Sharpe, CR, Hill, RA, Chappell, HM, Green, SE, Holden, K, Fergus, P, Chalmers, C and Stephens, PA

Increasing citizen scientist accuracy with artificial intelligence on UK camera-trap data

Remote Sensing in Ecology and Conservation Open Access ZSL

RESEARCH ARTICLE

Increasing citizen scientist accuracy with artificial intelligence on UK camera-trap data

C. R. Sharpe[1], R. A. Hill[2,3], H. M. Chappell[4], S. E. Green[2], K. Holden[5], P. Fergus[6], C. Chalmers[6] & P. A. Stephens[1,3]

[1]Conservation Ecology Group, Department of Biosciences, Durham University, Durham DH1 3LE, UK
[2]Department of Anthropology, Durham University, South Road, Durham DH1 3LE, UK
[3]MammalWeb, Durham Wildlife Trust, Chilton Moor, Houghton-Le-Spring, Tyne & Wear DH4 6PU, UK
[4]Rhombus Technology Ltd, 46 South Street, Durham DH1 4QP, UK
[5]Great North Museum: Hancock, Barras Bridge, Newcastle upon Tyne NE2 4PT, UK
[6]School of Computer Science and Mathematics, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK

Keywords
artificial intelligence, camera trapping, citizen science

Correspondence
C. R. Sharpe, Conservation Ecology Group, Department of Biosciences, Durham University, Durham DH1 3LE, UK. Tel.: +44 07939554364; Email: charlotte.sharpe@durham.ac.uk

Funding Information: This work is supported by ESRC IAA and EPSRC IAA, as well as funding from NERC (IAPETUS DTP PhD scholarship for S. Green; grant number NE/L002590/1). The Great North Museum: Hancock provided funding for camera traps and Mobile MammalWeb screens as well as facilitating the teacher training associated with this project. In addition, we thank the two anonymous reviewers and the associate editor who provided valuable feedback to support the production of the final version of this paper.

Editor: Dr. Marcus Rowcliffe
Associate Editor: Dr. Jorge Ahumada

Received: 24 December 2024; Revised: 4 April 2025; Accepted: 2 May 2025

doi: 10.1002/rse2.70012

Abstract

As camera traps have become more widely used, extracting information from images at the pace they are acquired has become challenging, resulting in backlogs that delay the communication of results and the use of data for conservation and management. To ameliorate this, artificial intelligence (AI), crowdsourcing to citizen scientists and combined approaches have surfaced as solutions. Using data from the UK mammal monitoring initiative MammalWeb, we assess the accuracies of classifications from registered citizen scientists, anonymous participants and a convolutional neural network (CNN). The engagement of anonymous volunteers was facilitated by the strategic placement of MammalWeb interfaces in a natural history museum with high footfall related to the 'Dippy on Tour' exhibition. The accuracy of anonymous volunteer classifications gathered through public interfaces has not been reported previously, and here we consider this form of citizen science in the context of alternative forms of data acquisition. While AI models have performed well at species identification in bespoke settings, here we report model performance on a dataset for which the model in question was not explicitly trained. We also consider combining AI output with that of human volunteers to demonstrate combined workflows that produce high accuracy predictions. We find the consensus of registered users has greater overall accuracy (97%) than the consensus from anonymous contributors (71%); AI accuracy lies in between (78%). A combined approach between registered citizen scientists and AI output provides an overall accuracy of 96%. Further, when the contributions of anonymous citizen scientists are concordant with AI output, 98% accuracy can be achieved. The generality of this last finding merits further investigation, given the potential to gather classifications much more rapidly if public displays are placed in areas of high footfall. We suggest that combined approaches to image classification are optimal when the minimisation of classification errors is desired.

Introduction

In the past 20 years, camera trapping has emerged as an efficient, low-impact method for mammal monitoring (Burton et al., 2015) that circumvents challenges of observing cryptic and nocturnal mammal species. The inferences that may be drawn from camera traps (CTs) extend to abundance estimation for species with either recognizable (Karanth, 1995) or unrecognisable (Mason et al., 2022; Rowcliffe et al., 2008) individuals, activity pattern analysis (Ridout & Linkie, 2009; Vazquez et al., 2019) and social network analysis (McCarthy et al., 2019). Increasing uptake of CTs for mammal research has been accompanied by advances in strategies for processing the images they collect (Green et al., 2020). One single trigger of a CT often yields a 'sequence' of images, so that each capture event (trigger of the motion sensor) represents a variable number of images generally assumed to contain the same subject (Hsing et al., 2022). While targeted surveys yield smaller datasets that can feasibly be classified by researchers, relying on experts restricts the geographic and temporal scope of monitoring. To this end, recent decades have seen three key innovations for processing images by (a) crowdsourcing to citizen scientists, (b) automated classification by artificial intelligence and (c) combined approaches of (a) and (b).

One breakthrough in image processing has been outsourcing the task to artificial intelligence (AI). Progress has been rapid since initial attempts at this task produced poor species recognition accuracy of around 38% (Chen et al., 2014). Subsequent innovations in AI, such as including a degree of manual image processing prior to automated processing by the AI model itself, have seen accuracy rise to 88% (Gomez Villa et al., 2017). More recently, an AI model trained for all species in the Snapshot Serengeti dataset achieved human-level accuracy (96.6%) for over 99% of the data (Norouzzadeh et al., 2018). AI models have performed with accuracy exceeding 95% for data from the United States and Europe, accompanied by advances in out-of-sample accuracy (i.e., accuracy on sites not included in the training data) (Böhner et al., 2023; Rigoudy et al., 2023; Tabak et al., 2020; Whytock et al., 2021). Despite the past decade of research advancing this field, AI outputs are not widely applied in ecological analyses or monitoring schemes. While the performance of models within sites they are trained on has evolved to produce accuracies of over 90% for North America, Central Africa and Europe (Fergus et al., 2024; Norouzzadeh et al., 2018; Rigoudy et al., 2023; Tabak et al., 2019; Whytock et al., 2021), performance is negatively affected when models are applied to new sites (Beery et al., 2018). Although some advances have been made in this regard (Rigoudy et al., 2023; Tabak et al., 2019; Whytock et al., 2021), it has so far not been possible for models to maintain their accuracy when applied to new sites, limiting their transferability and constraining the automation of image processing. Consequently, researchers typically still review AI predictions (Vélez et al., 2023).

Rather than striving for entirely automated approaches, many researchers recognize that maximizing time savings as well as accuracy might best be achieved by integrating the outputs of AI models with those of human classifiers (Adam et al., 2021; Green et al., 2020; Norouzzadeh et al., 2018; Whytock et al., 2021; Willi et al., 2019). For ecological monitoring, the integration of AI and citizen science offers fast-tracked data processing to accelerate positive conservation outcomes (McClure et al., 2020). It has been suggested that highly accurate human consensus could be used to train AI models and save experts the job of labelling training data (Willi et al., 2019), or that AI could reduce manual processing time by screening datasets for images devoid of animals or containing humans (Loos et al., 2018; McShea et al., 2016; Norouzzadeh et al., 2018; Willi et al., 2019). Combined approaches have also been reported to improve accuracy beyond what is possible from either individual method (Norouzzadeh et al., 2018; Trouille et al., 2019; Willi et al., 2019).

Parallel to the development of automated image processing, crowdsourcing has grown in its potential to be applied to CT data. In only 3 days, the Snapshot Serengeti project based in Tanzania achieved over 10.8 million online classifications from volunteers for over 1 million image sequences (Swanson et al., 2015). On average, each sequence received 27 classifications, and when aggregated, the species assignment matched expert labels with 98% accuracy, albeit with variation between species (Swanson et al., 2016). Through a similar UK-based platform, MammalWeb, it has been demonstrated that c. 9 concordant species classifications produce a consensus in which researchers can have 99% confidence (Hsing et al., 2018). Similarly, a US-based project found the agreement of 7 volunteers conveyed a classification deemed accurate 98% of the time (Rivera et al., 2024). For ecosystems where volunteers can expect to see highly charismatic species, as in Snapshot Serengeti, consensus classifications can be reached swiftly. However, recruitment is relatively more challenging in locations with less charismatic fauna, such as the UK. It is therefore desirable to identify species-and location-specific workflows that represent the most economical and accurate approach where fewer volunteers are available (Hsing et al., 2018). Further, the type of habitat surveyed and the camera settings used can influence the accuracy of human classifiers; volunteers have proven more accurate when three images are included per camera trigger compared with a single image and when videos are used instead of still images, and are more likely to provide false classifications in open-grassy habitats (Egna et al., 2020; Green et al., 2023).

MammalWeb is a network in which citizen scientists can become involved in both the deployment of CTs ('Trappers') as well as the classification of the images they produce ('Spotters') (Hsing et al., 2022). Since its inception in 2015, engagement with MammalWeb has expanded its geographic scope from the north-east of England to a national initiative. Typically, to contribute data, volunteers must register for an online account that ties their classifications to a traceable user ID. Registration potentially represents a barrier to participation for those who do not have access to computers or internet connection, as well as younger participants. Equally, it could deter first time users who would like to try the activity or people who do not wish to spend time setting up an account or share their contact details. Considering this, MammalWeb trialled the provision of public 'Mobile MammalWeb' terminals where individuals can contribute without the need for registration, potentially accelerating data acquisition (Hsing et al., 2022); however, there are legitimate concerns about the data quality, owing to the unknown motivations and expertise of short-term, low-engagement volunteers. Anonymous participation offers the easiest way for people to get involved and not only contribute data that can help to reach more rapid consensus classifications, but also widens access to engagement benefits of citizen science that can catalyse environmentally beneficial social change (Hesley et al., 2023; Jansen et al., 2024).

Here, we take advantage of the opportunity to compare the accuracies of classifications from different sources across a set of image sequences that have received classifications from registered citizen scientists, anonymous participants and an AI model, using a subset of MammalWeb images available via public Mobile MammalWeb terminals. Previous research suggests there is no difference in the accuracies of citizen scientists depending on whether they were logged in or contributing anonymously (through personal devices) to a particle physics project hosted on Zooniverse (Jackson et al., 2018). We test anonymous contributions in a different context; contributors were not actively seeking out the classification page, provided classifications through a public terminal rather than a personal device, and nor do we have evidence regarding their appreciation of the purpose of the task. This, therefore, represents a novel low-investment and short-term form of citizen science. In addition to comparing the classification accuracies of anonymous participants, registered citizen scientists and an AI model over an identical set of image sequences, we design two classification workflows that rely on input from either anonymous participants or registered citizen scientists, respectively, combined with output from the AI model. We predict that registered citizen scientists provide more accurate classifications than anonymous individuals due to a higher degree of engagement with the project. Further, we predict that combining both forms of citizen science with AI output will improve the predictive accuracy of outsourced classifications.

Materials and Methods

Timed to coincide with hosting the Natural History Museum's famous cast of a *Diplodocus* fossil between 18 May and 6 October 2019, the Great North Museum: Hancock in Newcastle, UK, installed several new exhibits. Among these, were five Mobile MammalWeb terminals located in different habitat sections of the Natural Northumbria gallery. Each screen was associated with one of the focal habitats in the gallery (lowlands, uplands, woodlands, coasts and urban). The screens hosted a modified MammalWeb interface, through which users could classify images from MammalWeb. The images, in turn, were contributed to MammalWeb by schools participating in the 'Dippy Schools Programme'. Fifty participating schools were each provided with a CT and, within a broader programme of continuing professional development, teachers attended a session on how and why to deploy CTs. Based on their locations, the schools were each allocated to one of the five major habitats, and it was the images uploaded from the relevant schools that appeared on each of the five touch-screen units. Upland habitats were under-represented, so images for that screen were augmented by those obtained from upland habitats during a simultaneous survey of County Durham (Mason et al., 2022). This set-up represents a method of increasing participation by eliminating the requirement for a personal device to contribute data. Camera settings and deployment characteristics are known to influence classification accuracy of the public (Egna et al., 2020), but we don't expect variation in these features identical camera models were used at each site. Further, teachers were given comprehensive guidance on the settings to use and how to deploy cameras in the field.

Starting with the image sequences that had been made available through terminals placed in the Great North Museum: Hancock, we identified a subset of sequences that had been classified by anonymous users, users registered through the MammalWeb platform, and the AI model to which the MammalWeb platform is linked (see 'AI Classifications', below). Some sequences on MammalWeb already had 'expert classifications', provided by affiliated researchers (Hsing et al., 2018). This was the case for most sequences in our focal subset. C. Sharpe (lead author) provided expert classifications for the small number of sequences missing expert labels. Although the potential for expert classifications to be biased is acknowledged, these labels are taken as the gold standard here, as has been protocol in similar analyses (Torney et al., 2019). C.S. also reviewed a small number of sequences for which there were multiple expert species labels to discern whether there were truly multiple species present, or whether there was a need for resolution of disagreement between experts. For sequences with multiple species present, all labels were preserved in the gold standard dataset. Images were programmed to appear randomly on Mobile MammalWeb units, and since we had no prior knowledge of the likely uptake or accuracy of anonymous classifiers, they were not set to retire after a given number of classifications as is protocol with image classification, due to our uncertainty regarding what would constitute a number of classifications that was both achievable and adequate.

Assessing accuracy

In the fields of machine learning and CT image classification, commonly reported metrics of accuracy are overall accuracy (overall proportion of predictions that were correct), as well as class-specific metrics of precision (proportion of positive identifications per class that are correct), recall (proportion of actual positives per class identified by the classifier) and F1 score (metric that describes harmonic mean of precision and recall). Overall accuracy is heavily influenced by the most popular classes (Norman et al., 2023) so here, class-specific precision is also reported to provide a more detailed picture of performance. For the purposes of biodiversity monitoring, researchers should strive for classification sources with high precision across species since precision reflects how often, when a user identifies species X, that species X is present (Rigoudy et al., 2023). The precision metric was deemed an appropriate measure to report in this investigation since it provides an intuitive measure of how confident researchers can be that a label provided by any classification approach is correct:

$$\text{Precision} = \frac{\text{True positives}}{\text{True positives} + \text{False positives}}$$

All analyses were conducted in R version 4.4.1 (R Core Team, 2024). Precision values and overall accuracy values were obtained using a multi-class confusion matrix in R, treating expert as ground truth and registered, anonymous and AI classifications as predicted labels, respectively, using the R package ‘ConfusionTableR’ (Hutson, 2021). Since the primary focus of the camera trapping is mammals, species-level bird classifications were condensed into one all-encompassing ‘Bird’ category. In addition to investigating the precision of classifiers, we include results characterizing the types of errors made by each group in the supplementary material. The following sections and Table 1 describe the acquisition of classifications from each source considered in this analysis.

Registered users

The classifications of registered users can be traced to their unique numerical identifiers. For this group, the analysis was conducted on three levels. First, every individual classification for each sequence in the data was compared with the gold standard to calculate class-specific precision scores that reflect the average precision of individual classifiers (individual level). Second, motivated by the high accuracy of previous records of aggregated volunteer input (Hsing et al., 2018; Swanson et al., 2016), individual classifications were aggregated into a single consensus species for each sequence that had been classified by multiple users (Hsing et al., 2018; Swanson et al., 2016). The consensus is defined as the species with the most common classification (consensus level). For sequences that had equal votes for more than one species, the most recent classification was excluded to break the tie. This left a very small number of sequences that were still tied (with more than two species included

Table 1. Features of the sources of image classification considered in this analysis.

Source	Approach	Image label compared with expert
Registered	First	Earliest classification per sequence
	Individual	All individual user classifications considered, and average accuracy and precision reported
	Consensus	Single consensus species defined per sequence
Anonymous	Individual	All classifications considered, and average accuracy and precision reported
	Consensus	Single consensus species defined per sequence
AI	Maximum probability	Species with single highest probability value across all images in each sequence
	Highest summed probability	Species with highest summed probability value across all images in each sequence
Combined algorithm	1. First registered user = maximum probability AI	For sequences where the first registered user and maximum probability AI classification match, this label compared with expert
	2. Registered consensus = maximum probability AI	For sequences where the registered consensus and maximum probability AI classification match, this label compared with expert
	3. Anonymous consensus = maximum probability AI	For sequences where anonymous consensus and maximum probability AI classification match, this label compared with expert

in a tie). These were excluded, and any sequences with only one classification were excluded. The number of sequences excluded according to these criteria is reported in the results (Fig. 4). Third, the earliest classification provided for each sequence was isolated and compared with the gold standard (first user level).

Anonymous users

Classifications received via Mobile MammalWeb terminals cannot be attributed to unique users, as the user ID corresponds to the terminal rather than the human that provided the classification. Classifications from these users were considered in two ways. First, each individual classification for each sequence in the data was compared with the gold standard to produce a precision score that reflects the average precision of individual classifications (average classification precision). Second, individual classifications were aggregated into a consensus per sequence and compared with the gold standard to determine precision (consensus precision). For sequences where more than one species was tied for the consensus, the most recent classification for either of the tied species was excluded. Those that remained tied (since more than two species had the maximum number of votes) were excluded. Additionally, any sequences with only one vote were excluded. Again, as with the registered consensus approach, the number of sequences excluded according to these criteria is reported in the results (Fig. 4).

AI classifications

For image data uploaded to MammalWeb, species identifications are provided by passing images through Conservation AI's 'UK Mammals' model (www.conservationai.co.uk; a deep convolutional neural network based on Faster RCNN Resnet 10; Vélez et al., 2023; Fergus et al., 2024). The model classifies every image, in contrast to human classifiers, who provide classifications at the level of an image sequence. To compare the AI's output to the gold standard, the classifications were aggregated to a sequence level prediction either by selecting the species with the single highest assigned probability across all predictions for that sequence, or by selecting the species with the highest summed probability across the sequence. The AI model is not trained to recognize certain species identified by experts in our data, including brown (European) hare, horse, stoat, small rodent (unknown species), or brown rat, so it was not possible to assess AI accuracy for these species. Only predictions with a probability over 0.55 are returned by Conservation AI. Although it is common to report accuracy for the top-1 and top-5 classes predicted by AI models (Gomez Villa et al., 2017; Norman et al., 2023; Norouzzadeh et al., 2018; Tabak et al., 2020; Whytock et al., 2021), we do not report these metrics owing to the low number of species reported per sequence. Instead, the class-specific precisions for the AI's highest probability predictions are considered.

Combining AI and citizen science

We also considered the performance of three algorithms designed to integrate the inputs of citizen science and AI. Combination rule 1 considers concordance between the first registered user classification and the AI model, while combination rule 2 uses the consensus of registered users in agreement with the AI model. Combination rule 3 considers the anonymous user consensus in agreement with the AI. In all cases, the AI's sequence level prediction was taken to be the maximum probability AI species, since this classification resulted in higher precision for the most classes.

Combined workflows are intended to increase both the efficiency and accuracy of image classifications (Green et al., 2020), and due to differences between the accuracies of registered or anonymous users, our workflows differed depending on who provided the human classification (Table 1). While the consensus from either registered or anonymous users was considered in combination with AI in rules 2 and 3, for registered users, the finding that the first registered user to classify a sequence is typically highly accurate (Fig. 1; Hsing et al., 2018) led us to test a workflow that relied only upon one human classification in combination with AI (rule 1). While registered user consensus is proven to be highly accurate without the input of AI (Fig. 1), combined rule 1 represents a method of increasing efficiency relative to combination rule 2 by eliminating the requirement to wait for a citizen science consensus before sequences can be retired with high confidence.

Results

For registered and anonymous participants as well as the AI model, accuracy was assessed over 4849 sequences that had been classified by all groups. Expert classifications indicated that only 0.49% (24) of sequences contained more than one species. A mean of 6.67 registered users (95% CI = 6.54, 6.81) had classified each sequence, identifying a mean of 1.35 species per sequence (95% CI = 1.33, 1.37). By comparison, a mean of 8.08 anonymous users (95% CI = 7.75, 8.42) had classified the same sequences, identifying more than twice as many species per sequence (3.14, 95% CI = 3.07–3.21). Therefore, the classifications of anonymous users include more incorrect classifications per sequence. An average of 1.25 (95%

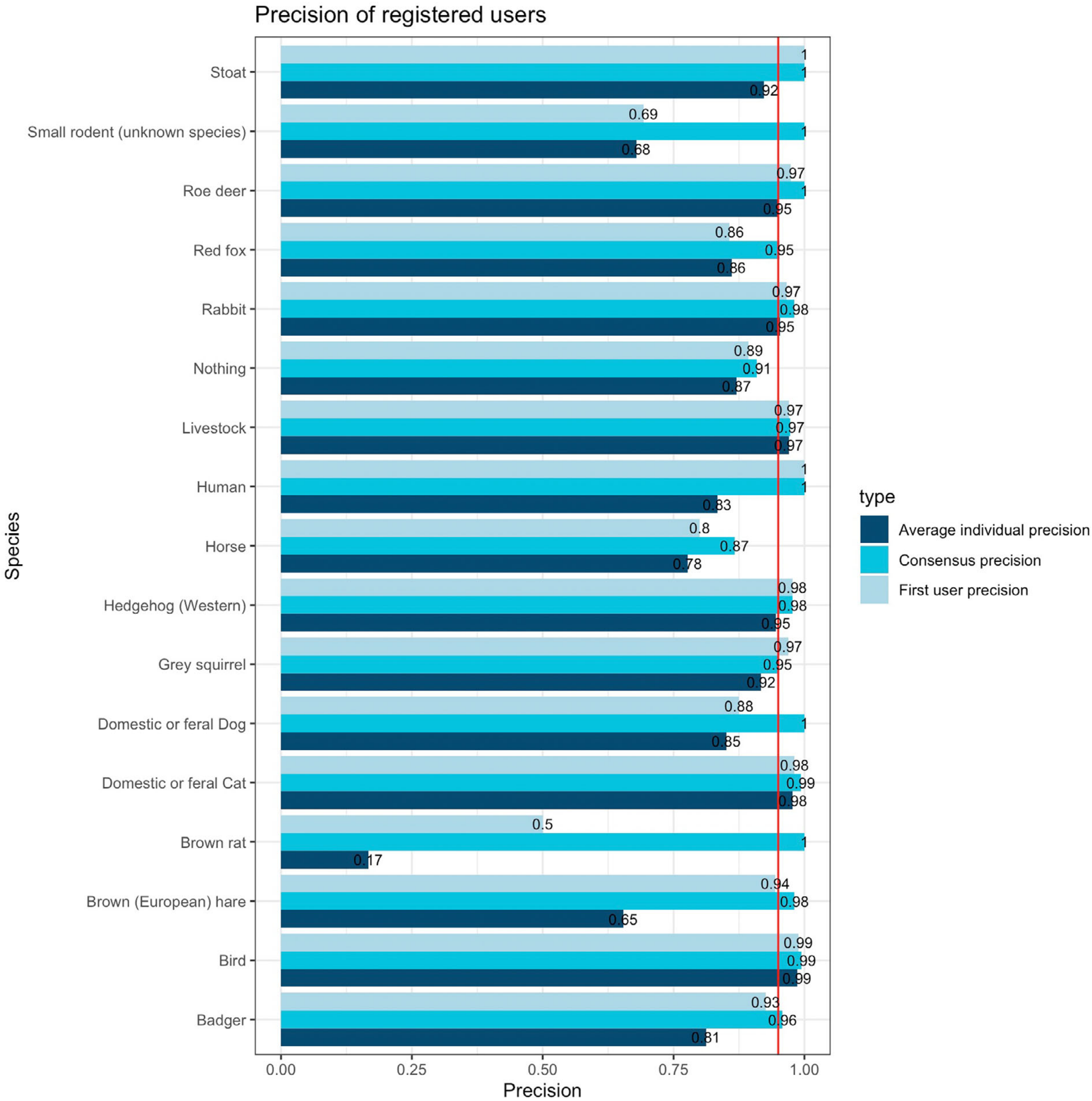

Figure 1. Class-specific precision values for registered users across 4849 sequences considering classifications at three levels: average individual precision—average precision of individual classifications provided by registered users, consensus precision—species with the greatest number of votes from registered users per sequence used to calculate precision, first user precision—only the earliest registered classification per sequence used to calculate precision. Vertical red line indicates 0.95 precision value.

CI = 1.23, 1.27) different species were identified per sequence by the AI model.

Registered users

At the individual level, the overall accuracy of registered users was 0.94, but average precision varied across classes (Fig. 1). Considering only the prediction of the first registered user to classify a sequence, the overall accuracy was 0.95 with similar variation in cross-class precision, but generally improved precision scores. Consensus classifications had an overall accuracy of 0.97, with only the classes 'Nothing' and 'Horse' having consensus precision below 0.95. The number of votes for the consensus was significantly positively associated with the classification being correct (OR = 1.16; 95% CI = 1.11, 1.21;

$P < 0.001$). For each additional vote, a given consensus had 16% improved odds of being correct. Once a consensus has 6 votes, the probability of it being correct was greater than 0.95.

Anonymous users

Overall accuracy averaged across all anonymous classifications for all sequences was 0.55, markedly lower than that of registered users. Average precision was below 0.5 for most classes (Fig. 2). Using the consensus approach improves the overall accuracy to 0.71, and the precision improves for every class except stoat, brown rat and bird. For each additional vote in favour, a given consensus classification had 36.8% greater odds of being correct (OR = 1.37; 95% CI = 1.32, 1.42; $P < 0.001$). To exceed 0.95 probability of any given anonymous consensus classification being correct, 12 votes were required. After

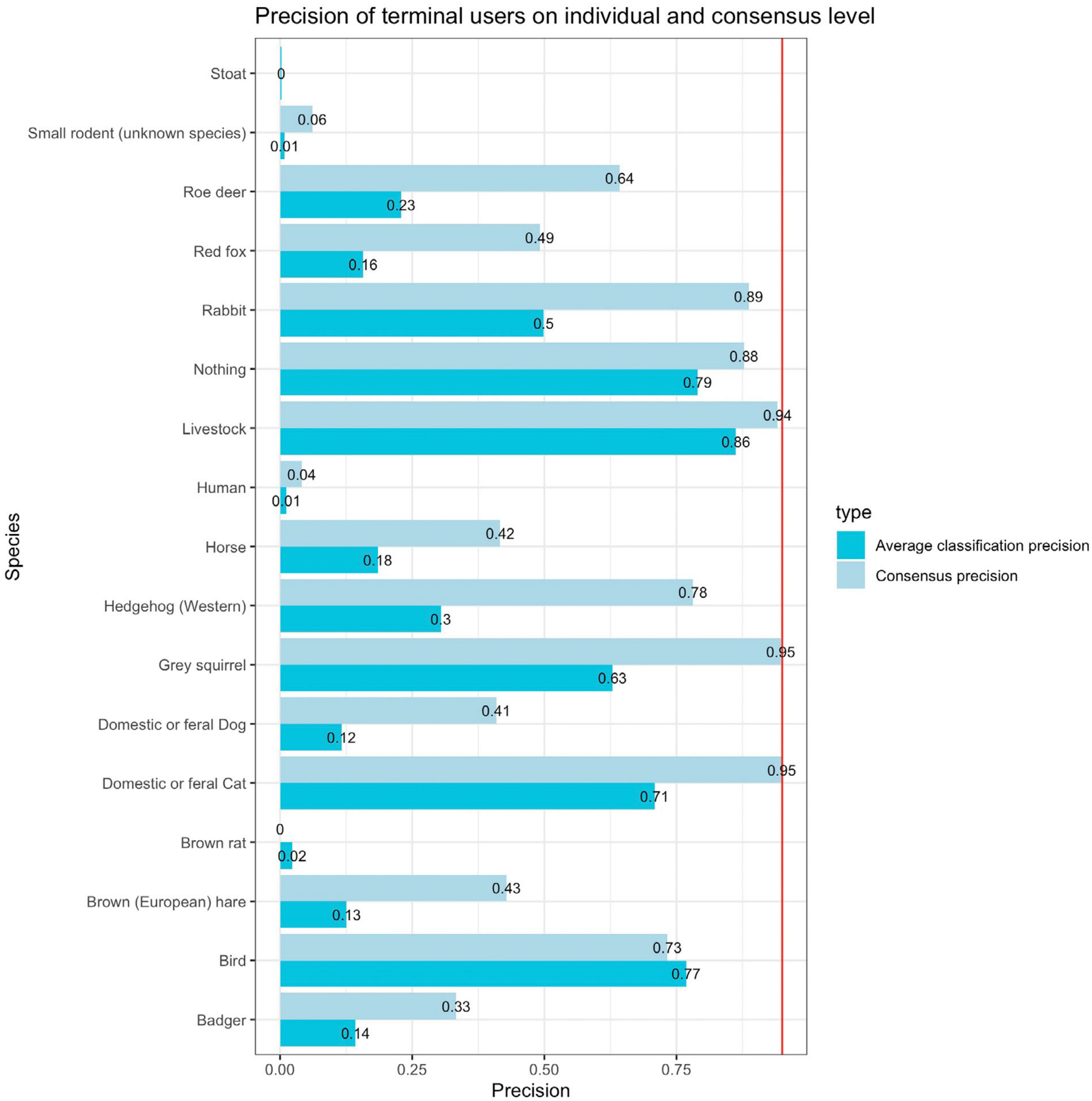

Figure 2. Class-specific precision values for anonymous users across 4849 sequences considering classifications at two levels: average classification precision—average precision of individual classifications provided via Mobile MammalWeb units, consensus precision—species with the greatest number of votes from anonymous classifications per sequence used to calculate precision. Vertical red line indicates 0.95 precision value.

adjusting for the number of votes per sequence, there was a significant difference between registered and anonymous users. An anonymous consensus had 85.8% lower odds of being correct than a registered consensus with the same number of votes (AOR = 0.14; 95% CI = 0.12, 0.17; $P < 0.001$).

AI classifications

Using the species classification with the maximum probability per sequence resulted in an overall accuracy of 0.78, although precision varied across species, and only the Bird category had precision above 0.95. The highest summed probability species showed similar results with an overall accuracy of 0.78 and similar between-species variation in precision, again with only the Bird category exceeding 0.95 precision. There was no significant difference in the odds of either the maximum probability species or highest summed probability species being correct (OR = 0.998, 95% CIs = 0.90, 1.10, $P = 0.97$).

Combining AI and citizen science

For combination rule 1, 3676 out of 4849 sequences had a classification from the first registered user that agreed with the maximum probability AI species per sequence. The overall accuracy compared with expert review across these sequences was 0.96 and the precision for every animal class exceeded 0.95 (Fig. 3), so for all sequences where the presence of an animal is agreed upon by the first registered user and the maximum probability AI prediction, the species assignment could be accepted with 95% confidence. Only the 'Nothing' class did not meet the 0.95 value for precision. Using this rule, 2230 labels could be accepted with 95% confidence.

Using combined rule 2, the registered consensus agreed with the maximum probability AI species per sequence for

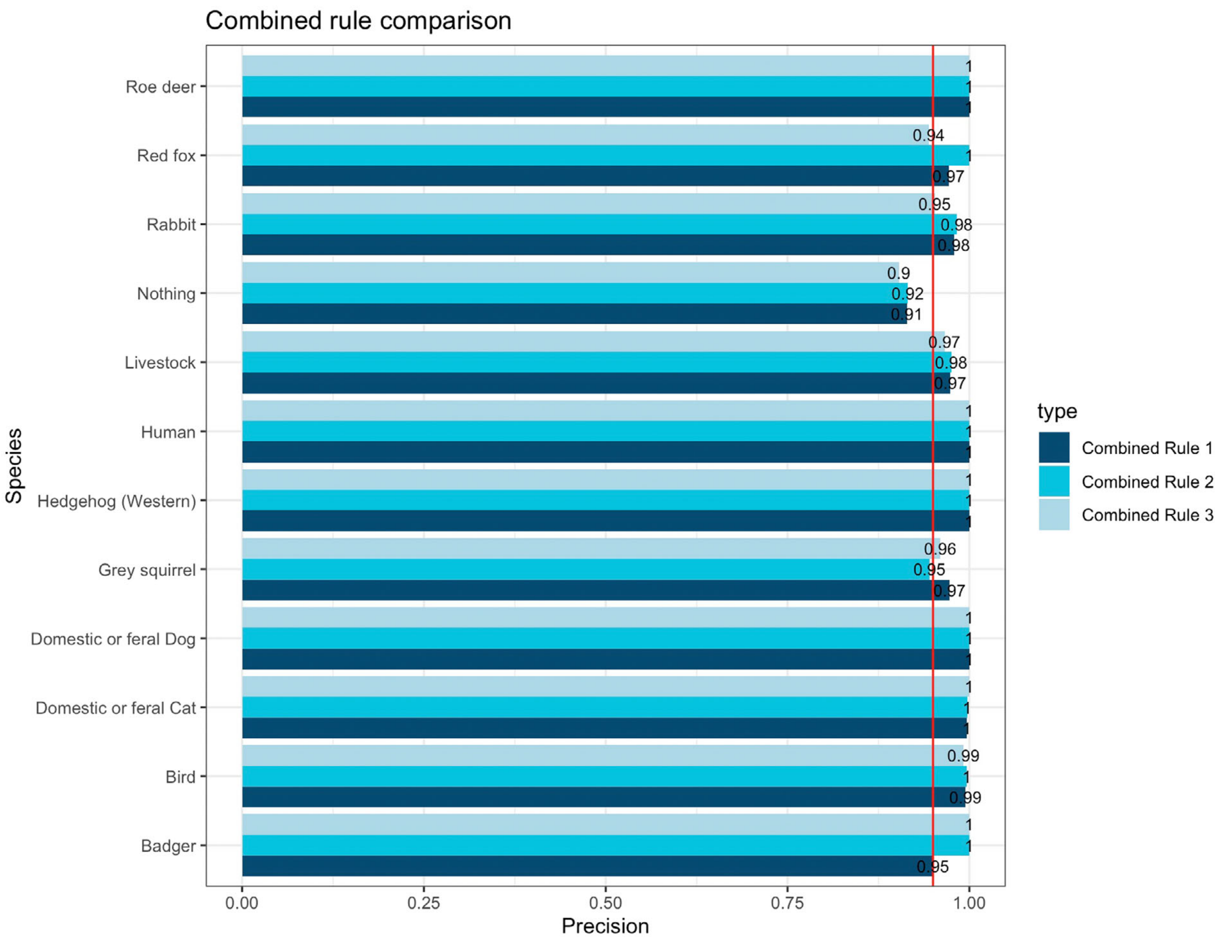

Figure 3. Comparison of precision scores for combination rules 1 (first registered user species = AI max probability species), 2 (registered consensus species = AI max probability species) and 3 (anonymous consensus species = AI maximum probability species).

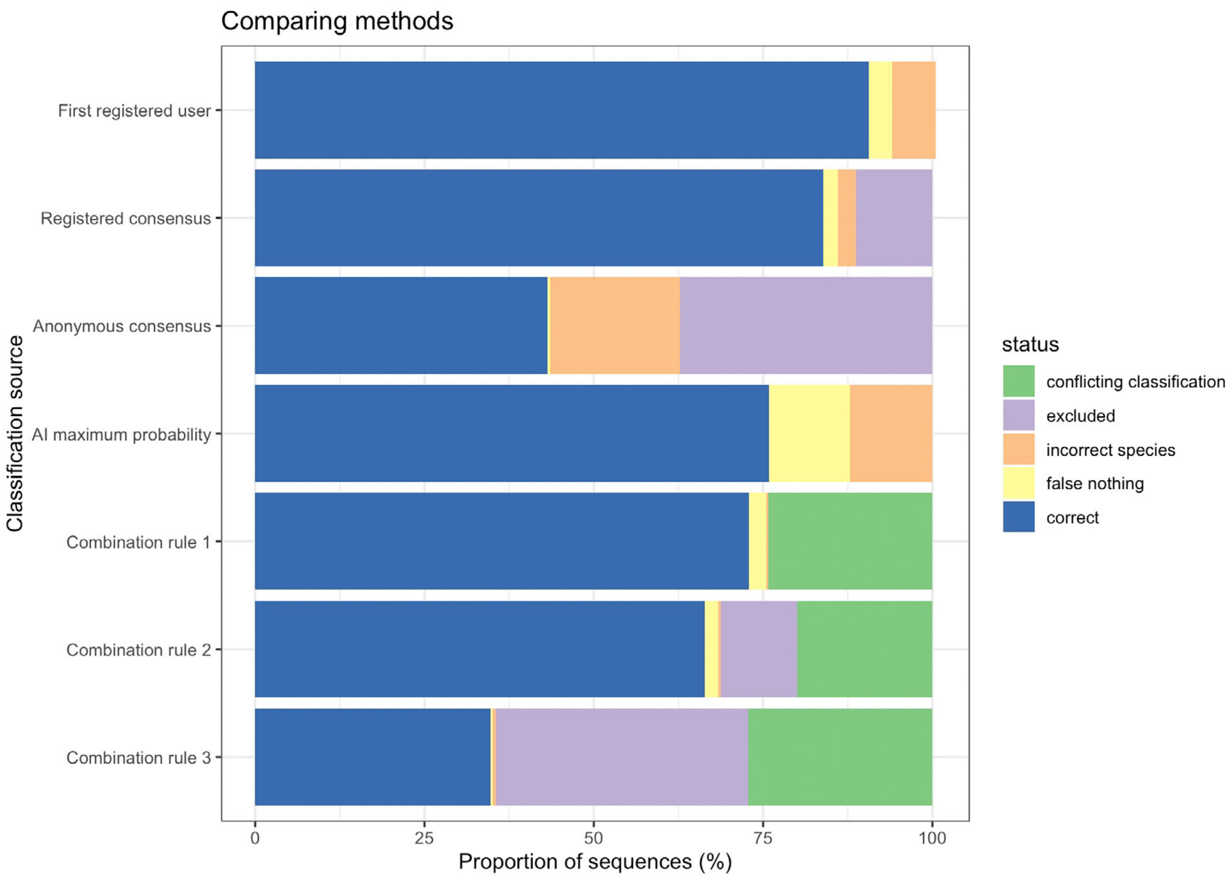

Figure 4. Status of 4849 sequences considered in this analysis by each method of acquiring classifications. Correct—label matches expert label, excluded—sequences excluded because they did not meet classification criteria to be analysed (see Methods), conflicting classification—AI and human labels are incongruous, false nothing—'nothing' identified but expert identified animal, incorrect species—animal identified that does not match expert label.

3321 sequences. The overall accuracy was 0.97 and the precision was 0.95 or higher for every animal class (Fig. 3), and only below 0.95 for Nothing, meaning that—as for combination rule 1—every animal classification can be accepted with at least 95% confidence in its accuracy (2141 sequences).

From combination rule 3, the terminal consensus matched the AI output for 1713 sequences. The overall accuracy was 0.98, and most classes had precision scores over 0.95, except for red fox and Nothing (Fig. 3). Despite the high accuracy, only 1567 sequences can be retired with this rule compared with 2230 and 2141 by rules 1 and 2, respectively, resulting in a greater proportion with conflicting classifications that require manual review (Fig. 4).

Discussion

Motivated by the pressing requirement for efficient methods of processing mounting volumes of CT data, we provide insight into the accuracies of different approaches to image classification. Here, we considered the accuracies of registered users, anonymous users and an AI model, along with two combined workflows, in classifying images from a UK mammal monitoring project. The accuracy of registered users was far greater than that of either anonymous citizen scientists or AI, particularly when individual labels were aggregated into a consensus. However, combined approaches yielded predictions of greater accuracy than is possible from any individual source. We discuss our results in relation to how the relative engagement of citizen scientists influences the accuracy of their classifications, the benefits of AI both as a standalone approach or a tool for enhancing citizen science, and finally, the potential for anonymous contributors to provide scientifically valuable data.

Influence of engagement on human accuracy

A clear outcome of our analysis was that the greater investment or interest of volunteers that register for a

MammalWeb account prior to providing their classifications translates into superior predictive accuracy compared with anonymous users. These results echo previous findings that registered citizen scientists are highly accurate via MammalWeb and other projects, particularly when their contributions are aggregated into a consensus (Hsing et al., 2018; Swanson et al., 2016). It has previously been reported, in research concerned with user experience, that the predictive accuracy of anonymous citizen scientists did not differ from those who were logged in through the Zooniverse website (Jackson et al., 2018). However, these participants actively sought out the classification task to which they contributed, whereas the anonymous MammalWeb participants were not specifically seeking to engage with the project and are likely to have had a relatively low degree of engagement. This could account for the difference observed here that was not observed via Zooniverse. Nevertheless, our results may prompt future consideration of the variation between forms of citizen science that, in the case of MammalWeb, at least, must be treated differently.

Relatively lower precision results for small body-sized mammals from both anonymous and registered citizen scientists likely reflect the reduced ability of conventional CTs to capture small mammals. This is well-established in CT research and has fuelled a separate strand of research optimizing CTs for small mammals (Glen et al., 2013; Littlewood et al., 2021; McCleery et al., 2014).

Value of AI as standalone and combined approach

Our finding that Conservation AI's UK Mammal model performed with an overall accuracy of 78% represents a slightly lower accuracy than expected since the model has been reported to operate with a mean average precision of 0.976 (Fergus et al., 2024), although it is acknowledged that the model is sensitive to variation in image quality. The deviation from expected accuracy is likely attributable to features of the images that come from a high diversity of sites that are not included in the training data for this model. The inability of AI models to transfer to new sites and maintain their accuracy (Shepley et al., 2021) remains a major obstacle to the introduction of large-scale automated monitoring. One suggestion has been to use higher resolution, publicly available imagery from sites such as FlickR and iNaturalist to train AI models, only later supplementing their training with CT data (Shepley et al., 2021). Attempts at supplementing training data with imagery such as that from iNaturalist have produced impressively accurate AI models for camera-trap data (Schneider et al., 2024), demonstrating promise in this approach. Using publicly available images for training can also provide a solution where species are not commonly captured by CTs, such as birds, which consequently do not have easily accessible training datasets (Chalmers et al., 2023). Improvements to the performance of AI models more generally could also be accessed by including metadata with training images, such as temperature, location and time (Tøn et al., 2024).

There is great potential in applying AI in combination with human classifications (Green et al., 2020). Combined approaches have enjoyed great success in other realms through the platforms eBird (Kelling et al., 2013) and iNaturalist (Ceccaroni et al., 2019). Additionally, the positive benefits of citizen science stretch beyond the scientific enterprise into social life (Eichholtzer et al., 2023), highlighting the wider value of reserving a space for citizen science in image classification workflows. A pervasive concern expressed in recent publications is that there is a fine line to tread in balancing the implementation of AI with the maintenance of a role for volunteers to provide citizen science outcomes that benefit science and society (Fortson et al., 2024; Pankiv & Kloetzer, 2024; Sharma et al., 2024). Rather than using AI to reduce the role of humans, we have demonstrated here that AI models can facilitate the use of contributions from low-expertise volunteers in monitoring schemes. We found that for most sequences where the species label was concordant between the anonymous consensus and the AI prediction, precision exceeded 0.95 (Fig. 4, combined rule 3). This supports the finding that although involving humans increases the time taken to obtain classifications, doing so reduces errors (Huebner et al., 2024). In addition, we demonstrated a workflow where accurate image labels can be gathered rapidly by relying on only a single classification from a registered MammalWeb user. Where the first registered MammalWeb Spotter identifies a species, and that species is also identified by the AI model, that label can be accepted with over 0.95 precision regardless of species. This result bolsters previous reports that only one or two human classifications that accord with AI can reduce processing time (Willi et al., 2019).

Considering classifications from registered MammalWeb classifiers, the first registered user yields a slightly greater proportion of sequences with correct classifications than the registered consensus, but the proportion of sequences with false nothing or incorrect species classifications is reduced when the consensus is used. Given the much greater speed of acquiring a single classification than waiting for a consensus, there is clearly a trade-off between speed and accuracy. Combined rules 1 and 2 that utilize input from the first registered user and registered consensus, respectively, yield a proportion of sequences with conflicting classifications between humans and AI that would require subsequent manual review but, crucially, they result in a reduction of incorrect species and false nothing classifications relative to the individual methods in isolation. Again, this points to a trade-off between speed and minimization of errors. A similar effect results from combined rule 3, which markedly lowers the proportion of incorrect species classifications compared with the anonymous consensus but leaves a high proportion of sequences unclassified either with no human consensus species or with conflicting labels between AI and humans (Fig. 4).

Although the precision of each combined method is even across species and consistently high (Fig. 3) different benefits are gained from the different approaches. Our recommendations regarding the best approach depend on what users stand to gain, or what projects aim to achieve. If the intention is to include contributions of anonymous users from public terminals in wildlife monitoring, the best approach would be to use the anonymous consensus in combination with AI, and sequences with conflicting classifications or no consensus species could be redirected to registered users for classification. If the first registered user classification is consistent with the AI prediction, the sequence can be retired, or else a registered consensus should be sought. Contrastingly, if the objective is to determine correct image labels as quickly as possible, retiring sequences after one registered user classification accords with the maximum probability AI prediction (Fig. 3, combined rule 1) offers the quickest approach, and sequences with conflicting classifications from this approach could also be redirected to registered users until a consensus is reached. However, there are circumstances where instead of prioritizing time savings, researchers may wish to minimize errors when aiming to detect individual species; for instance, if they are of conservation priority. The hedgehog *Erinaceus europaeus*, for example, is currently receiving focused monitoring effort with CTs due to alarming declines (Evans et al., 2024). In this instance, a combined approach would be favourable due to the lower proportion of incorrect species and false nothing classifications for this species (Supporting Information).

A consistent result is that precision for the Nothing and Horse classes never exceeded 0.95 by any classification approach. It has previously been noted that achieving a high precision for 'Nothing' sequences in MammalWeb takes large numbers of concordant votes (Hsing et al., 2018). This is because false-positives (i.e. suggesting that an animal is present when the sequence is in fact devoid of animals) are extremely rare. Consequently, a single vote for any species can undermine conviction in large numbers of votes for 'nothing'. By contrast, volunteers frequently make 'false nothing' classifications—perhaps because, in images where there is only a partial or blurred view of an animal, classifiers prefer to classify the image as 'Nothing' than guess the species identity (Swanson et al., 2016). Relatively low confidence in 'nothing' classifications represents a significant barrier to efficient workflows, and limits the generality of consensus algorithms by necessitating an alternative approach for retiring images alleged to contain nothing. One suggestion has been to ask those who upload data to screen their own images for blanks prior to upload to reduce room for error (Hsing et al., 2018), or to invest in AI models that can complete this task with high accuracy (Norouzzadeh et al., 2018); this supports the recent integration of MegaDetector (Beery et al., 2019) into the MammalWeb workflow. In previous analyses the agreement of nine or 10 volunteers has produced higher confidence in nothing classifications (Hsing et al., 2018; Swanson et al., 2016). Here, the consensus of registered or anonymous users was based on a median of six or four classifications, so further classifications for consensus 'Nothing' sequences could improve precision for this class of images. Regarding horses, assessment of sequences to produce the gold standard suggested that domestic horses were often visible in the background of images with other wild animals in the foreground that were clearly responsible for the trigger. In these instances, we suggest that volunteers have classified the animal that has more obviously triggered the camera, disregarding domestic horses. This is not a matter of huge concern, since domestic horses are not of particular interest in the context of monitoring wild animals.

Role of anonymous terminals in scientific engagement

Although the focus of this analysis is the accuracy of outsourced classifications, it is also important to consider how mobile MammalWeb units facilitate greater engagement with science. Interacting with MammalWeb has previously been demonstrated to have positive impacts on school students by improving knowledge of local wildlife and providing the satisfaction of contributing to science (Hsing et al., 2020). Out with the MammalWeb project, there is broader evidence that participating in citizen science increases participants knowledge base (Masters et al., 2016), inspires advocacy for nature (Forrester et al., 2017) and provides enjoyment to those who take part (Jansen et al., 2024), as well as providing experiences with nature that are becoming less accessible (Schuttler et al., 2018). By placing units in a public place, engagement with MammalWeb is extended beyond those who have access to a personal device and to those who would not otherwise be aware of the project. During the period that MammalWeb units were installed in the Great North Museum: Hancock, S.G. completed surveys with participants that indicated participants had enjoyed taking part and frequently had no previous experience of an equivalent task, evidencing the ability of the scheme to engage new audiences and generate further engagement. This contributes to a more diverse set of users and increases the accessibility of citizen science and its benefits.

To date, the accuracy of anonymous contributors has not been explored. The lower degree of investment and engagement of anonymous contributors appears to translate to lower accuracy. There is, thus, a challenge to be overcome in providing a level of instruction that increases accuracy without undermining the ease of use associated with a public terminal. Despite this, we demonstrated that an accurate signal could be extracted from anonymously contributed data by relying on concordance with the maximum probability AI classification. $\sim$32% of the 4849 sequences considered here could be retired with over 0.95 precision using this workflow, although a greater proportion had correct labels when registered user contributions were used (Fig. 4). Since the probability of an anonymous consensus classification being correct increases by 36.8% with each additional vote in agreement, it would be possible for the accuracy of anonymous contributions to be shifted closer to that of registered users if the number of classifications gathered can be increased. The importance of placing anonymous devices in areas of high throughput is therefore paramount in extracting an accurate signal. Overall, there is scientific value to be gained from the contributions of anonymous citizen scientists, in addition to the societal benefits of public engagement, encouraging future use of similar devices at appropriate locations such as natural history museums and science fairs. Public interfaces such as the Mobile MammalWeb units described here are not widely used, despite their ability to engage a wider group of volunteers, and we suggest that they could be used to scale the engagement of anonymous participants in camera-trap image classification. Aside from locations in urban centres such as museums and science centres that could host Mobile MammalWeb units, there are alternative locations that could host units such as nature reserve visitor centres, bird hides, community hubs, libraries, or local fairs. To this end, representatives from MammalWeb have been in discussion with several national NGOs that have widespread visitor centres, which seems a promising avenue to scale public terminal engagement, especially if visitors to those locations can assist with classifying data from the locations they are visiting. Alternatively, Mobile MammalWeb units, or tablets, could be loaned to schools where teachers could perhaps integrate a MammalWeb lesson into the curriculum (Hsing et al., 2020) in which the tablet screen could be projected to the class and images collaboratively classified with students to teach identification of UK mammal species.

Rather than discarding data owing to concerns regarding the accuracy of anonymous participants, we have demonstrated that highly precise image classifications can be redeemed when anonymous contributions are considered in tandem with AI output, encouraging further initiatives that engage anonymous volunteers in science. In the grander scheme of CT image processing, the findings reported here should provide assurance that outsourcing to citizen science and AI networks can alleviate classification bottlenecks with trustworthy classifications, particularly when combined workflows are implemented.

Acknowledgements

We thank S. Mason and participating schools for contributing images to MammalWeb that were analysed in this research. This work is supported by ESRC IAA and EPSRC IAA, as well as funding from NERC (IAPETUS DTP PhD scholarship for S. Green; grant number NE/L002590/1). The Great North Museum: Hancock provided funding for CTs and Mobile MammalWeb screens as well as facilitating the teacher training associated with this project. In addition, we thank the two anonymous reviewers and the associate editor who provided valuable feedback to support the production of the final version of this paper.

Author Contributions

C. R. Sharpe: Formal analysis; writing – original draft; conceptualization. **R. A. Hill:** Conceptualization; supervision; writing – review and editing. **H. M. Chappell:** Data curation; writing – review and editing. **S. E. Green:** Data curation; writing – review and editing. **K. Holden:** Data curation; writing – review and editing. **P. Fergus:** Software; writing – review and editing. **C. Chalmers:** Software; writing – review and editing. **P. A. Stephens:** Conceptualization; supervision; writing – review and editing.

Disclosure

Intellectual contributions from anonymous participants in the north-east of England were facilitated by the provision of public mobile MammalWeb terminals that allow individuals to anonymously contribute species classifications to camera-trap data that was collected by local school groups.

Data Availability Statement

Data available from the Figshare Repository https://figshare.com/projects/Supporting_data_for_Sharpe_et_al_

2025_Increasing_citizen_scientist_accuracy_with_artificial_intelligence_on_UK_camera_trap_data_published_in_Remote_Sensing_in_Ecology_and_Conservation_/247712.

References

Adam, M., Tomášek, P., Lehejček, J., Trojan, J. & Jůnek, T. (2021) The role of citizen science and deep learning in camera trapping. *Sustainability*, **13**(18), 10287. https://doi.org/10.3390/su131810287

Beery, S., Morris, D. &Yang, S. (2019) Efficient pipeline for camera trap image review. *arXiv*. https://doi.org/10.48550/arXiv.1907.06772.

Beery, S., Van Horn, G. & Perona, P. (2018) Recognition in Terra Incognita. In: *Computer vision—ECCV 2018. European conference on computer vision*. Cham: Springer, pp. 472–489. https://doi.org/10.1007/978-3-030-01270-0_28

Böhner, H., Kleiven, E.F., Ims, R.A. & Soininen, E.M. (2023) A semi-automatic workflow to process images from small mammal camera traps. *Ecological Informatics*, **76**, 102150. https://doi.org/10.1016/j.ecoinf.2023.102150

Burton, A.C., Neilson, E., Moreira, D., Ladle, A., Steenweg, R., Fisher, J.T. et al. (2015) REVIEW: wildlife camera trapping: a review and recommendations for linking surveys to ecological processes. *Journal of Applied Ecology*, **52**(3), 675–685. https://doi.org/10.1111/1365-2664.12432

Ceccaroni, L., Bibby, J., Roger, E., Flemons, P., Michael, K., Fagan, L. et al. (2019) Opportunities and risks for citizen science in the age of artificial intelligence. *Citizen Science: Theory and Practice*, **4**(1), 29. https://doi.org/10.5334/cstp.241

Chalmers, C., Fergus, P., Wich, S., Longmore, S.N., Walsh, N.D., Stephens, P.A. et al. (2023) Removing human bottlenecks in bird classification using camera trap images and deep learning. *Remote Sensing*, **15**(10), 2638. https://doi.org/10.3390/rs15102638

Chen, G., Han, T.X., He, Z., Kays, R. & Forrester, T. (2014) Deep convolutional neural network based species recognition for wild animal monitoring. In: *2014 IEEE international conference on image processing (ICIP)*. Paris, France. New York City: Institute of Electrical and Electronics Engineers, pp. 858–862. Available from: https://doi.org/10.1109/ICIP.2014.7025172

Egna, N., O'Connor, D., Stacy-Dawes, J., Tobler, M.W., Pilfold, N., Neilson, K. et al. (2020) Camera settings and biome influence the accuracy of citizen science approaches to camera trap image classification. *Ecology and Evolution*, **10**(21), 11954–11965. https://doi.org/10.1002/ece3.6722

Eichholtzer, A.C., Driscoll, D.A., Patrick, R., Galletta, L. & Lawson, J. (2023) The co-benefits of biodiversity citizen science for well-being and nature relatedness. *Applied Psychology. Health and Well-Being*, **16**(2), 515–536. https://doi.org/10.1111/aphw.12502

Evans, B.C., Rowcliffe M., Carbone C., Cartledge E.L., Al-Fulaij N., Pringle H. et al. (2024) Lens on the wild: innovations in wildlife monitoring with machine learning. Available from: https://www.the-ies.org/analysis/lens-wild-innovations-wildlife-0 [Accessed 19 December 2024].

Fergus, P., Chalmers, C., Longmore, S. & Wich, S. (2024) Harnessing artificial intelligence for wildlife conservation. *Conservation*, **4**(4), 685–702. https://doi.org/10.3390/conservation4040041

Forrester, T.D., Baker, M., Costello, R., Kays, R., Parsons, A.W. & McShea, W.J. (2017) Creating advocates for mammal conservation through citizen science. *Biological Conservation*, **208**, 98–105. https://doi.org/10.1016/j.biocon.2016.06.025

Fortson, L., Crowston, K., Kloetzer, L. & Ponti, M. (2024) Artificial intelligence and the future of citizen science. *Citizen Science: Theory and Practice*, **9**(1), 32. https://doi.org/10.5334/cstp.812

Glen, A.S., Cockburn, S., Nichols, M., Ekanayake, J. & Warburton, B. (2013) Optimising camera traps for monitoring small mammals. *PLoS One*, **8**(6), e67940. https://doi.org/10.1371/journal.pone.0067940

Gomez Villa, A., Salazar, A. & Vargas, F. (2017) Towards automatic wild animal monitoring: identification of animal species in camera-trap images using very deep convolutional neural networks. *Ecological Informatics*, **41**, 24–32. https://doi.org/10.1016/j.ecoinf.2017.07.004

Green, S.E., Rees, J.P., Stephens, P.A., Hill, R.A. & Giordano, A.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

Green, S.E., Stephens, P.A., Whittingham, M.J. & Hill, R.A. (2023) Camera trapping with photos and videos: implications for ecology and citizen science. *Remote Sensing in Ecology and Conservation*, **9**(2), 268–283. https://doi.org/10.1002/rse2.309

Hesley, D., Kaufman, M. & Lirman, D. (2023) Citizen science benefits coral reefs and community members alike. *Frontiers in Environmental Science*, **11**, 1–18. Available from: https://doi.org/10.3389/fenvs.2023.1250464

Hsing, P.-Y., Bradley, S., Kent, V.T., Hill, R.A., Smith, G.C., Whittingham, M.J. et al. (2018) Economical crowdsourcing for camera trap image classification. *Remote Sensing in Ecology and Conservation*, **4**(4), 361–374. https://doi.org/10.1002/rse2.84

Hsing, P.-Y., Coghill, L., Ryder, J., Austin, M., Dooley, S., Ellison, A. et al. (2020) Citizen scientists: school students conducting, contributing to and communicating ecological research – experiences of a school–university partnership. *School Science Review*, **101**(376), 67–74.

Hsing, P.-Y., Hill, R.A., Smith, G.C., Bradley, S., Green, S.E., Kent, V.T. et al. (2022) Large-scale mammal monitoring: the potential of a citizen science camera-trapping project in the United Kingdom. *Ecological Solutions and Evidence*, **3**(4), e12180. https://doi.org/10.1002/2688-8319.12180

Huebner, S.E., Palmer, M.S. & Packer, C. (2024) Human supervision is key to achieving accurate AI-assisted wildlife identifications in camera trap images. *Citizen Science: Theory and Practice*, **9**(1), 1–12. Available from: https://doi.org/10.5334/cstp.752

Hutson G. (2021) ConfusionTableR: Confusion matrix toolset. Manual https://CRAN.R-project.org/package=ConfusionTableR

Jackson, C.B., Crowston, K. & Østerlund, C. (2018) Did they login? Patterns of anonymous contributions in online communities. *Proceedings of the ACM on Human-Computer Interaction*, **2**(CSCW), 1–16. https://doi.org/10.1145/3274346

Jansen, M., Beukes, M., Weiland, C., Blumer, M., Rudolfi, M., Poerting, J. et al. (2024) Engaging citizen scientists in biodiversity monitoring: insights from the WildLIVE! Project. *Citizen Science: Theory and Practice*, **9**(1), 6. https://doi.org/10.5334/cstp.665

Karanth, K.U. (1995) Estimating tiger Panthera tigris populations from camera-trap data using capture—recapture models. *Biological Conservation*, **71**(3), 333–338. https://doi.org/10.1016/0006-3207(94)00057-W

Kelling, S., Lagoze, C., Wong, W.-.K., Yu, J., Damoulas, T., Gerbracht, J. et al. (2013) eBird: a human/computer learning network to improve biodiversity conservation and research. *AI Magazine*, **34**(1), 10. https://doi.org/10.1609/aimag.v34i1.2431

Littlewood, N.A., Hancock, M.H., Newey, S., Shackelford, G. & Toney, R. (2021) Use of a novel camera trapping approach to measure small mammal responses to peatland restoration. *European Journal of Wildlife Research*, **67**(1), 12. https://doi.org/10.1007/s10344-020-01449-z

Loos, A., Weigel, C. & Koehler, M. (2018) Towards automatic detection of animals in camera-trap images In: 2018 26th European signal processing conference (EUSIPCO), 2018 26th European Signal Processing Conference (EUSIPCO), Rome, Italy, IEEE, pp. 1805–1809. https://doi.org/10.23919/EUSIPCO.2018.8553439.

Mason, S.S., Hill, R.A., Whittingham, M.J., Cokill, J., Smith, G.C. & Stephens, P.A. (2022) Camera trap distance sampling for terrestrial mammal population monitoring: lessons learnt from a UK case study. *Remote Sensing in Ecology and Conservation*, **8**(5), 717–730. https://doi.org/10.1002/rse2.272

Masters, K., Oh, E.Y., Cox, J., Simmons, B., Lintott, C., Graham, G. et al. (2016) Science learning via participation in online citizen science. *Journal of Science Communication*, **15**(3), A07.

McCarthy, M.S., Després-Einspenner, M.L., Farine, D.R., Samuni, L., Angedakin, S., Arandjelovic, M. et al. (2019) Camera traps provide a robust alternative to direct observations for constructing social networks of wild chimpanzees. *Animal Behaviour*, **157**, 227–238. https://doi.org/10.1016/j.anbehav.2019.08.008

McCleery, R.A., Zweig, C.L., Desa, M.A., Hunt, R., Kitchens, W.M. & Percival, H.F. (2014) A novel method for camera-trapping small mammals. *Wildlife Society Bulletin*, **38**(4), 887–891. https://doi.org/10.1002/wsb.447

McClure, E.C., Sievers, M., Brown, C.J., Buelow, C.A., Ditria, E.M., Hayes, M.A. et al. (2020) Artificial intelligence meets citizen science to supercharge ecological monitoring. *Patterns*, **1**(7), 100109. https://doi.org/10.1016/j.patter.2020.100109

McShea, W.J., Forrester, T., Costello, R., He, Z. & Kays, R. (2016) Volunteer-run cameras as distributed sensors for macrosystem mammal research. *Landscape Ecology*, **31**(1), 55–66. https://doi.org/10.1007/s10980-015-0262-9

Norman, D.L., Bischoff, P.H., Wearn, O.R., Ewers, R.M., Rowcliffe, J.M., Evans, B. et al. (2023) Can CNN-based species classification generalise across variation in habitat within a camera trap survey? *Methods in Ecology and Evolution*, **14**(1), 242–251. https://doi.org/10.1111/2041-210X.14031

Norouzzadeh, M.S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M.S., Packer, C. et al. (2018) Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences of the United States of America*, **115**(25), E5716–E5725. https://doi.org/10.1073/pnas.1719367115

Pankiv, K. & Kloetzer, L. (2024) Does using artificial intelligence in citizen science support volunteers’ learning? An experimental study in ornithology. *Citizen Science: Theory and Practice*, **9**(1), 36. https://doi.org/10.5334/cstp.733

R Core Team. (2024) *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ridout, M.S. & Linkie, M. (2009) Estimating overlap of daily activity patterns from camera trap data. *Journal of Agricultural Biological and Environmental Statistics*, **14**(3), 322–337. https://doi.org/10.1198/jabes.2009.08038

Rigoudy, N., Dussert, G., Benyoub, A., Besnard, A., Birck, C., Boyer, J. et al. (2023) The DeepFaune initiative: a collaborative effort towards the automatic identification of European fauna in camera trap images. *European Journal of Wildlife Research*, **69**(6), 113. https://doi.org/10.1007/s10344-023-01742-7

Rivera, K., Fidino, M., Lehrer, E.W., Torsey, H.R., Allen, S., Trouille, L. et al. (2024) Optimizing community science contributions in ecology: a case study on Zooniverse's "Chicago wildlife watch". *Biological Conservation*, **292**, 110490. https://doi.org/10.1016/j.biocon.2024.110490

Rowcliffe, J.M., Field, J., Turvey, S.T. & Carbone, C. (2008) Estimating animal density using camera traps without the need for individual recognition. *Journal of Applied Ecology*, **45**(4), 1228–1236.

Schneider, D., Lindner, K., Vogelbacher, M., Bellafkir, H., Farwig, N. & Freisleben, B. (2024) Recognition of European mammals and birds in camera trap images using deep neural networks. *IET Computer Vision*, **18**(8), 1–31. https://doi.org/10.1049/cvi2.12294

Schuttler, S.G., Sorensen, A.E., Jordan, R.C., Cooper, C. & Shwartz, A. (2018) Bridging the nature gap: can citizen science reverse the extinction of experience? *Frontiers in Ecology and the Environment*, **16**(7), 405–411. https://doi.org/10.1002/fee.1826

Sharma, N., Colucci-Gray, L., Lakeman-Fraser, P., Robinson, A., Newman, J., van der Wal, R. et al. (2024) Image recognition as a "dialogic AI partner" within biodiversity citizen science—an empirical investigation. *Citizen Science: Theory and Practice*, **9**(1), 735. https://doi.org/10.5334/cstp.735

Shepley, A., Falzon, G., Meek, P. & Kwan, P. (2021) Automated location invariant animal detection in camera trap images using publicly available data sources. *Ecology and Evolution*, **11**(9), 4494–4506. https://doi.org/10.1002/ece3.7344

Swanson, A., Kosmala, M., Lintott, C. & Packer, C. (2016) A generalized approach for producing, quantifying, and validating citizen science data from wildlife images. *Conservation Biology*, **30**(3), 520–531. https://doi.org/10.1111/cobi.12695

Swanson, A., Kosmala, M., Lintott, C., Simpson, R., Smith, A. & Packer, C. (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

Tabak, M.A., Norouzzadeh, M.S., Wolfson, D.W., Newton, E.J., Boughton, R.K., Ivan, J.S. et al. (2020) Improving the accessibility and transferability of machine learning algorithms for identification of animals in camera trap images: MLWIC2. *Ecology and Evolution*, **10**(19), 10374–10383. https://doi.org/10.1002/ece3.6692

Tabak, M.A., Norouzzadeh, M.S., Wolfson, D.W., Sweeney, S.J., Vercauteren, K.C., Snow, N.P. et al. (2019) Machine learning to classify animal species in camera trap images: applications in ecology. *Methods in Ecology and Evolution*, **10**(4), 585–590. https://doi.org/10.1111/2041-210X.13120

Tøn, A., Ahmed, A., Imran, A.S., Ullah, M. & Azad, R.M.A. (2024) Metadata augmented deep neural networks for wild animal classification. *Ecological Informatics*, **83**, 102805. https://doi.org/10.1016/j.ecoinf.2024.102805

Torney, C.J., Lloyd-Jones, D.J., Chevallier, M., Moyer, D.C., Maliti, H.T., Mwita, M. et al. (2019) A comparison of deep learning and citizen science techniques for counting wildlife in aerial survey images. *Methods in Ecology and Evolution*, **10**(6), 779–787. https://doi.org/10.1111/2041-210X.13165

Trouille, L., Lintott, C.J. & Fortson, L.F. (2019) Citizen science frontiers: efficiency, engagement, and serendipitous discovery with human–machine systems. *Proceedings of the National Academy of Sciences of the United States of America*, **116**(6), 1902–1909. https://doi.org/10.1073/pnas.1807190116

Vazquez, C., Rowcliffe, J.M., Spoelstra, K. & Jansen, P.A. (2019) Comparing diel activity patterns of wildlife across latitudes and seasons: time transformations using day length. *Methods in Ecology and Evolution*, **10**(12), 2057–2066. https://doi.org/10.1111/2041-210X.13290

Vélez, J., McShea, W., Shamon, H., Castiblanco-Camacho, P.J., Tabak, M.A., Chalmers, C. et al. (2023) An evaluation of platforms for processing camera-trap data using artificial intelligence. *Methods in Ecology and Evolution*, **14**(2), 459–477. https://doi.org/10.1111/2041-210X.14044

Whytock, R.C., Świeżewski, J., Zwerts, J.A., Bara-Słupski, T., Koumba Pambo, A.F., Rogala, M. et al. (2021) Robust ecological analysis of camera trap data labelled by a machine learning model. *Methods in Ecology and Evolution*, **12**(6), 1080–1092. https://doi.org/10.1111/2041-210X.13576

Willi, M., Pitman, R.T., Cardoso, A.W., Locke, C., Swanson, A., Boyer, A. et al. (2019) Identifying animal species in camera trap images using deep learning and citizen science. *Methods in Ecology and Evolution*, **10**(1), 80–91. https://doi.org/10.1111/2041-210X.13099

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Data S1.