Biological Conservation 292 (2024) 110490

Contents lists available at ScienceDirect

Biological Conservation

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

Optimizing community science contributions in ecology: A case study on Zooniverse's 'Chicago wildlife watch'

Kimberly Rivera [a,*], Mason Fidino [a], Elizabeth W. Lehrer [a], Holly R. Torsey [b], Sarah Allen [c], Laura Trouille [d], Seth B. Magle [a]

[a] *Department of Conservation and Science, Lincoln Park Zoo, Chicago, IL, USA*
[b] *Zooniverse, c/o Department of Conservation and Science, Lincoln Park Zoo, Chicago, IL, USA*
[c] *3dna Corp. dba NationBuilder, Los Angeles, CA, USA*
[d] *Science Engagement Division and Zooniverse, The Adler Planetarium, Chicago, IL, USA*

ARTICLE INFO

Keywords:
Camera trap
Community science
Accuracy
Image classification

ABSTRACT

Public participation in research, or community science (CS), has an important role in advancing ecological research, especially data processing. CS contributions to camera trap studies have supported wildlife conservation through the rapid processing of images and videos. However, more studies are needed to quantify the accuracy and efficiency of CS participation. We used a case study from Chicago Wildlife Watch, a Zooniverse project, to explore variability in image classification accuracy and assess efficiency of responsive retirement rules which dictate how many times an image is viewed and annotated. We found that CS participants were highly accurate when classifying empty (96.0 %) and commonly photographed species in our study area (60.14 % across all species). User agreement on species images most impacted classification accuracy, though accuracy was higher for those containing larger species and those annotated by more engaged participants. With respect to efficiency, we found that three consecutive 'empty' classifications from participants led to over 95 % classification accuracy in empty images and if 7 participants agreed on a species present in an image, they were accurate 98 % of the time, on average. These results further support the value of CS in ecological research and the value of applying unique project designs which consider occurrence of regional species and field systems (e.g. camera placement or ecosystem). Given these results, we encourage scientists to continue applying quantitative techniques to custom design projects to effectively use CS participants' time and maximize data accuracy.

1. Introduction

Participatory community science (hereafter CS) has revolutionized ecological research and environmental education (Fraisl et al., 2022). Together, participants and scientists work to expand the collection of ecological data, data processing, and scientific engagement (Bonney et al., 2009; Frigerio et al., 2018). CS takes many forms, from active data collection to online data processing, (Eitzel et al., 2017) and has local to global impacts on ecological knowledge, conservation initiatives and community involvement in ecology (Louvrier et al., 2022; Sullivan et al., 2009; Bonney et al., 2009). For example, iNaturalist is an online CS platform where participants report and identify observations of wildlife in their communities. This global dataset, unachievable without large-scale participation, supports scientists to track species phenology (Di Cecco et al., 2021), the occurrence of rare wildlife (Wilson et al., 2020), and species interactions (Gazdic and Groom, 2019). Zooniverse is another CS platform that has made public participation in ecological research (among other fields) widely accessible by offering an online interface where researchers create projects where public members can contribute to data processing by annotating and classifying images or videos (Simpson et al., 2014). The proliferation of projects like these has led to many exciting opportunities to collect and curate ecological data and information on motivations of CS volunteers.

Though CS projects and platforms like these have undoubtedly become important to scientific research (Bonney et al., 2009), we require processes to validate and integrate these datasets into scientific workflows (Wiggins et al., 2011). An increasingly popular method for people to participate in CS ecology, which has lacked such validating processes, is the annotation of camera trap data (Green et al., 2020). Camera traps are a widely used, affordable tool that allows ecologists to

* Corresponding author at: 2001 N. Clark St., Chicago, IL 60614, USA.
E-mail address: kimberly.rivera22194@gmail.com (K. Rivera).

https://doi.org/10.1016/j.biocon.2024.110490
Received 14 September 2023; Received in revised form 29 January 2024; Accepted 5 February 2024
Available online 5 March 2024

study spatiotemporal, behavioral, and ecological patterns in wildlife to support conservation and management (Rovero et al., 2013). CS participation in camera data annotation is especially valuable to scientists as classifying these data is a time-consuming endeavor that typically requires professionals to visually examine thousands of individual images or videos (Swanson et al., 2015). Through this work, participants have supported scientists' efforts to understand animal distributions, movement, and behaviors (Lasky et al., 2021; Swanson et al., 2015; McCarthy et al., 2021). Participants can also benefit by gaining ecological knowledge and becoming more engaged with the scientific community via two-way communication and collaboration opportunities (Lasky et al., 2021; Cox et al., 2015).

Various platforms, including Zooniverse, have made CS participation in camera trap research especially successful (McShea et al., 2016). On Zooniverse, participants help to annotate species, individuals, or individual characteristics such as injuries or age class (McCarthy et al., 2021; Jones et al., 2018; Thel et al., 2021). Though growing literature in this field has helped to remove bottlenecks in camera trap data processing (Egna et al., 2020; Gadsden et al., 2021), evaluating the accuracy of participants and classification workflows remains an important component to this research. Validating CS data processing ensures efficient use of participants' time while maximizing the accuracy of an image's classification. Conventionally, camera trap projects hosted on Zooniverse set their images to retire between 15 and 25 aggregated classifications and may adopt additional 'responsive retirement rules' which allow images to retire more quickly if they reach certain consensus thresholds (personal communication; Gadsden et al., 2021). The use of custom retirement rules, which dictate how many times an image is classified, can help to increase classification efficiency, especially as the difficulty of species and empty image identification can greatly vary (Potter et al., 2019; Gadsden et al., 2021). Additionally, factors such as animal body size, distinguishing traits, or image quality can impact overall classification accuracy (Potter et al., 2019; Gadsden et al., 2021; Swanson et al., 2016; Murray et al., 2021). As such, there may be times when requiring a specific aggregated classification count is too much or too little. This may depend on the species in an image or the overall quality of the image itself.

To further explore how variation in camera images impact classification accuracy and the utility of responsive retirement rules, we conducted a study within Chicago Wildlife Watch (hereafter CWW), a Zooniverse project. CWW is a camera trapping project led by the Lincoln Park Zoo to monitor biodiversity in the Chicago Metropolitan area, USA. We used these data to quantify how species, image, and volunteer traits were associated with classification accuracy. We hypothesized that 1) if the average size of species makes them easier to detect in an image, then participant accuracy will increase as the average body size of species increases, 2) if poor image quality can make it more difficult to either locate or identify a species, then participant accuracy will be lower in images that are blurred relative to clear images, and 3) if participants improve their classification accuracy through experiences (e.g. time), then participants with the most classifications will be the most accurate. In addition to testing these hypotheses, we also evaluated the efficiency of CWW's retirement rules which follow commonly retirement patterns of other camera trapping Zooniverse projects (Willi et al., 2018; personal communication).

Our work aims to evaluate the biases in CS-based camera trap image classifications and outline additional analytical frameworks for researchers interested in incorporating CS into camera trapping projects reliably.

2. Materials and methods

2.1. Chicago wildlife watch

CWW (www.chicagowildlifewatch.org) was developed by the Urban Wildlife Institute (UWI) at the Lincoln Park Zoo, in collaboration with Zooniverse staff via the Project Builder Platform in 2014. Data on CWW are collected and managed by UWI as part of their long-term urban biodiversity monitoring project (see Magle et al., 2016 for sampling methodologies). Like other Zooniverse projects, CWW allows participants to identify wildlife species in photos. In addition to annotating camera trap images, CWW participants can learn about the project team, program objectives, and direct questions to the research team and one another in an open forum. While identifying images, participants can tag a single or multiple species (if they are present), indicate if young are present, or, if an image has no species, mark it as 'empty'. Participants are not able to submit 'unknown' classifications and instead are encouraged to make their best guess about what the species may be. The species available for participants to annotate were chosen based on prior detections in the study area (2010–2014): human, beaver (*Castor canadensis)*, skunk (*Mephitis mephitis*), bird (any species), livestock (any species), flying squirrel (*Glaucomys volans*), domestic cat (*Felis catus*), mink (*Neovison vison*), fox squirrel (*Sciurus niger*), chipmunk (*Tamias striatus*), mouse (family *Muridae*), gray squirrel (*Sciurus carolinensis*), coyote (*Canis latrans*), mower (human or mower visible), melanistic gray squirrel, deer (*Odocoileus virginianus*), muskrat (*Ondatra zibethicus)*, tree squirrel (if squirrels cannot be identified to species), domestic dog (*Canis lupus familiaris*), opossum (*Didelphis virginiana*), weasel (if weasels cannot be identified to species), gray fox (*Urocyon cinereoargenteus)*, rabbit (*Sylvilagus floridanus)*, woodchuck (*Marmota monax)*, red fox (*Vulpes vulpes*), raccoon (*Procyon lotor*), horse (*Equus caballus*), and rat (*Rattus norvegicus*).

2.2. Image accuracy analysis

To evaluate Zooniverse participant accuracy, classifications were validated against trained experts. Experts used a custom database to upload and classify camera trap images. Under this process, two expert reviewers examined each uploaded image. If both experts agreed on an image's classification, the image would retire as that classification. If there was disagreement between the two experts, the image moved to a validation step where a third expert would review the image and determine the final classification. We limited our analysis to images that were either empty or had a single species (98.7 % of images). Likewise, we limited Zooniverse classifications to only those associated with a Zooniverse account (87.3 % of participants), which allowed us to link classifications to a unique individual. Data from this study were classified by participants between June 2020 and May 2021.

To address our hypotheses on how animal size, image quality, and participant experience impacted classification accuracy, we conducted two analyses using binomial generalized linear mixed models. First, we fitted a model to only images that trained experts classified as empty. This was done because empty images lack species trait data, e.g. animal weight. Second, we fitted a model to images trained experts had classified as not empty (i.e., wildlife was present). For both steps, the binary response value was 1 if a Zooniverse participant's classification matched that of the experts, and 0 if it did not. Likewise, we added a participant-level and site-level random effect to both models to account for variation among participants and across sampling sites. We also account for variation across species with a random effect in the species-level model.

With respect to covariates included in the models, both models included a metric for image blurriness, participant engagement, and evenness of classification accuracy. Image blurriness was quantified using Python ver 3.10 with the OpenCV package ver 4.3.0. Images were converted into grayscale and a Laplacian kernel was applied to calculate the sharpness of image lines where lower Laplacian values indicate increased blurriness (Woods, 2012; Murray et al., 2021). As participants who engage more often on CWW may have greater classification accuracy, we created a continuous count variable to summarize participant engagement. To do this, we grouped data by participant and sorted them sequentially from the first image they classified to their last. We then numbered these images to represent the cumulative number of images each participant saw while tagging images on CWW. We assumed that participants with higher counts were more engaged on CWW than participants with low counts. We also used Pielou's evenness index to measure agreement on a single image's classification (Pielou, 1966). For this metric, a value of 1 indicates high evenness, or low classification agreement, while a 0 index indicates perfect agreement. In addition to these covariates, the species model also included the average log body weight of the species present in the image.

2.3. Retirement accuracy analysis

We developed CWW with custom retirement rules to determine when an image will be removed from the viewing or classification pool. We aimed to evaluate the effectiveness of retirement rules with consideration of their conditional pattern (see below). To evaluate the accuracy of retirement rules, we use a binomial generalized linear model with a logit link. We used the same binary response as above, where the response value was a 1 if the retired classification matched that of the experts, and a 0 if it did not. We used the categorical retirement rules as predictor variables to estimate the accuracy of each rule and added a site-level random effect to account for variation across sampling sites. The retirement classifications were conditionally implemented, occurring sequentially. The retirement conditions are as follows:

1. If any participant classified a subject as 'human', the image retired as human, hereafter 'human.'
2. If the first 3 participants classified the image as empty (consecutively), the image retired as empty, hereafter '3 empty.'
3. If 5 participants classified the image as empty (non-consecutively), the image retired as 'empty', hereafter '5 empty.'
4. If 7 participants classified the same species (non-consecutively), the image retired when the seventh participant classified the consensus species, hereafter '7 species.'
5. If images were not retired under the above conditions, images were classified by 15 participants and were retired as the classification with majority consensus, hereafter 'classification count.'

Given this retirement pattern, the image identification difficulty may vary between different retirement steps and thus are non-independent. For example, images that contain non-human species and do not retire as a species under the '7 species' rule, are likely more difficult to identify species or images.

All models were fitted with the lme4 package in R ver. 4.1.1 (Bates et al., 2014). To assess model fit, we calculated MacFadden's pseudo-R^2. Values >0.2 indicate a strong model fit (Domencich and McFadden, 1975; Louviere et al., 2000).

3. Results

We analyzed 142,062 unique images which were annotated by 2755 Zooniverse participants. The three most common image classifications by Zooniverse participants were: empty (n = 278,030), eastern gray squirrel (n = 76,774), and raccoon (n = 43,944), which matched the top three classifications among experts (n = 240,646, n = 100,180, n = 52,168; respectively). Note that unique images are classified by multiple participants on Zooniverse, thus the number of classifications is greater than the number of images (see retirement rules above). The three least commonly classified species by Zooniverse participants were: horse (n = 68), livestock (n = 161), and muskrat (n = 368). This differed from experts who least often classified: gray fox (n = 11), North American beaver (n = 14) and woodchuck (n = 46). CS participants' classification accuracy varied largely across species with flying squirrel, weasel (genus *Mustela*), and mouse as the most misidentified species (13.43 %, 22.73 %, and 26.33 % correct respectively; Appendix S). Although common urban mammals like white-tailed deer or raccoon often had the highest classification accuracy, less common species with distinguishing features, like the striped skunk, were also classified by Zooniverse participants with high accuracy (79.10 %).

3.1. Image accuracy analysis

We found the empty image model to strongly fit the data with a MacFadden's pseudo-R^2 of 0.53. We note that this analysis had a large sample size, thus the errors associated with our fixed effects are likely over-precise. Overall, Zooniverse participants were 99.62 % accurate at classifying empty images. Based on our binomial generalized linear mixed model, we found support that evenness was negatively associated with participant accuracy of empty photos (β = −1.67, 95 % CI = −1.70, −1.64; Fig. 1). For example, an empty image with perfect agreement among participants (i.e., evenness = 0) had a mean accuracy of 1.00, (95 % CI = 1.00, 0.99), which was roughly 2.83 times higher than an empty image with total disagreement among participants (i.e., evenness = 1; 0.36, 95 % CI = 0.80, 0.07). We found that participant engagement had a significant, but relatively small negative effect on classification accuracy (β = −0.17, 95 % CI = −0.23, −0.11) while image blurriness had a significant, but small positive effect on accuracy (β = 0.08, 95 % CI = 0.05, 0.12). Through the application of random effects, we found that the classification accuracy of empty images varied more by participants (sd = 1.42) than by differences in camera locations (sd = 0.28).

For our species image analysis, we calculated a MacFadden's pseudo-R^2 of 0.24, indicating a good model fit. We found that evenness, participant engagement, and animal weight were strong predictors of participant accuracy. Image blurriness had a positive but small effect on accuracy (β =0 0.06, 95 % CI = 0.04, 0.07). Evenness negatively affected the accuracy of species classifications (β = −1.10, 95 % CI = −1.12, −1.09) whereas increasing animal weight generally increased classification accuracy (β =0 0.51, 95 % CI = 0.24, 0.77; Fig. 2). As predicted, classification accuracy increased with participant engagement (β = 0.34, 95 % CI = 0.32, 0.36). The random effect structure of our species-level model revealed that classification accuracy varied most among participants (sd = 0.86, assuming normal variation in accuracy on the logit scale) and less so across species (sd = 0.60) and camera locations (sd =

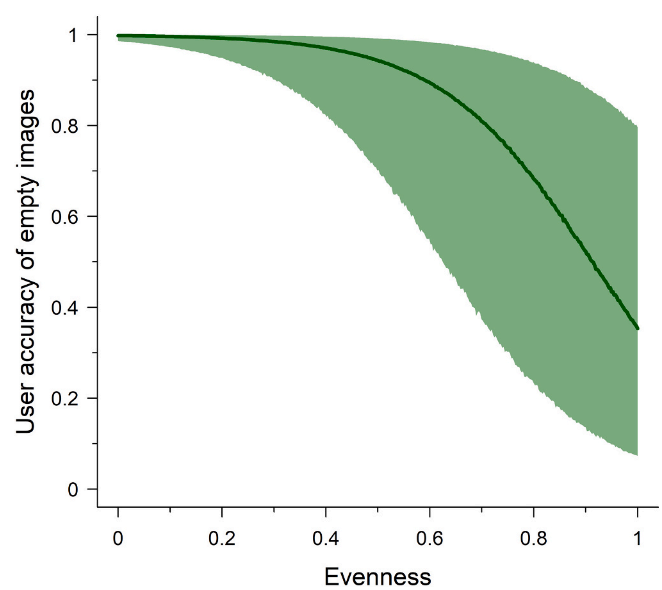

Fig. 1. Zooniverse participants more accurately classified empty images correctly when the aggregated image classifications among participants had more agreement (i.e., evenness = 0) whereas images with more disagreement among participants (i.e. evenness =1) were about 2.8 times less accurate. The solid horizontal line of this figure represents the median estimate from the model while the shaded ribbon is the 95 % confidence interval. Other model covariates were held at their mean value.

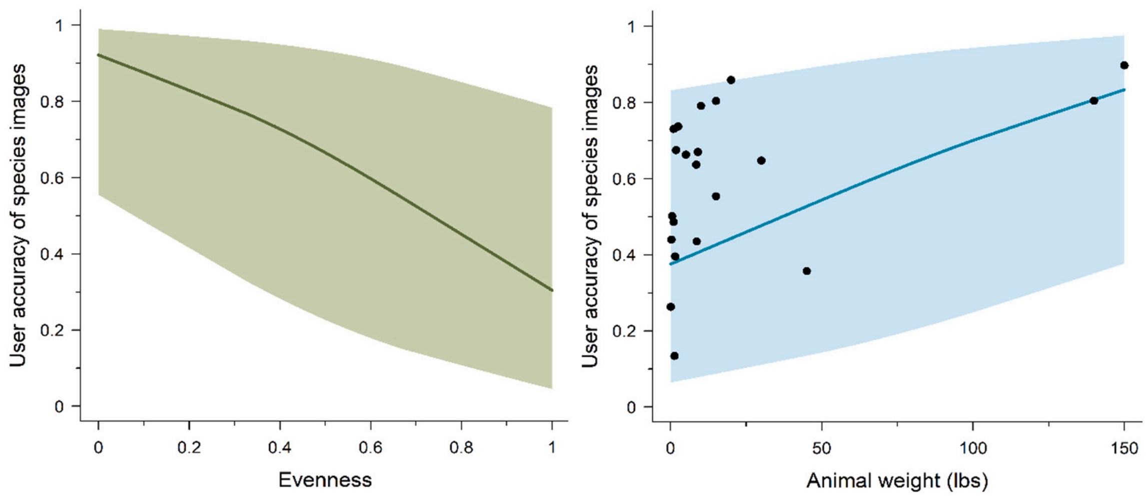

Fig. 2. Zooniverse participants were less accurate with classifying photos with species if the aggregated participant classifications had high evenness (i.e., more disagreement). Furthermore, Zooniverse participants more accurately classified species that are larger. The solid horizontal lines of each subplot represent median estimates from the model while the shaded ribbons are 95 % confidence intervals. Points indicate individual species weights. Other model covariates were held at their mean value.

0.31).

3.2. Retirement accuracy analysis

The most common retirement rule implemented was '3 empty' (the first 3 participants classified the image as empty), followed by '7 species' (7 participants classified the same animal), 'human' (any participant classified a subject as 'human'), '5 empty' (5 participants classified the image as empty), and 'classification count' (15 participants classified an image; Table 1). If we had implemented the '3 empty' rule alone, which was 95.96 % accurate, Zooniverse participants would have saved over 1,000,000 views of camera trap images then if these images has retired solely under the 'classification count' rule (e.g. sending these images to 15 participants). We calculated a MacFadden's pseudo-R^2 of 0.23 for the retirement model, indicating a good fit. We found that the three most commonly implemented retirement rules, '3 empty', '7 species', and 'human', had over 90 % classification accuracy by participants (Table 1). Species which retired under the '7 species' rule were viewed by <9 people on average (mean = 8.8 classifications). Accuracy was lowest for the '5 empty' retirement rule (57.23 %), indicating that if an image does not retire from the '3 empty' rule, it likely contains an animal species. Therefore, the '5 empty' rule may be less useful, especially as it is the second least frequently used retirement rule.

Accuracy varied among species retired with the '7 species' and 'classification count' rules (Fig. 3; Appendix S). For the '7 species' rule, species with the lowest accuracy (<65 %) were predominantly rare species, or species with fewer than 200 total image detections. This included North American beaver, gray fox, red fox and American mink. However, one rare species, woodchuck, and one rare family, weasel (Mustelids), had perfect accuracy ($p = 1.0$) within the '7 species' rule. Other species identified at high accuracy within this rule included striped skunk, raccoon, white-tailed deer, domestic dog, Virginia opossum, and Eastern cottontail (accuracy >0.99). For the 'classification count' rule, accuracy of species classifications by participants was highly variable, indicating this retirement rule was less reliable than the others. Four species were classified with perfect accuracy within this rule: American mink, domestic dog, Eastern chipmunk and North American beaver. There was no clear pattern in how rarity of species may impact the accuracy of retirement within this rule.

Table 1
Table summarizing retirement rule frequency (the count of images retired out of 141,994 images), accuracy (estimate of participant accuracy), and confidence intervals around that accuracy estimate.

Retirement rules	Rule frequency	Accuracy	95 % CI lower	95 % CI upper
3 empty	81,602	95.96	95.32	96.52
5 empty	8104	57.23	53.33	61.03
7 species	35,486	98.23	97.92	98.50
Classification count	725	83.82	80.46	86.70
Human	16,077	90.42	88.93	91.74

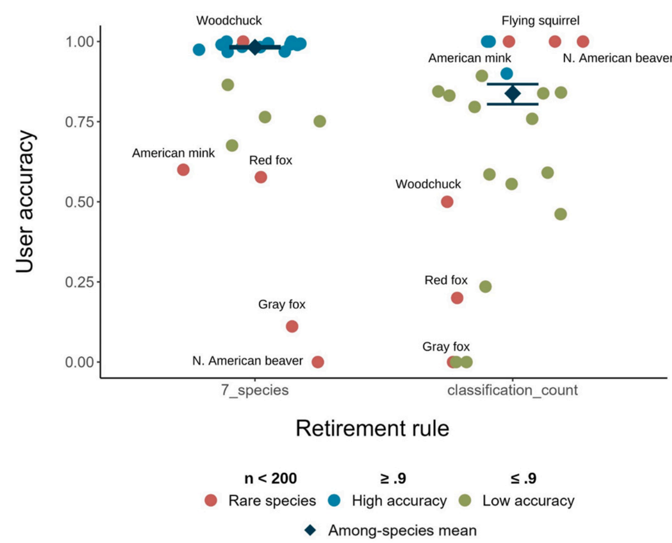

Fig. 3. The variation of participant accuracy, or the probability of Zooniverse participants correctly classifying images, across species for the retirement rules '7 species' and 'classification count'. Colors of points indicate whether the species had low accuracy (p < 0.9), high accuracy ($p \geq 0.9$), or were rare species (n < 200 images). Rare species are labeled. An among-species mean is also included with horizontal lines above and below to indicate the 95 % confidence interval.

4. Discussion

Engaging public participants in CS camera trap projects positively contributes to ecological research, specifically through data processing (McCarthy et al., 2021; Swanson et al., 2016; Gadsden et al., 2021). Participants on CWW were accurate image classifiers. As few as three annotations from Zooniverse participants, for example, could produce highly accurate classifications for empty images. As empty images represent the majority of images collected during camera trapping surveys, this represents a phenomenal increase in the speed CS participants could collectively process images with little decrease in accuracy. Additionally, species which retired under the '7 species' rule were also highly accurate for commonly seen (e.g. white-tailed deer) and distinguishable species (e.g. striped skunk). These results highlight Zooniverse participant's ability to annotate images (for at least our study species) efficiently. Thus, customizable retirement rules are central to improving CS processing time of empty images, as well as common and distinguishable species.

In testing our hypotheses, we found that image evenness, or measure of agreement, had an inverse relationship with classification accuracy (i.e., as agreement decreased, accuracy decreased) for both empty and species-present images. These results align with previous studies (Swanson et al., 2016; Gadsden et al., 2021) and we encourage future projects to use evenness as a threshold to accept aggregated classifications by participants. Exact thresholds should be determined by project specific goals and rarity of species of interest (Swanson et al., 2016). Though we confirmed that increasing animal weight improved overall accuracy (Potter et al., 2019), we believe evenness is a more reliable measure given that certain species characteristics, such as rarity, highly impact classification accuracy. For example, American beavers were one of the largest species captured in CWW images but had <0.4 accuracy in our species-level model likely due to their rarity ($n = 5$ unique images). The importance of rarity, as highlighted in other studies (Clare et al., 2019), was further emphasized in our retirement accuracy analysis, specifically with the '7 species' retirement rule. We found that species with the lowest accuracy (bottom four) were all rare species, except for woodchuck and weasel (not classified to species), which had perfect detection accuracy. These results indicate that rare species generally require trained expert review. We therefore encourage future studies to generate regionally specific lists of common and rare species and use these as guides to either accept or review CS classified images.

Given that our case study is limited to one ecological system, other CS camera trap projects should conduct similar statistical analyses that are curated towards their study species and sampling environments. Our analysis suggests the utility of two unique retirement rules, one for empty images and another for species-present images. Given the high classification accuracy for empty images, we believe the '3 empty' retirement rule, where three people consecutively annotate and retire an image as 'empty', may be effective for most studies. Researchers who collect images with distant field-of-views (cameras placed high to capture distant background areas), may need to consider a higher threshold (such as 5 consecutive empties) to account for increased likelihood of false 'empty' classifications (Egna et al., 2020). We found that if there was disagreement within the first three images, there was likely an animal present, as indicated by the low accuracy found in the '5 empty' rule. Low accuracy in the '5 empty' rule may also indicate these images are more difficult to identify as empty, or that there are inaccuracies within expert classifications. However, participants failing to detect an animal are more common than experts falsely detecting an animal in a truly empty image (personal observation). For species-present images, we found that an average of 9 classifiers was sufficient to accurately classify the majority of species, similar to results found by Swanson et al. (2016). In our study, the '7 species' retirement rule was highly effective, leaving few images to retire under the last, 'catch-all' rule, 'classification count' (725 of 140 thousand images). Though these retirement rules cannot be compared independently given the conditional nature of their implementation, our results indicate that most images are retired before images reach 15 classifiers. We note that images which retired under the 'classification count' rule are likely to be difficult to annotate images and thus require expert review. As an alternative to increasing an image's viewing pool (or number of classifiers), we recommend calculating evenness across aggregated images and using this value as a threshold to reject or accept CS participant classifications for a given image. To determine this threshold and the accuracy of the retirement rules we used, we recommend all projects to maintain a pool of expertly classified data (classified by >2 experts) for comparative analyses (Swanson et al., 2016).

In summary, we recommend the following steps. First project managers should maintain an expert pool of annotated data (where at least two experts annotate a single image). We recognize that expertly classified data is also subject to misclassification, and suggest that depending on downstream objectives and analyses, it may be useful to account for this through model-based solutions (Clare et al., 2021). The proportion of expertly annotated images has varied study to study, where Swanson et al., analyzed 0.35 % of their 1.2 million images and Gadsden et al., 56.22 % of 10,199 images. Here, we recommend studies with <10,000 images to annotate 50 % of images and studies >10,000 images to maintain a minimum of 5000 images with the addition of any images annotated with rare species. A rare species list should be generated by project managers based on their knowledge of the area and history of annotated species. Projects may begin by implementing one to two types of retirement rules, one for empty images (3–5 consecutive 'empty' tags), and another for species. The agreement of 7 species classifications was sufficient for our ecosystem, however new projects may find it useful to implement only an empty image retirement rule while retiring all other images at 10 classifications to determine a species-specific retirement rule most useful for their ecosystem and project aims (such as 7 participants classifying the same species). Additionally, any images which retire as a rare species or fall above a relevant evenness threshold, should be reviewed by experts. The expert pool of data can be used to determine an appropriate evenness threshold, though <0.5 has been recommended by others (Swanson et al., 2016).

While we did not assess this with our study, it is necessary to mention that Machine Learning (ML) is an additional tool that can be considered in CS camera trapping workflows. ML in camera trapping commonly uses deep learning algorithms trained on large datasets of images to rapidly classify camera trap data (Norouzzadeh et al., 2020). Various organizations, including Zooniverse, are harnessing ML to support efficient classifications of project image data (Vélez et al., 2022a). However, the success of ML is limited by the diversity represented in trained image datasets, both in terms of species present and geographic coverage. For example, geographical variation between trained and applied datasets can impact classification accuracy which may be driven by differences in backgrounds or dominant features (habitat type, presence of rocks or trees, etc.) or environmental conditions (light or climate; Schneider et al., 2020). Therefore, CS participation on untrained datasets is especially useful and can be applied in tandem with ML. This dual application can speed up image processing and allow participants to grow trained datasets to verify and improve ML classification accuracy for their specific project (Vélez et al., 2022b). For example, to further increase the annotation of images in a CS project, a researcher may want to use ML to filter out empty images which were identified with high confidence, send the remaining images to CS participants for further validation, and then manually validate images with low participant agreement or of rare species.

Overall, our results highlight the value of public participation in camera data processing. We believe that successful CS projects value participants' contributions, and project leaders can support their participants by developing streamlined workflows and improving two-way communication. We encourage project leaders to provide training tools and regular feedback as to how their work is contributing to ongoing research to maximize project success and engagement. Specifically, communication with highly engaged participants could significantly contribute to participant accuracy as found in this study. Examples could include development of thorough tutorials (Cox et al., 2015), regular communication with participants via blog posts, forums, or access to scientific communications (advertising presentations, workshops, etc.) that address project goals, updates, successes, and impacts on ecology and conservation. Together, community participants and ecologists have power neither group has separately, and the more efficient these collaborations, the more we can accelerate scientific discovery and innovation.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Kimberly Rivera: Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Mason Fidino:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Elizabeth W. Lehrer:** Writing – review & editing, Project administration, Investigation, Conceptualization. **Holly R. Torsey:** Writing – review & editing, Project administration. **Sarah Allen:** Writing – review & editing, Project administration. **Laura Trouille:** Writing – review & editing, Project administration. **Seth B. Magle:** Writing – review & editing, Project administration, Investigation, Conceptualization.

Declaration of competing interest

These authors have no competing interests to declare.

Data availability

Custom code and data for this analysis is available on GitHub: https://github.com/karivera2194/Zooniverse_manuscript.git.

Acknowledgements

We would like to recognize the fantastic Zooniverse team who supported the building of Chicago Wildlife Watch in addition to feedback on manuscript drafts. This research would not be possible without the help of thousands of Zooniverse participants who spent their time and energy helping to classify images and moderate questions. We are incredibly thankful to these participants and will continue to support and recognize this community. We also thank members of Urban Wildlife Institute and the Lincoln Park Zoo for their help and support. Funding for the Urban Wildlife Institute was provided by the Grainger Foundation, the Abra Prentice Foundation, and the Davee Foundation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.biocon.2024.110490.

References

Bates, M. Douglas, Maechler, Matin, Bolker, Ben, Walker, Steve, 2014. lme4: linear mixed-effects models using Eigen and S4. R package version 1.1-7. http://CRAN.R-project.org/package=lme4.

Bonney, Rick, Cooper, Caren B., Dickinson, Janis, Kelling, Steve, Phillips, Tina, Rosenberg, Kenneth V., Shirk, Jennifer, 2009. Citizen science: a developing tool for expanding science knowledge and scientific literacy. BioScience 59 (11), 977–984. https://doi.org/10.1525/bio.2009.59.11.9.

Cecco, Di, Grace, J., Barve, Vijay, Belitz, Michael W., Stucky, Brian J., Guralnick, Robert P., Hurlbert, Allen H., 2021. Observing the observers: how participants contribute data to INaturalist and implications for biodiversity science. BioScience 71 (11), 1179–1188. https://doi.org/10.1093/biosci/biab093.

Clare, D.J., John, Philip A., Townsend, Christine Anhalt-Depies, Locke, Christina, Stenglein, Jennifer L., Frett, Susan, Martin, Karl J., Singh, Aditya, Van Deelen, Timothy R., R, and Benjamin Zuckerberg., 2019. Making inference with messy (citizen science) data: when are data accurate enough and how can they be improved? Ecol. Appl. 29 (2), e01849.

Clare, J.D., John, Philip A., Townsend, and Benjamin Zuckerberg., 2021. Generalized model-based solutions to false-positive error in species detection/nondetection data. Ecology 102 (2), e03241.

Cox, Joe, Eun Young, Oh., Simmons, Brooke, Lintott, Chris, Masters, Karen, Graham, Gary, Greenhill, Anita, Holmes, Kate, 2015. Defining and measuring success in online citizen science: a case study of Zooniverse projects. Computing in Science & Engineering 17 (4). https://doi.org/10.1109/MCSE.2015.65.

Domencich, Tom, McFadden, Daniel L., 1975. Urban Travel Demands: A Behavioral Analysis. North Holland Publishing, Amsterdam.

Egna, Nicole, O'Connor, David, Stacy-Dawes, Jenna, Tobler, Mathias W., Pilfold, Nicholas, Neilson, Kristin, Simmons, Brooke, Davis, Elizabeth O., Bowler, Mark, Fennessy, Julian, Glikman, Jenny A., Larpei, Lexson, Lekalgitele, Jesus, Lekupanai, Ruth, Lekushan, Johnson, Lemingani, Lekuran, Lemirgishan, Joseph, Lenaipa, Daniel, Lenyakopiro, Jonathan, Lenalakiti, Ranis, Lororua, Masenge, Muneza, Arthur, Rabhayo, Sebastian, Ranah, Symon Masiaine Ole, Ruppert, Kirstie, Owen, Megan, 2020. Camera settings and biome influence the accuracy of citizen science approaches to camera trap image classification. Ecol. Evol. 10, 11954–11965. https://doi.org/10.1002/ece3.6722.

Eitzel, M.V., Cappadonna, Jessica L., Santos-Lang, Chris, Duerr, Ruth Ellen, Virapongse, Arika, West, Sarah Elizabeth, Kyba, Christopher Conrad Maximillian, Bowser, Anne, Cooper, Caren Beth, Sforzi, Andrea, Metcalfe, Anya Nova, Harris, Edward S., Thiel, Martin, Haklay, Mordechai, Ponciano, Lesandro, Roche, Joseph, Ceccaroni, Luigi, Shilling, Fraser Mark, Dörler, Daniel, Heigl, Florian, Kiessling, Tim, Davis, Brittany Y., Jiang, Qijun, 2017. Citizen science terminology matters: exploring key terms. Citizen Science: Theory and Practice 2 (1), 1–20. https://doi.org/10.5334/cstp.96.

Fraisl, Dilek, Hager, Gerid, Bedessem, Baptiste, Gold, Margaret, Hsing, Pen-Yuan, Danielsen, Finn, Hitchcock, Colleen B., Hulbert, Joseph M., Piera, Jaume, Spiers, Helen, Thiel, Martin, Haklay, Mordechai, 2022. Citizen science in environmental and ecological sciences. Nature Reviews Methods Primers 2 (1), 1–20. https://doi.org/10.1038/s43586-022-00144-4.

Frigerio, Didone, Pipek, Pavel, Kimmig, Sophia, Winter, Silvia, Melzheimer, Jörg, Diblíková, Lucie, Wachter, Bettina, Richter, Anett, 2018. Citizen science and wildlife biology: synergies and challenges. Ethology 124 (6), 365–377. https://doi.org/10.1111/eth.12746.

Gadsden, Gabriel I., Malhotra, Rumaan, Schell, Justin, Carey, Tiffany, Harris, Nyeema C., 2021. Michigan ZoomIN: validating crowd-sourcing to identify mammals from camera surveys. Wildl. Soc. Bull. 45 (2), 221–229. https://doi.org/10.1002/wsb.1175.

Gazdic, Milan, Groom, Quentin, 2019. INaturalist is an unexploited source of plant-insect interaction data. Biodiversity Information Science and Standards 3, e37303. https://doi.org/10.3897/biss.3.37303.

Green, Siân E., Rees, Jonathan P., Stephens, Philip A., Hill, Russell A., Giordano, Anthony J., 2020. Innovations in camera trapping technology and approaches: the integration of citizen science and artificial intelligence. Animals 10 (1), 132. https://doi.org/10.3390/ani10010132.

Jones, Fiona M., Allen, Campbell, Arteta, Carlos, Arthur, Joan, Black, Caitlin, Emmerson, Louise M., Freeman, Robin, Hines, Greg, Lintott, Chris J., Macháčková, Zuzana, Miller, Grant, Simpson, Rob, Southwell, Colin, Torsey, Holly R., Zisserman, Andrew, Hart, Tom, 2018. Time-lapse imagery and volunteer classifications from the Zooniverse penguin watch project. Scientific Data 5 (1), 180124. https://doi.org/10.1038/sdata.2018.124.

Lasky, Monica, Parsons, Arielle, Schuttler, Stephanie, Mash, Alexandra, Larson, Lincoln, Norton, Ben, Pease, Brent, Boone, Hailey, Gatens, Lisa, Kays, Roland, 2021. Candid critters: challenges and solutions in a large-scale citizen science camera trap project. Citizen Science: Theory and Practice 6 (1), 4. https://doi.org/10.5334/cstp.343.

Louviere, Jordan J., Hensher, David A., Swait, Joffre D., 2000. Stated Choice Methods. Cambridge University Press, New York.

Louvrier, Julie L.P., Planillo, Aimara, Stillfried, Milena, Hagen, Robert, Börner, Konstantin, Kimmig, Sophia, Ortmann, Sylvia, Schumann, Anke, Brandt, Miriam, Kramer-Schadt, Stephanie, 2022. Spatiotemporal interactions of a novel mesocarnivore community in an urban environment before and during SARS-CoV-2 lockdown. J. Anim. Ecol. 91 (2), 367–380. https://doi.org/10.1111/1365-2656.13635.

Magle, Seth B., Lehrer, Elizabeth W., Fidino, Mason, 2016. Urban Mesopredator distribution: examining the relative effects of landscape and socioeconomic factors. Anim. Conserv. 19 (2), 163–175. https://doi.org/10.1111/acv.12231.

McCarthy, Maureen S., Stephens, Colleen, Dieguez, Paula, Samuni, Liran, Després-Einspenner, Marie-Lyne, Harder, Briana, Landsmann, Anja, Lynn, Laura K., Maldonado, Nuria, Ročkaiová, Zuzana, Widness, Jane, Wittig, Roman M., Boesch, Christophe, Kühl, Hjalmar S., Arandjelovic, Mimi, 2021. Chimpanzee identification and social network construction through an online citizen science platform. Ecol. Evol. 11 (4), 1598–1608. https://doi.org/10.1002/ece3.7128.

McShea, William J., Forrester, Tavis, Costello, Robert, He, Zhihai, Kays, Roland, 2016. Volunteer-run cameras as distributed sensors for macrosystem mammal research. Landsc. Ecol. 31 (1), 55–66. https://doi.org/10.1007/s10980-015-0262-9.

Murray, H. Maureen, Fidino, Mason, Lehrer, Elizabeth W., Simonis, Juniper L., Magle, Seth B., 2021. A multi-state occupancy model to non-invasively monitor visible signs of wildlife health with camera traps that accounts for image quality. J. Anim. Ecol. 90 (8), 1973–1984. https://doi.org/10.1111/1365-2656.13515.

Norouzzadeh, S. Mohammad, Morris, Dan, Beery, Sara, Joshi, Neel, Jojic, Nebojsa, Clune, Jeff, 2020. A deep active learning system for species identification and counting in camera trap images. Methods Ecol. Evol. 12 (1), 150–161. https://doi.org/10.1111/2041-210X.13504.

Pielou, C. Evelyn, 1966. The measurement of diversity in different types of biological collections. J. Theor. Biol. 13, 131–144.

Potter, Larrisa, Brady, Christopher J., Murphy, Brett P., 2019. Accuracy of identifications of mammal species from camera trap images: a northern Australian case study. Austral Ecol. 44, 473–483. https://doi.org/10.1111/aec.12681.

Rovero, Francesco, Zimmermann, Fridolin, Berzi, Duccio, Meek, Paul, 2013. "Which camera trap type and how many do I need?" a review of camera features and study designs for a range of wildlife research applications. Hystrix the Italian Journal of Mammalogy 24 (2), 148–156. https://doi.org/10.4404/hystrix-24.2-8789.

Schneider, Stefan, Greenberg, Saul, Taylor, Graham W., Kremer, Stefan C., 2020. Three critical factors affecting automated image species recognition performance for camera traps. Ecol. Evol. 10, 3503–3516. https://doi.org/10.1002/ece3.6147.

Simpson, Robert, Page, Kevin R., De Roure, David, 2014. Zooniverse: Observing the World's Largest Citizen Science Platform. WWW '14 Companion: Proceedings of the 23rd International Conference on World Wide Web, pp. 1049–1054. https://doi.org/10.1145/2567948.2579215.

Sullivan, Brian L., Wood, Christopher L., Iliff, Marshall J., Bonney, Rick E., Fink, Daniel, Kelling, Steve, 2009. EBird: a citizen-based bird observation network in the biological sciences. Biol. Conserv. 142 (10), 2282–2292. https://doi.org/10.1016/j.biocon.2009.05.006.

Swanson, Alexandra, Kosmala, Margaret, Lintott, Chris, Simpson, Robert, Smith, Arfon, Packer, Craig, 2015. Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. Scientific Data 2 (1), 150026. https://doi.org/10.1038/sdata.2015.26.

Swanson, Alexandra, Kosmala, Margaret, Lintott, Chris, Packer, Craig, 2016. A generalized approach for producing, quantifying, and validating citizen science data from wildlife images. Conserv. Biol. 30 (3), 520–531. https://doi.org/10.1111/cobi.12695.

Thel, Lucie, Chamaillé-Jammes, Simon, Keurinck, Léa, Catala, Maxime, Packer, Craig, Huebner, Sarah E., Bonenfant, Christophe, 2021. Can citizen science analysis of camera trap data be used to study reproduction? Lessons from snapshot Serengeti program. Wildl. Biol. 2021 (2), wlb.00833 https://doi.org/10.2981/wlb.00833.

Vélez, Juliana, McShea, William, Shamon, Hila, Castiblanco-Camacho, Paula J., Tabak, Michael A., Chalmers, Carl, Fergus, Paul, Fieberg, John, 2022a. An evaluation of platforms for processing camera-trap data using artificial intelligence. Methods Ecol. Evol. 14 (2), 459–477.

Vélez, Juliana, Castiblanco-Camacho, Paula J., Tabak, Michael A., Chalmers, Carl, Fergus, Paul, Fieberg, John, 2022b. Choosing an Appropriate Platform and Workflow for Processing Camera Trap Data Using Artificial Intelligence. Preprint. https://doi.org/10.48550/arXiv.2202.02283.

Wiggins, Andrea, Newman, Greg, Robert Stevenson, D., Crowston, Kevin, 2011. Mechanisms for data quality and validation in citizen science. In: IEEE 7th International Conference on E-Science. https://doi.org/10.1109/eScienceW.2011.27.

Willi, Marco, Pitman, Ross T., Cardoso, Anabelle W., Lock, Christina, Swanson, Alexandra, Boyer, Amy, Veldthuis, Marten, Fortson, Lucy, 2018. Identifying animal species in camera trap images using deep learning and citizen science. Methods Ecol. Evol. 10 (1), 80–91. https://doi.org/10.1111/2041-210X.13099.

Wilson, Joseph S., Pan, Aaron D., General, David Emmanuel M., Koch, Jonathan B., 2020. More eyes on the prize: an observation of a very rare, threatened species of Philippine bumble bee, Bombus irisanensis, on INaturalist and the importance of citizen science in conservation biology. J. Insect Conserv. 24 (4), 727–729. https://doi.org/10.1007/s10841-020-00233-3.

Woods, John W., 2012. Image Enhancement and Analysis. Multidimensional Signal, Image, and Video Processing and Coding, Second edition, pp. 223–256. https://doi.org/10.1016/B978-0-12-381420-3.00007-2.