6.44S	3.41R/5.69A	3.10E/3.95M/4.95H
SLIC	3.29	2.91F/3.15NF/4.78U	3.23R/5.72A	2.86E/4.12M/4.87H
IFT-SLIC	3.54	3.34F/3.21NF/4.86U	3.51R/4.99A	3.16E/4.50M/5.11H
MaskSLIC	5.31	4.84F/5.73NF/6.07U/6.36S	5.30R/8.33A	4.30E/6.40M/8.71H

The answers that received Undefined or Small classification, anonymous volunteers, and hard tasks took longer. Considering the difficulty level is possible to see that the average time increases as the difficulty increases, once tasks with medium difficulty took longer than the easy ones, and the hardest took longer than the medium.

The MaskSLIC workflow had an average time greater than SLIC, which has the same image resolution. This could have happened due to a small segment size in MaskSLIC, joined to user interface characteristics, since it requires analysis of three images instead of two as the other workflows. Also, from Table 5 this last workflow presented a percentage of hard tasks closer to the 2013 and 2016 workflows than the other workflows with $30m$ resolution (SLIC and IFT-SLIC).

The 2016 workflow may have had a longer average time because the volunteers had to flip the images to see both color compositions while the other workflows displayed the color compositions side-by-side as default. Both 2016 and 2013 workflows also didn't have a default zoom for the tasks, different from the subsequent workflows. Therefore, the volunteers needed to zoom in to get closer to the tasks, which could dispense some time. These conditions could explain why 2016 and 2013 workflows have almost 6s/7s as average time while the new tasks from refined2016 and refined2013 spent only 4s.

5.6. Accuracy Regarding Task's Difficulty Level and HoR

Figure 9 shows the accuracy for different GT (GT-U and GT-M) by considering the tasks' difficulty level and HoR. It can be seen that for both GTs, the accuracies drop as the difficulty arises for most workflows (Figures 9 A and C). As for HoR, for GT-M (Figure 9 D), the accuracy increases as the segments are more homogeneous, except for MaskSLIC once it had a reasonable Undefined class accuracy, especially with segments with HoR between 0.6 and 0.7.

For GT-U (Figure 9 B), the volunteers had better accuracy for segments with HoR below 0.6 than segments with HoR between 0.6 and 0.7 once as less homogeneous the segment as better the volunteers classify them as Undefined. However, the accuracy is still low, reinforcing the need to early notice Undefined segments, to perform segmentation refinements, and only after that, send them to the volunteers.

This behavior does not occur for MaskSLIC, wherein GT-U accuracy for HoR between 0.6 and 0.7 was much higher than the first range (HoR below 0.6). This could have happened due to these tasks' visualization (i.e. color or distribution of the pixels and segments' shapes), which enabled the volunteers to notice the mixture between Forest and Non-Forest pixels. More investigation is planned to understand this abnormal outcome.
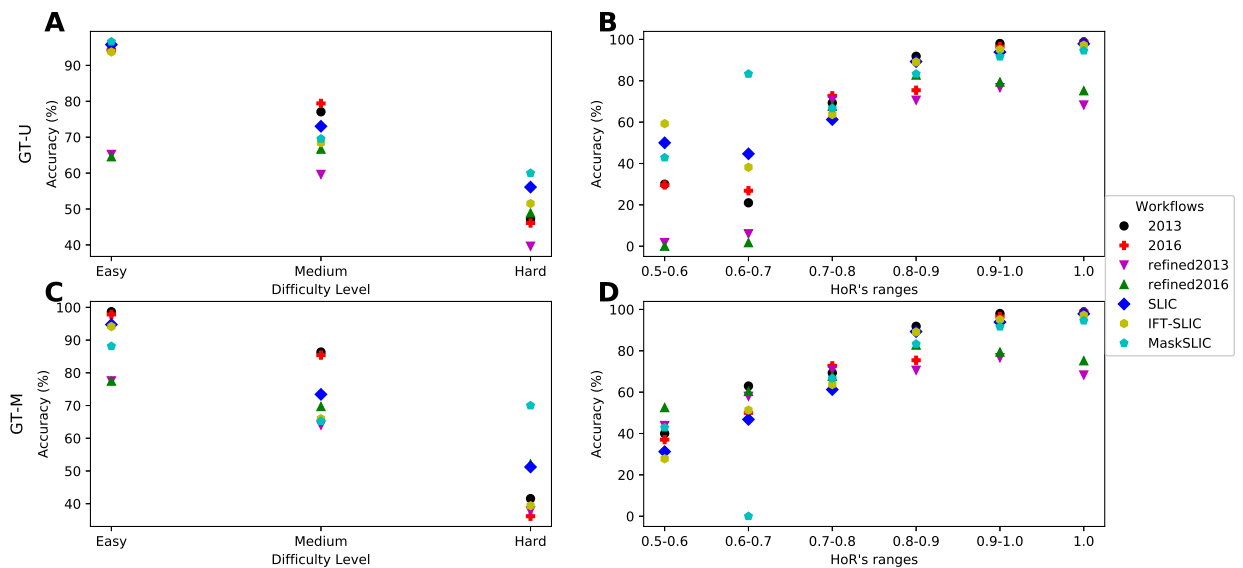

Figure 9. Accuracies for each workflow considering difficulty level and HoR.

5.7. Volunteers' Score

The volunteers' scores, according to the metrics defined in Section 3.3.5, were evaluated, generating a volunteers' ranking where the first five positions are shown in Table 8. The consensus hits at the third column show the number of contributions in agreement to the consensus. The hit rate (HR) consensus represents the same value as a relative measurement. The fifth column presents the Volunteer Score (VS) value. The last two columns show the HR related to GT-U and GT-M.

Table 8. Volunteers' ranking for ForestEyes Project.

User_id	Number answers	Consensus hits	HR consensus	VS	HR GT-U	HR GT-M
1805790	4,459	4,194	94.06%	4273.5	86.61%	86.41%
1917133	3,255	3,065	94.16%	3122	78.59%	82.33%
1756349	3,266	2,950	90.32%	3044.8	71.28%	73.48%
1733572	3,266	2,860	87.57%	2981.8	77.56%	78.44%
1061480	2,633	2,480	94.19%	2525.9	86.14%	90.32%
...	...	...	...	...	...	...

By dividing the volunteers into registered and anonymous, the average HRs for consensus, GT-U, and GT-M were calculated and are presented in Table 9. Besides the registered volunteers contributing more than the anonymous, as seen in Table 2, they also achieve better performance.

Table 9. Average HRs for registered and anonymous volunteers.

Average HR	Registered	Anonymous
Consensus	79.22%	75.62%
GT-U	69.21%	63.13%
GT-M	72.24%	69.03%

6. Conclusion

This paper described the ForestEyes Project and the steps taken to attest its feasibility regarding the volunteers' contributions which encompassed the following modules: Pre-processing, Citizen Science, and Organization and Selection. The remote sensing images used were from Landsat-8 satellites, and PRODES-based data were used as groundtruth.

The six official CS campaigns covered variations in image resolution ($60m$ and $30m$), segmentation algorithm (SLIC, IFT-SLIC, and MaskSLIC), and user interfaces. More than $86,000$ contributions were received from 644 distinct volunteers, being 174 anonymous and 470 registered. The registered volunteers contributed much more than the anonymous and achieved higher average HRs for both consensus and GTs.

The upgrade of the image resolution combined with the user interface enhancement seems to improve the volunteers' contributions, reaching higher accuracy regarding

the Undefined class and decreasing the percentage of hard tasks. However, more improvements are needed to identify Undefined segments before sending them to volunteers. The recognition of Undefined segments is a very important issue as it enables the refinement of these segments and thus improves the volunteers' accuracy regarding GT-PRODES.

The good results obtained in this paper show the power of the CS and the Zooniverse.org community. The continuous improvements in the ForestEyes Project could enable the use of CS data as a possible complement to official forest monitoring programs, observing recently deforested areas and contributing to the preservation of the world's biggest cradle of biodiversity.

7. Future Work

As future work to turn the ForestEyes Project into an operational system, some challenges need to be overcome, such as:

- Pre-Processing Module: new segmentation methods (for example, Deep Learning for semantic segmentation [L]), automatic pre-identification of Undefined segments, processing to be performed with new types of Non-Forest (as clouds and rivers).
- Citizen Science Module: the selection of segments to be sent to the volunteers, less dependence on PRODES classification, investigations into the volunteers' perceptions or cognition, and visualization enhancements.
- Machine Learning Module: the development of this module is ongoing, where the volunteers' classifications will be used with a machine learning paradigm, known as Active Learning (AL) [T], to build a robust training set. A previous work [D] showed the feasibility of using AL with PRODES data, where the training set was incremented with pixels. This new training set will be composed of segments, so a new type of representation is required, wherein Haralick texture features [H] and deep learning features can be applied.
- Next campaigns: to check the deforestation pattern in other forest areas, new geographical regions, new satellites, and to apply this project to other application domain as the use and occupancy of urban land in hillsides in the fountainhead protection area.

References

Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence*, 34(11):2274–2282, 2012.

C Adam-Bourdarios, D Cameron, A Filipčič, E Lancon, Wenjing Wu, et al. Atlas@ home: harnessing volunteer computing for hep. In *Journal of Physics: Conference Series*, volume 664, page 022009. IOP Publishing, 2015.

Aneta Afelt, Roger Frutos, and Christian Devaux. Bats, coronaviruses, and defor-estation: Toward the emergence of novel infectious diseases? *Frontiers in microbiology*, 9:702, 2018.

Eduardo Barreto Alexandre, Ananda Shankar Chowdhury, Alexandre Xavier Falcao, and Paulo A Vechiatto Miranda. Ift-slic: A general framework for superpixel generation based on simple linear iterative clustering and image foresting transform. In *2015 28th SIBGRAPI Conference on Graphics, Patterns and Images*, pages 337–344. IEEE, 2015.

Cláudio Aparecido de Almeida, Alexandre Camargo Coutinho, Júlio César Dalla Mora Esquerdo, Marcos Adami, Adriano Venturieri, Cesar Guerreiro Diniz, Nadine Dessay, Laurent Durieux, and Alessandra Rodrigues Gomes. High spatial resolution land use and land cover mapping of the brazilian legal amazon in 2008 using landsat-5/tm and modis data. *Acta Amazonica*, 46(3):291–302, 2016.

David P Anderson, Jeff Cobb, Eric Korpela, Matt Lebofsky, and Dan Werthimer. Seti@ home: an experiment in public-resource computing. *Communications of the ACM*, 45(11):56–61, 2002.

Jeferson S Arcanjo, Eduardo FP Luz, Álvaro L Fazenda, and Fernando M Ramos. Methods for evaluating volunteers' contributions in a deforestation detection citizen science project. *Future Generation Computer Systems*, 56:550–557, 2016.

Rick Bonney, Caren Cooper, and Heidi Ballard. The theory and practice of citizen science: Launching a new journal. *Citizen Science: Theory and Practice*, 1(1), 2016.

Bianca Marissa Bradford and Glenn D Israel. *Evaluating volunteer motivation for sea turtle conservation in Florida*. University of Florida Cooperative Extension Service, Institute of Food and . . . , 2004.

Jeffrey P Cohn. Citizen science: Can volunteers do real research? *BioScience*, 58 (3):192–197, 2008.

Seth Cooper, Firas Khatib, Adrien Treuille, Janos Barbero, Jeehyung Lee, Michael Beenen, Andrew Leaver-Fay, David Baker, Zoran Popović, et al. Predicting protein structures with a multiplayer online game. *Nature*, 466(7307):756, 2010.

José Rodrigues Coura and José Borges-Pereira. Chagas disease: 100 years after its discovery. A systemic review. *Acta tropica*, 115(1-2):5–13, 2010.

Justin Cranshaw and Aniket Kittur. The polymath project: lessons from a successful online collaboration in mathematics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1865–1874, 2011.

Vickie Curtis. Motivation to participate in an online citizen science game: A study of foldit. *Science Communication*, 37(6):723–746, 2015.

F.B.J.R. Dallaqua, A.L. Fazenda, and F.A. Faria. ForestEyes project: Can citizen

scientists help rainforests? In *IEEE 15th International Conference on eScience*, pages 18–27. IEEE, 9 2019.

Fernanda B. J. R. Dallaqua, Fabio Augusto Faria, and Alvaro L Fazenda. Active learning approaches for deforested area classification. In *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 48–55. IEEE, 2018.

Luiz Fernando Ferreira Gomes de Assis, Karine Reis Ferreira, Lúbia Vinhas, Luis Maurano, Cláudio Aparecido de Almeida, Jether Rodrigues Nascimento, André Fernandes Araújo de Carvalho, Claudinei Camargo, and Adeline Marinho Maciel. Terrabrasilis: A spatial data infrastructure for disseminating deforestation data from brazil.

Carlos M de Souza Jr, Sanae Hayashi, and Adalberto Veríssimo. 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.

Cesar Guerreiro Diniz, Arleson Antonio de Almeida Souza, Diogo Corrêa Santos, Mirian Correa Dias, Nelton Cavalcante da Luz, Douglas Rafael Vidal de Moraes, Janaina Sant'Ana Maia, Alessandra Rodrigues Gomes, Igor da Silva Narvaes, Dalton M Valeriano, 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*, 8(7):3619–3628, 2015.

Debra A Fischer, Megan E Schwamb, Kevin Schawinski, Chris Lintott, John Brewer, Matt Giguere, Stuart Lynn, Michael Parrish, Thibault Sartori, Robert Simpson, et al. Planet hunters: the first two planet candidates identified by the public using the kepler public archive data. *Monthly Notices of the Royal Astronomical Society*, 419 (4):2900–2911, 2012.

Steffen Fritz, Cidália Fonte, and Linda See. The role of citizen science in earth observation, 2017.

Steffen Fritz et al. Geo-Wiki: An online platform for improving global land cover. *Environmental Modelling & Software*, 31:110–123, 2012.

Alessandra R Gomes, César G Diniz, and Cláudio A Almeida. Amazon regional center (inpe/cra) actions for brazilian amazon forest: Terraclass and capacity building projects. *Interdisciplinary Analysis and Modeling of Carbon-Optimized Land Management Strategies for Southern Amazonia*, page 101, 2014.

François Grey. Viewpoint: The age of citizen cyberscience. *Cern Courier*, 29, 2009.

Christi J Guerrini, Mary A Majumder, Meaganne J Lewellyn, and Amy L McGuire. Citizen science, public policy. *Science*, 361(6398):134–136, 2018.

Product Guide. Landsat surface reflectance-derived spectral indices; 3.6 version. *Department of the Interior US Geological Survey (USGS): Reston, VA, USA*, 2017.

Trisha Gura. Citizen science: amateur experts. *Nature*, 496(7444):259–261, 2013.

Muki Haklay. Citizen science and volunteered geographic information: Overview and typology of participation. In *Crowdsourcing geographic knowledge*, pages 105–122. Springer, 2013.

Matthew C Hansen, Yosio E Shimabukuro, Peter Potapov, and Kyle Pittman. Comparing annual modis and prodes forest cover change data for advancing monitoring of brazilian forest cover. *Remote Sensing of Environment*, 112(10):3784–3793, 2008.

Matthew C Hansen, Peter V Potapov, Rebecca Moore, Matt Hancher, SAA Tu-rubanova, Alexandra Tyukavina, David Thau, SV Stehman, SJ Goetz, Thomas R Loveland, et al. High-resolution global maps of 21st-century forest cover change. *science*, 342(6160):850–853, 2013.

Matthew C Hansen, Alexander Krylov, Alexandra Tyukavina, Peter V Potapov, Svetlana Turubanova, Bryan Zutta, Suspense Ifo, Belinda Margono, Fred Stolle, and Rebecca Moore. Humid tropical forest disturbance alerts using landsat data. *Environmental Research Letters*, 11(3):034008, 2016.

Robert M Haralick, Karthikeyan Shanmugam, et al. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, (6):610–621, 1973.

Simon Haykin. *Neural networks: a comprehensive foundation*. Prentice Hall PTR, 1994.

Wesley M Hochachka, Daniel Fink, Rebecca A Hutchinson, Daniel Sheldon, Weng-Keen Wong, and Steve Kelling. Data-intensive science applied to broad-scale citizen science. *Trends in ecology & evolution*, 27(2):130–137, 2012.

Anne Holohan and Anurag Garg. Collaboration online: The example of distributed computing. *Journal of Computer-Mediated Communication*, 10(4):JCMC10415, 2005.

INPE. A taxa consolidada de desmatamento por corte raso para os nove esta-dos da Amazônia Legal (AC, AM, AP, MA, MT, PA, RO, RR e TO) em 2019 é de $10.129 km^2$. `http://www.inpe.br/noticias/noticia.php?Cod_Noticia=5465`, 2020. Accessed: 2021-02-13.

Benjamin Irving. maskslic: regional superpixel generation with application to local pathology characterisation in medical images. *arXiv preprint arXiv:1606.09518*, 2016.

Aisling Irwin. No phds needed: how citizen science is transforming re-search. *Nature*, 562(7728):480—482, October 2018. ISSN 0028-0836. doi: 10.1038/d41586-018-07106-5. URL `https://doi.org/10.1038/d41586-018-07106-5`.

Christopher O Justice, Eric Vermote, John RG Townshend, Ruth Defries, David P Roy, Dorothy K Hall, Vincent V Salomonson, Jeffrey L Privette, George Riggs, Alan Strahler, et al. The moderate resolution imaging spectroradiometer (modis): Land remote sensing for global change research. *IEEE transactions on geoscience and remote sensing*, 36(4):1228–1249, 1998.

Karen N King and Cynthia V Lynch. The motivation of volunteers in the nature conservancy-ohio chapter, a non-profit environmental organization. *Journal of Volunteer Administration*, 16(2):5–11, 1998.

Trupti M Kodinariya and Prashant R Makwana. Review on determining number of cluster in k-means clustering. *International Journal*, 1(6):90–95, 2013.

Margaret Kosmala, Andrea Wiggins, Alexandra Swanson, and Brooke Simmons. Assessing data quality in citizen science. *Frontiers in Ecology and the Environment*, 14(10):551–560, 2016.

Donald T Lauer, Stanley A Morain, and Vincent V Salomonson. The landsat pro-gram: Its origins, evolution, and impacts. *Photogrammetric Engineering and Re-mote Sensing*, 63(7):831–838, 1997.

Pietro De Lellis, Shinnosuke Nakayama, and Maurizio Porfiri. Using demographics toward efficient data classification in citizen science: a bayesian approach. *PeerJ Computer Science*, 5:e239, nov 2019. doi: 10.7717/peerj-cs.239. URL `https://doi.org/10.7717%2Fpeerj-cs.239`.

Jianhua Lin. Divergence measures based on the shannon entropy. *IEEE Transactions on Information theory*, 37(1):145–151, 1991.

Chris J Lintott, Kevin Schawinski, Anže Slosar, Kate Land, Steven Bamford, Daniel Thomas, M Jordan Raddick, Robert C Nichol, Alex Szalay, Dan Andreescu, et al. Galaxy zoo: morphologies derived from visual inspection of galaxies from the sloan digital sky survey. *Monthly Notices of the Royal Astronomical Society*, 389(3):1179–1189, 2008.

Stuart Lloyd. Least squares quantization in pcm. *IEEE transactions on information theory*, 28(2):129–137, 1982.

D Lombraña González, A Harutyunyan, B Segal, I Zacharov, E McIntosh, PL Jones, M Giovannozzi, L Rivkin, MA Marquina, P Skands, et al. Lhc@ home: a volunteer computing system for massive numerical simulations of beam dynamics and high energy physics events. In *Conf. Proc.*, volume 1205201, pages 505–507, 2012.

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*, pages 3431–3440, 2015.

Thomas E. Lovejoy and Carlos Nobre. Amazon tipping point. *Science Advances*, 4(2), 2018. doi: 10.1126/sciadv.aat2340. URL `https://advances.sciencemag.org/content/4/2/eaat2340`.

Eduardo FP Luz, Felipe RS Correa, Daniel L González, François Grey, and Fernando M Ramos. The forestwatchers: a citizen cyberscience project for deforestation monitoring in the tropics. *Human Computation*, 1(2):137–145, 2014.

K MacDicken, Ö Jonsson, L Piña, S Maulo, V Contessa, Y Adikari, M Garzuglia, E Lindquist, G Reams, and R D'Annunzio. Global forest resources assessment 2015: how are the world's forests changing? 2016.

Adeline Marinho Maciel and Lubia Vinhas. Reasoning about deforestation trajectories in Pará state, brazil. In *Anais do XIX Simpósio Brasileiro de Sensoriamento Remoto*, 2019.

Nicolas Maisonneuve, Matthias Stevens, Maria E Niessen, and Luc Steels. Noisetube: Measuring and mapping noise pollution with mobile phones. In *Information technologies in environmental engineering*, pages 215–228. Springer, 2009.

N Justin Marshall, Diana A Kleine, and Angela J Dean. Coralwatch: education, monitoring, and sustainability through citizen science. *Frontiers in Ecology and the Environment*, 10(6):332–334, 2012.

Claude Martin. *On the Edge: The State and Fate of the World's Tropical Rainforests.* Greystone Books Ltd, 2015.

Alana Kasahara Neves, Thales Sehn Korting, Leila Maria Garcia Fonseca, Gilberto Ribeiro de Queiroz, Lúbia Vinhas, Karine Reis Ferreira, and Maria Isabel Sobral Escada. Terraclass x mapbiomas: Comparative assessment of legend and mapping agreement analysis. In *GEOINFO*, pages 295–300, 2017.

Mabel Ortega Adarme, Raul Queiroz Feitosa, Patrick Nigri Happ, Claudio Aparecido De Almeida, and Alessandra Rodrigues Gomes. Evaluation of deep learning techniques for deforestation detection in the brazilian amazon and cerrado biomes

from remote sensing imagery. *Remote Sensing*, 12(6), 2020. ISSN 2072-4292. doi: 10.3390/rs12060910. URL `https://www.mdpi.com/2072-4292/12/6/910`.

Rachael Petersen, Lilian Pintea, and Liz Bourgault. Forest Watcher Brings Data Straight to Environmental Defenders. http://blog.globalforestwatch.org/people/forest-watcher-brings-data-straight-to-environmental-defenders/, 2017. Accessed: 02-15-2020.

M Jordan Raddick, Georgia Bracey, Pamela L Gay, Chris J Lintott, Phil Murray, Kevin Schawinski, Alexander S Szalay, and Jan Vandenberg. Galaxy zoo: Exploring the motivations of citizen science volunteers. *arXiv preprint arXiv:0909.2925*, 2009.

Dmitry Schepaschenko et al. Recent advances in forest observation with visual interpretation of very high-resolution imagery. *Surveys in Geophysics*, 40(4):839–862, 2019.

Jonathan Silvertown. A new dawn for citizen science. *Trends in ecology & evolution*, 24(9):467–471, 2009.

Arfon M Smith, Stuart Lynn, and Chris J Lintott. An introduction to the zooniverse. In *First AAAI conference on human computation and crowdsourcing*, 2013.

Marinalva Dias Soares, Rafael Santos, Nandamudi Vijaykumar, and Luciano Dutra. Citizen science-based labeling of imprecisely segmented images: Case study and preliminary results. In *2010 Brazilian Symposium on Collaborative Systems-Simposio Brasileiro de Sistemas Colaborativos*, pages 87–94. IEEE, 2010.

Arlesson Souza, Antônio Miguel Vieira Monteiro, Camilo Daleles Rennó, Cláudio Aparecido Almeida, Dalton de Morisson Valeriano, Fabiano Morelli, Lubia Vinhas, Luis Eduardo P. Maurano, Marcos Adami, Maria Isabel Sobral Escada, Marisa da Motta, and Silvana Amaral. Metodologia utilizada nos projetos PRODES e DETER. *São José dos Campos: INPE*, 2019.

C Souza and Tasso Azevedo. Mapbiomas general handbook. *MapBiomas: São Paulo, Brazil*, pages 1–23, 2017.

Dave Stainforth, Jamie Kettleborough, Andrew P Martin, Andrew Simpson, Richard Gillis, Ali Akkas, Richard Gault, Mat Collins, David Gavaghan, and Myles Allen. Climateprediction. net: Design principles for publicresource modeling research. In *IASTED PDCS*, pages 32–38. Citeseer, 2002.

Brian L Sullivan, Christopher L Wood, Marshall J Iliff, Rick E Bonney, Daniel Fink, and Steve Kelling. ebird: A citizen-based bird observation network in the biological sciences. *Biological Conservation*, 142(10):2282–2292, 2009.

Brian L Sullivan, Jocelyn L Aycrigg, Jessie H Barry, Rick E Bonney, Nicholas Bruns, Caren B Cooper, Theo Damoulas, André A Dhondt, Tom Dietterich, Andrew Farnsworth, et al. The ebird enterprise: an integrated approach to development and application of citizen science. *Biological Conservation*, 169:31–40, 2014.

Devis Tuia, Michele Volpi, Loris Copa, Mikhail Kanevski, and Jordi Munoz-Mari. A survey of active learning algorithms for supervised remote sensing image classification. *IEEE Journal of Selected Topics in Signal Processing*, 5(3):606–617, 2011.

Wayne S Walker, Claudia M Stickler, Josef M Kellndorfer, Katie M Kirsch, and Daniel C Nepstad. Large-area classification and mapping of forest and land cover in the brazilian amazon: A comparative analysis of alos/palsar and landsat data sources. *IEEE Journal of selected topics in applied earth observations and remote sensing*, 3(4):594–604, 2010.

Andrew J Westphal, Anna L Butterworth, Christopher J Snead, Nahide Craig, David Anderson, Steven M Jones, Donald E Brownlee, Richard Farnsworth, and Michael E Zolensky. Stardust@ home: a massively distributed public search for interstellar dust in the stardust interstellar dust collector. 2005.