Heteromation in Citizen Science: The Division of Labor Between Citizens, Experts, and Machines

Alena Seredko[1], Anna Jia Gander[2], Marisa Ponti[2]
[1]Department of Education, Communication and. Learning
[2]Department of Applied Information Technology
University of Gothenburg, Sweden
Corresponding author: marisa.ponti@ait.gu.se

Abstract.

Citizen science is a promising field for the creation of human-machine systems with increasing computational abilities, as several projects generate large datasets that can be used as training materials for machine learning models. This paper aims to identify the forms of human-machine learning integration in citizen science projects. The fifty articles examined in this systematic review report on projects combining human and machine efforts for analyzing, coding, classifying, and clustering data provided, for example, by cameras and telescope images. Machine learning is used at various stages of the data life cycle, through algorithms that perform tasks like classification, regression, clustering, and association. The findings highlight the character of the projects as heteromated systems, wherein human participation remains crucial and volunteer and ML efforts are often positioned as complementary rather than mutually exclusive. While leveraging the complementarity of strengths is one of the main arguments to combine humans and machines and enhance their respective capabilities, essentializing the attributes of humans and machines should be avoided. Treating these attributes as stable and natural does not take into account that cognitive work will be shifting between humans and machines, as the list of research tasks that machines can do is growing, although algorithms are still second to humans on recognizing patterns and they have longer learning curves. The findings can help researchers and practitioners to better understand human-machine integration in citizen science and point to unexplored areas. The emerging academic field of collective intelligence, increasingly interested in combining human intelligence with AI, can also find the review relevant.

Keywords: citizen science; collective intelligence; function allocation; heteromation; human-machine integration; machine learning.

Introduction

Artificial intelligence (AI) has been used in citizen science (CS) for about 20 years (Ceccaroni et al. 2019). A subfield of AI, machine learning (ML), in particular, supervised learning, has been able to learn input/output relationships from data for years and thus solve problems such as classification or regression tasks with high accuracy. The creation of more complex human-machine systems with increasing computational abilities opens up new ways of collaboration between humans and machines. Citizen science is a promising arena for the design and implementation of these systems (Lintott & Reed, 2013), as several projects have generated large datasets that can be used as training materials for ML models. Human-machine integration can help process these massive amounts of data more efficiently and accurately, and monitor the results. As ML enters the citizen science scene, which activities performed by citizens are potentially more automatable? Or, will humans continue to work alongside machines in heteromated projects, meaning that they combine capacities of humans and machines (Ekbia and

Nardi 2017). With this in mind, this paper aims to identify articles that report on forms of human-machine learning integration in citizen science projects. In this section, we briefly introduce the notion of human-machine integration and we present the rationale and research questions of this paper. In the Literature Search Process section, we elaborate on the methodology we used for collecting and assessing the papers under review. In the Results and Discussion sections, we present, visualize and reflect on the review findings. The last section presents conclusions from this study and points to future directions.

Conceptualizations of human-machine integration

Our review focuses on human-machine learning integration. The choice of what term to use has been an issue, as we are aware that terms should be chosen carefully and their use explained because no single term is appropriate for all contexts. In the articles reviewed, we did not encounter closely related terms such as human-machine cooperation or collaboration but only the term 'human-machine integration' (Trouille, Lintott, and Fortson 2019). In Kelling et al. (2013), we found the term human-computer learning network in connection with the eBird project. Zevin et al. (2017) discuss the development of a socio-computational system leveraging the power of human and machine computation in the Gravity Spy project. While several other articles do not employ explicit terms, they use specific vocabulary pointing at the joint efforts of humans and machines. The most commonly used words include *combine* and its derivatives (e.g., Wright et al. 2017; Coughlin et al. 2019; Keshavan et al. 2019), *synergy* (Beaumont et al. 2014), and *symbiotic relationship* (Wright et al. 2019; Crowston et al. 2020). A summary table with the terms found in the literature is in Supplementary File 1. In this paper, we use the term *human-machine transformation* because, in our reading, it broadly refers to humans working alongside machines.

Rationale

A systematic review of human-machine learning integration in citizen science is timely and relevant for at least two reasons. First, a comprehensive background is important for taking stock of current knowledge and highlighting the significance of human-machine integration in this field. Second, this review can help researchers and practitioners to identify gaps in the knowledge of this topic, thus helping them to identify questions or formulate hypotheses for future research. Furthermore, a better understanding of human-machine integration in citizen science could be also relevant for the emerging academic field of collective intelligence, which is increasingly interested in combining human intelligence with AI (Mulgan 2018).

Aim of the paper

The purpose of this paper is, through a systematic review of papers in citizen science, to identify projects using machine learning, and critically analyze forms of human-machine integration. This review can benefit diverse stakeholders, including citizen science designers, researchers, and practitioners, by providing insights into the distribution of tasks between humans and algorithms to achieve projects' stated goals.

This systematic review aims to produce an in-depth account of reported forms of human-machine integration in citizen science by collecting evidence about the tasks performed and the role played by citizen scientists, experts (e.g., professional researchers), and algorithms. We focused on tasks because function allocation, that is, deciding which tasks or jobs should be allocated to humans and which ones to machines is a crucial step in the design of human-machine systems (de Winter & Dodou 2014). Therefore, the following research question (RQ) was addressed:

- RQ 1: what tasks citizen scientists, experts, and algorithms performed to achieve the goals of citizen science projects, and why?

Literature Search Process

This section outlines the search procedure and sampling of relevant articles across three databases: SCOPUS, Web of Science, and the ACM Digital Library. These databases are well-established, multi-disciplinary research platforms, including a wide variety of peer-reviewed journals, and they are being updated regularly. These three databases were used they constitute a baseline for search of published peer-reviewed articles. However, we are aware that, in the case of citizen science publications, the true extent of these publications can be larger as studies can be published in non-peer-reviewed literature sources and would not be referred in these three databases. We did not include preprints.

We reviewed the existing literature for articles reporting on examples of human-machine integration in citizen science projects. Two search procedures were conducted. First, we searched for articles containing "citizen science" and "artificial intelligence", or "machine learning" in the title, abstract, and keyword sections. However, after a brief initial scanning of the resulting articles, we added several other search terms to include the most widely used machine learning paradigms and methods, such as supervised learning, unsupervised learning, reinforcement learning, reinforcement algorithm, deep learning, neural network(s), and transfer learning. The search was limited to articles written in English and published until July 2020. A total of 170 results were collected across the three databases, 100 of which were unique. The search terms used and the number of results per database are presented in Table 1.

Table 1. Search procedure 1. Databases, search strings. and result count.

Database	Search strings	Results
Web of science	TOPIC: ("citizen science" OR "citizen scientist*") AND TOPIC: ("artificial intelligence" OR "machine learning" OR "supervised learning" OR "unsupervised learning" OR "reinforcement learning" OR "reinforcement algorithm" OR "deep learning" OR "neural network*" OR "transfer learning") Refined by: LANGUAGES: (ENGLISH) AND DOCUMENT TYPES: (ARTICLE) Timespan: All years. Indexes: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI.	83
SCOPUS	(TITLE-ABS-KEY (*"citizen science"* OR *"citizen scientist*"*) AND TITLE-ABS-KEY (*"artificial intelligence"* OR *"machine learning"* OR *"supervised learning"* OR *"unsupervised learning"* OR *"reinforcement learning"* OR *"reinforcement algorithm"* OR *"deep learning"* OR *"neural network*"* OR *"transfer learning"*)) AND (LIMIT-TO (DOCTYPE , *"ar"*)) AND (LIMIT-TO (LANGUAGE , *"English"*))	86

Database	Search strings	Results
ACM	[[Abstract: "citizen science"] OR [Abstract: "citizen scientist"] OR [Abstract: "citizen scientists"]] AND [[Abstract: "artificial intelligence"] OR [Abstract: "machine learning"] OR [Abstract: "supervised learning"] OR [Abstract: "unsupervised learning"] OR [Abstract: "reinforcement learning"] OR [Abstract: "reinforcement algorithm"] OR [Abstract: "deep learning"] OR [Abstract: "neural network"] OR [Abstract: "neural networks"] OR [Abstract: "transfer learning"]] Applied filters: Research article, Journals	1
Total		**170**

The initial examination of the resulting articles revealed that the chosen search strategy did not fully cover the articles focused on citizen science games, such as EteRNA or Eyewire. This could be because some authors choose to use the game title in the abstract rather than refer to it as a “citizen science game”. To overcome this limitation, we used a list of citizen science games (Baert 2019) as a source of search terms for the second search procedure. Table 2 contains the employed search strings and the number of generated results. While we tested searches with all of the game titles, only those that generated at least one result were included. The search using the ACM Digital Library did not produce any results, thus it was excluded from the table. The search generated 17 results, nine of which were unique. Out of the remaining nine articles, five were not covered by the first search procedure.

Table 2. Search procedure 2 (citizen science games). Databases, search strings, and result count.

Database	**Search strings**	**Results**
Web of science	TOPIC: (Foldit OR Eyewire OR “Project Discovery” OR “NeMO-net”) AND TOPIC: ("artificial intelligence" OR "machine learning" OR "supervised learning" OR "unsupervised learning" OR "reinforcement learning" OR "reinforcement algorithm" OR "deep learning" OR "neural network*" OR "transfer learning") Refined by: LANGUAGES: (ENGLISH) AND DOCUMENT TYPES: (ARTICLE) Timespan: All years. Indexes: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI.	10
SCOPUS	(TITLE-ABS-KEY (*Foldit* OR *Eyewire* OR *“Project Discovery”* OR *“NeMO-net”*) AND TITLE-ABS-KEY (*"artificial intelligence"* OR *"machine learning"* OR *"supervised learning"* OR *"unsupervised learning"* OR *"reinforcement learning"* OR *"reinforcement algorithm"* OR *"deep learning"* OR *"neural network*"* OR *"transfer*	7

	learning")) AND (LIMIT-TO (DOCTYPE , *"ar"*)) AND (LIMIT-TO (LANGUAGE , *"English"*))	
Total		**17**

Table 3. Exclusion criteria and number of excluded articles.

Exclusion criteria	**Count**
Not related to artificial intelligence	9
Not related to citizen science	18
Citizen science data and machine learning procedures are independent of each other	12
Citizen science data/procedure are simulated	3
Not empirical studies	7
Irrelevant topics	6
Total	**55**

Since we aimed to focus on empirical evidence from the projects combining citizen scientists and machine learning efforts, several selection criteria were applied to filter out irrelevant and theoretical articles. Papers were assessed based on their titles, abstracts, keywords, and, if necessary, full-text reading. We excluded non-empirical papers as well as articles that do not apply or develop ML methods or do not report citizen science projects. We also filtered out those articles that compare the results produced by citizen scientists with the outcomes of ML algorithms trained on expert-produced data. Table 3 presents the exclusion criteria and the number of excluded articles. As a result of the selection process, 50 articles were selected for the review (The list of the 50 articles is in Supplementary file 2).

Results

The results section is divided into three parts. First, we provide a descriptive overview of the dataset, including some basic characteristics of the reviewed publications and the fields of citizen science projects addressed in them. Second, we present the analysis of the 50 reviewed papers in terms of tasks performed by humans, including citizen scientists and experts, and machine learning algorithms. Third, we present two major approaches to integrating the described tasks into a single workflow within the reviewed projects.

Overview of the dataset

First, we present a descriptive overview of the dataset. The reviewed articles were published between 2011 and 2020, with substantial growth in the number observed starting from 2018: 35 out of the 50 articles were published between 2018 and 2020 (Supplementary file 2 contains the list of the 50 articles). The increasing interest in combining machine learning (ML) and citizen science (CS) is also evident from the growing diversity of research fields with which the described CS projects are associated. The review demonstrated a considerable variety of citizen science projects (n=42) with some papers reporting on using data from several projects. The three main areas that attract the most attention across the whole timespan are astronomy and astrophysics (e.g., Galaxy Zoo, Gravity Spy, and Supernova Hunters), biology (EteRNA, EyeWire, and Project Discovery), and ecology and biodiversity (e.g., eBird, Bat Detective). However, starting from 2017, we observe a larger variation including archeology (e.g., Heritage Quest, and field expeditions), neuroimaging (Braindr.us), seismology (MyShake app), and environmental issues (recruiting volunteers to measure the quality of air or water).

Tasks performed by citizen scientists

The two main categories of tasks performed by citizen scientists are collecting data in the field and classifying observations. Other tasks include generating new taxonomies, validating the algorithm classification results, solving in-game puzzles, and going through training. Supplementary file 3 presents a summary table of the references regarding these tasks. Here, we describe these tasks in more detail.

Collecting data in the field is a set of tasks widely assigned to citizen scientists in the areas of ecology, biodiversity, and environmental monitoring. Delegating the collection of data to volunteers allows researchers to map geographical distributions and spatial variation in unprecedented scope and details, which is especially relevant when monitoring by researchers is not feasible or efficient enough. The most common types of data contributed by volunteers include photos of plants or animals, accompanied by some context information (such as location and date/time of observation), and sometimes by a description (e.g., Derville et al. 2018; Capinha 2019). Less common types are videos and audio recordings (e.g., Zilli et al. 2014; Hardison et al. 2019). These observations were often accompanied by species classification, as citizen scientists were asked to submit observations of a particular species (e.g., Jackson, Gergel, and Martin 2015). Alternatively, volunteers submitted observations that they classified with the help of an instrument, such as using mobile app suggestions, without attaching a photo, as in the eBird app (Curry et al. 2018). Several articles also reported on citizens sending a specimen to researchers, e.g., bee trap nests (Kerkow et al. 2020; Everaars et al. 2011). Another type is a relatively passive data collection that does not require analysis on the part of citizens. Lim et al. (2019) and Adams et al. (2020) reported on projects aimed at sampling air quality: volunteers were equipped with AirBeam sensors and asked to sample several routes by walking or cycling there. In Winter et al. (2019), an Android app was presented that allowed for identifying and classifying charged particles in camera image sensors. The only task outsourced to the citizens, in this case, was installing the app.

The second popular set of tasks performed by citizen scientists is related to image analysis and includes classifying images into predefined categories, describing objects by choosing all relevant categories from a predefined list, as well as identifying and counting objects. The research fields setting up citizen science projects to outsource these tasks to volunteers include astronomy and astrophysics, ecology and biodiversity, archeology, biology, and neuroimaging. The tasks are performed in web interfaces: the majority of the projects run on the Zooniverse platform, but there are also separate initiatives such as the Braindr.us website (Keshavan, Yeatman, and Rokem 2019) and the Project Discovery implemented in the Eve online game (Sullivan et al. 2018). Allocating classification tasks to citizen scientists is often related to the extremely large size of currently available datasets that makes expert classification unfeasible. The projects leverage human ability for pattern recognition and benefit from the scope of citizen science projects. The resulting classifications constitute training datasets for ML analysis. Citizen scientists classified objects from images into predefined categories. It can be a binary classification task, e.g., citizens decided whether a supernova candidate is a real or a 'bogus' detection (Wright et al. 2017; Wright et al. 2019). Alternatively, there could be a larger number of categories. For example, four studies reported on the Gravity Spy project, where users were presented with spectrograms and asked to classify glitches into predefined categories according to their morphology (Bahaadini et al. 2018; Crowston et al. 2020; Jackson et al. 2020; Zevin et al. 2017). Another task performed by citizen scientists was about describing an object in an image using a set of predefined characteristics. Examples include describing circumstellar debris disk candidates (Nguyen et al. 2018); classification of protein localization patterns in microscopy images (Sullivan et al. 2018); and morphological classification of galaxies (Jiménez et al. 2020; Kuminski et al. 2014; Shamir, Diamond, and Wallin 2016). Last, the projects benefiting from citizen scientists identifying and counting objects asked citizens to identify and locate animals of particular species (Bowley et al. 2019; Torney et al. 2019); mark potential archeological sites (Lambers et al. 2019), and identify and locate Moon craters (Tar et al. 2017), and interstellar bubbles on images (Beaumont et al. 2014; Duo and Offner 2017).

Another task outsourced to citizen scientists is related to generating new taxonomies of objects. Coughlin et al. (2019) discussed that Gravity Spy project volunteers did not only classify spectrograms into already known classes of glitches but can also suggested new classes, being aided by the initial ML clustering of morphologically similar objects. Citizen scientists also performed validation of algorithm classification or object detection results. Participants of the Leafsnap project submitted photos of leaves, and if the shape-matching algorithm did not classify the plant with high enough probability, citizens were offered several options to choose from (Kress et al., 2018). A larger-scale validation procedure was reported by Lambers et al. (2019). The reported algorithm detected potential archeological objects in the remotely sensed data. Then, citizen scientists together with heritage managers and/or academic researchers participated in field expeditions to validate the ML results.

A distinct type of citizen science projects are games wherein citizens were asked to solve in-game puzzles. In a number of these games, the tasks performed by volunteers differ considerably when compared to other projects. Kim et al. (2014) reported on the EyeWire game project, where players contributed to mapping 3D structures of retinal neurons by coloring the area that belongs to one neuron and avoiding coloring other neurons on a 2D slice image. Koodli et al. (2019) and Lee et al. (2014) discussed the EteRNA project, where players solved two-dimensional puzzles to design sequences that can fold into a target RNA structure.

A standard task required of citizen scientists is going through training, which was sometimes done face-to-face if citizens were asked to collect specific types of data in the field (Hardison et al. 2019). In other cases, this training occurred mainly online, as citizens went through guidelines prepared by project authors (Keshavan et al. 2019). The training could also be guided, facilitated, and assessed using ML algorithms (Zevin et al. 2017). While we may suggest that all citizen scientists go through some kind of training, not all of the reviewed articles included related information.

Tasks performed by experts and researchers

Tasks performed by researchers and field experts are the most varied. They include collecting and processing the original data before it is presented to volunteers or algorithms, creating the gold-standard datasets, processing and curating the data collected or classified by citizen scientists, and preparing the training datasets for ML. Several tasks are related to recruiting, training, and supporting volunteers. Finally, researchers are involved in the evaluation and validation of results. It is important to note that some tasks performed by researchers may not be discussed in papers in detail, since they occur naturally in every project, or because they may not be relevant for the discussion. Therefore, this section outlines only those tasks that are discussed in sufficient detail.

Several studies on biodiversity reported on researchers collecting observation data of species occurrence in the field (Derville et al. 2018; Jackson et al. 2015; Zilli et al. 2014). Researchers also obtained pre-classified data from external sources, such as the records of ladybirds sourced from the UK Biological Records Centre (Terry et al. 2020). These observations together with observational data collected by citizen scientists were further used to train and test ML algorithms. In cases when ML methods were used to predict species distribution or environmental conditions (e.g., coral bleaching), researchers were also responsible for sourcing data related to the characteristics of the environment. Examples of such data were mean temperature and precipitation (Capinha 2019; Jackson et al. 2015), and geospatial data including roads and types of land usage (Lim et al. 2019). Additionally, original data obtained from cameras or sensors were preprocessed by researchers to be further presented to citizen scientists. For example, the audio recordingsfrom bat observations were split into short sound clips and converted to spectrograms for the Bat Detective project (Mac Aodha et al. 2018), while in the Serengeti Wildebeest Count project, images from trap cameras were filtered to remove the empty ones and thereby reduce the number of images for citizen scientists to classify (Torney et al. 2019).

One of the tasks traditionally performed by researchers and field experts is creating the gold-standard dataset. However, the application of these datasets varied. Expert classifications were used to perform the initial training of the algorithm (Crowston et al. 2020; Jackson et al. 2020), to calibrate and fine-tune the machine learning performance (Beaumont et al. 2014; Jiménez et al. 2020), or to provide testing set for ML classification methods (Crowston et al. 2020; Tar et al. 2017). Expert-labeled data was also included in the guidelines for volunteers (Keshavan et al. 2019), and used to assess the accuracy of citizen scientists classifications and give feedback to volunteers (Jackson et al. 2020; Zevin et al. 2017), as well as to create the weighting of each citizen scientist's vote in the final label based on how much their labels corresponded to the gold-standard set (Keshavan et al. 2019).

Researchers involved in the development of citizen science projects were responsible for recruiting, training, and supporting volunteers. In those projects where volunteers were asked to collect data in a specific location (e.g., air quality measurements along certain routes, or coral bleaching measures on specific beaches), researchers recruited volunteers and performed face-to-face training (Adams et al. 2020; Hardison et al. 2019; Kumagai et al. 2017). When citizen participation was not bound to a particular space, volunteers received written guidelines (Bowley et al. 2019; Torney et al. 2019; Wright et al. 2017). Supporting user motivation and engagement was another task performed by researchers. Examples include ensuring that volunteers were involved in real classification tasks that led to the advancement of the project (Crowston et al. 2020). In projects that required volunteers to collect observations in the field, researchers followed up on citizens' contributions (Jackson et al. 2015; Kerkow et al. 2020), and provided online support and feedback (Lambers et al. 2019). Kim et al. (2014) also reported on integrating leaderboard and online chats in the EyeWire project to motivate players to compete and communicate with each other.

The data provided by citizen scientists, be it observations or classifications, was further processed and curated by researchers. The observations provided by citizen scientists, such as cicada call recordings or ladybird recordings, were classified by field experts to be further used by an ML algorithm (Terry et al. 2020; Zilli et al. 2014). Other related tasks included processing citizen scientist contributions (e.g., returned bee nests) for future analysis (Everaars et al. 2011; Kerkow et al. 2020); deciding on the number of volunteer votes required before the final classification label for an image was generated and used to train or test ML algorithm (Lambers et al. 2019; Sullivan et al. 2018; Wright et al. 2019; Wright et al. 2017); and choosing a limited amount of volunteer-produced data for training an algorithm (Koodli et al. 2019). Preparing the training dataset for machine learning also included such tasks as generating pseudo-absences when the information provided by volunteers only indicates presences observed (Jackson et al. 2015); generating synthetic observations of bubbles in dust emission to improve ML classification (Dou and Offner 2017); or augmenting the training dataset by transforming existing images to increase the accuracy of ML classification (Dou and Offner 2017).

Researchers were also involved in the evaluation of results, while the actual tasks depended on the goals of the studies. Researchers evaluated the predictive accuracy of species distribution models using other occurrence datasets (Botella et al. 2018; Kerkow et al. 2020). Bowley et al. (2019) reported on comparing the results of ML training using citizen science data and using expert classifications. A low error level demonstrated the viability of engaging citizen scientists to produce training data for ML. Several articles reported on comparing the performance of different ML and statistical models to predict species distribution (Curry et al. 2018; Jackson et al. 2015). Lee et al. (2014) discussed that EteRNA volunteer players outperformed previous algorithms in discovering RNA design rules. Furthermore, the results of ML classifications were compared with manual classifications done by field experts (Nguyen et al. 2018; Pearse et al. 2018; Wright et al. 2017). A similar approach was reported on by Kress et al. (2018) in relation to the Leafsnap app. However, since Leafsnap participants needed to confirm the classification suggestions of an algorithm, the validation of accuracy referred to the results from both citizen scientists and ML. Unique were the validation procedures reported by Lambers et al. (2019), as experts together with citizen scientists validated the new potential archaeological objects identified using ML by going into the field.

Tasks performed by machine learning algorithms

Tasks assigned to algorithms can be subdivided into several categories, such as classification and object detection; clustering; improving the performance by learning from citizen science feedback; modeling species distribution, predicting the environmental conditions (e.g., air quality, or variations in the data); addressing biases in the original data or in CS classification and detection results; guiding citizen science training; and finally learning from player moves in a citizen science game.

Classification and object detection are the most popular tasks performed by ML algorithms in a variety of projects in the fields of ecology and biodiversity, astronomy and astrophysics, etc. Examples include an algorithm trained on citizen science and expert labels and used to classify galaxy images (Jiménez et al. 2020), or an algorithm trained on Serengeti Wildebeest Count project data used for counting wildlife in aerial survey images (Torney et al. 2019). It is argued that with the limited number of citizen scientists and increasingly large databases of images, ML offers an approach to scale up data processing, overcome the analysis 'bottleneck' problem, and also relieve some burden from researchers and citizen scientists who would only have to classify enough images for ML training and not the whole dataset (Torney et al. 2019; Wright et al. 2017).

Clustering is another task performed by ML (Coughlin et al. 2019; Wright et al. 2019). Coughlin et al. (2019) reported on the algorithm facilitating the discovery of new glitches by citizen scientists in the Gravity Spy project. Due to the sheer volume of available images, it is extremely difficult for volunteers to identify new classes by finding a sufficient number of similar objects that do not belong to any of the known classes. Thus, an ML algorithm clustered similar images together and offered this set to volunteers to make their judgment. Wright et al. (2019) reported on using ML clustering to produce an initial grouping of similar images. Grouped images were shown to citizen scientists in the Supernova Hunters project, who had to mark all of the objects belonging to one glitch class. Then, citizen scientists' labels were fed back to the algorithm to make the clustering purer, thus the algorithm learned from citizen science feedback. Compared to the standard image-by-image presentation, this approach allowed to considerably reduce the amount of volunteer effort operationalized as the number of clicks required to classify a dataset. Lambers et al. (2019) also reported on improving the results of the algorithm using contributions from citizen scientists. Volunteers participated in field expeditions to validate archeological objects detected by ML, and the results were used to tune the algorithm object detection results.

The articles related to the EteRNA game also discussed the development of ML algorithms that learn from player moves. Lee et al. (2014) developed the EteRNABot algorithm trained using design rules uncovered by players and used to predict the design of RNA structures. Koodli et al. (2019) described the development of EternaBrain-SAP algorithm trained on the compilation of expert player moves.

ML was also employed in biodiversity and ecology-related projects to model species distribution, e.g., to predict the probability of White-taled Ptarmigan occurrence over Vancouver Island (Jackson et al. 2015), or to predict the distribution area of the Asian bush mosquito (Kerkow et al. 2020). The datasets used for training were usually combined from different sources: observations collected and reported by citizen scientists and sometimes by experts as well, and environmental or climate data extracted by the researchers. Another ML task was to predict water quality (Thornhill et al. 2017), air quality (Lim et al. 2019), or coral bleaching (Kumagai et al. 2017), using data collected by citizen scientists in the field and environmental or urban data collected by scientists.

ML methods were used to mitigate biases or errors in the data collected and/or classified by citizen scientists. Adams et al. (2020) reported on developing an ML model to adjust the AirBeam sensor observations that demonstrated errors in high humidity period.

Another issue addressed by using ML is that citizen scientists' participation in collecting data is not uniformly distributed in space and is naturally biased towards recording observations rather than absences. For instance, Derville et al. (2018) compared several algorithms modeling species distribution based on how they account for the sampling bias present in the nonsystemic citizen science observations of humpback whales. Several articles also reported on employing transfer learning techniques, if only a small training dataset was available in a particular project. Willi et al. (2019) trained a model on the data from the larger Snapshot Serengeti citizen science project, and then applied transfer learning approach to achieve higher accuracy when only smaller datasets were available as in the Camera CATalogue project.

One of the arguments for applying ML is that volunteers are different in terms of expertise levels and are prone to human error, thus they may misclassify data. This issue is addressed in several ways. Tar et al. (2017) used predictive error modeling to evaluate the false positive contamination in the Moon Zoo project classifications. Keshavan et al. (2019) fitted the model with citizen science ratings weighted by how much their classifications matched the golden standard created by experts. Shamir et al. (2016) suggested using a pattern recognition algorithm to evaluate the consistency of annotations made by individual volunteers in the Galaxy Zoo project.

Kelling et al. (2013) used a probabilistic machine learning approach to measure the expertise of eBird volunteer observers. They employed the occupancy-detection experience model to measure the probability of a particular species being identified in a particular site, and to distinguish expert observers from novice observers who are more likely to misclassify common bird species. This approach allowed researchers to provide volunteers with feedback on their observation accuracy, and also to improve the quality of a training dataset for an ML algorithm. Gravity Spy is one of the few projects (in addition to eBird mentioned above), wherein the information about the quality of citizen contribution was used to give feedback and training to volunteers (Crowston et al. 2020; Jackson et al. 2020; Zevin et al. 2017). ML system scaffolded volunteers by guiding them through several training levels: at first showing glitches belonging to two classes with a high level of ML-determined probability and then - as a volunteer learned to classify them - increasing the number of possible classes and offering images with lower ML confidence score. Crowston et al. (2020) also described the development of a model to trace volunteer learning based on the agreement of CS classifications with ML classifications.

Integrating tasks into one workflow

The human and machine tasks described in the previous sections are integrated into a single workflow in different ways. The most common type of integration of human and machine efforts is when citizen scientists make some sort of contribution, which is then used to train and test one or several machine learning approaches. It could include citizens collecting the data and ML algorithms classifying the data or modeling species distribution. Alternatively, citizens can classify data, and ML trained on these classifications is used to classify a larger dataset. The roles of researchers and experts here are related to the orchestration of the whole process. They are responsible for recruiting and training volunteers, collecting more data, creating the gold-standard dataset of classifications, as well as testing and evaluating the overall performance of the systems. When these studies aim to address the biases and errors in citizen scientists' observations and classifications, the methods include employing specific ML approaches, generating synthetic training data for an algorithm, or weighing citizen scientist labels based on their levels of performance. When existing datasets are employed, sometimes from finished CS projects as well, volunteers' input is included only in the first stage of the process, when they provide or process the original data.

Although the goal of several ML approaches is to reduce the required volunteer efforts (Willi et al. 2019), or overcome the need to rely on citizen scientists' contributions (Sullivan et al. 2018), volunteer and ML efforts are often positioned as complementary. The discussions of leveraging the complementary strengths of humans and machines are particularly relevant for studies that have classification or object count as a shared task (Keshavan et al. 2019; Lukic et al. 2018; Torney et al. 2019). It is argued that volunteers are better at harder pattern recognition tasks (Kuminski et al. 2014; Lukic et al. 2018), need less training examples to make judgments about new images (Wright et al. 2017), and are capable of serendipitous discoveries (Beaumont et al. 2014). Meanwhile, algorithms allow to process extremely large datasets (Kuminski et al. 2014; Wright et al. 2017) and are capable of reproducible analysis (Beaumont et al. 2014). Thus, it is generally argued that synergy is the most productive strategy (Beaumont et al. 2014) and that it is not yet possible to replace human input (Lukic et al. 2018).

However, several articles report on keeping citizen scientists involved in the process beyond the stage of initial contributions. This is another type of integrating human and machine efforts into one workflow. The articles can be grouped into three subsets of studies on the ground of tasks performed by different agents: (1) algorithm used to train and provide feedback to volunteers; (2) algorithm used to help volunteers perform their tasks efficiently; (3) volunteers validating ML results.

The first subset includes articles on eBird and Gravity Spy projects. Kelling et al. (2013) described the human-machine combination as a human/computer learning network that benefited from ML speed and scalability and volunteers' identification of bird species in the eBird project. The study described a feedback loop, as ML was used to evaluate the accuracy of citizens' observations and provide feedback to volunteers. This, in turn, was argued to increase their level of expertise and improve the quality of the training dataset. Moreover, the algorithm was also used to address the sampling bias by suggesting volunteers where to sample next. In Gravity Spy, the combination of human and machine efforts to classify glitches is related to algorithms guiding the learning of citizen scientists (Crowston et al. 2020; Jackson et al. 2020; Zevin et al. 2017). Volunteers were progressing from the beginner to the advanced level, as the algorithm compared their classification accuracy with the gold-standard and gradually offered images with lower ML confidence scores.

The second subset of studies is characterized by ML algorithms helping volunteers to perform their tasks most efficiently. Coughlin et al. (2019) reported on using ML to cluster similar images to make it easier for volunteers in Gravity Spy to identify new glitch classes. Wright et al. (2019) described the symbiotic relationship between ML and citizen scientists in the Supernova Hunters project. Since ML required an extensive training dataset, it could take a long time for volunteers to classify enough images. Thus, an initial clustering of similar images was suggested to help reduce volunteer effort in labeling the dataset.

The third subset refers to studies that report on volunteers validating the ML results. Users of the Leafsnap app uploaded photos of leaves, and if the shape-matching algorithm classification did not reach the required level of confidence, users were given several options to help identify the species (Bowley et al. 2019). Lambers et al. (2019) presented an integrated approach in the domain of archeology. The algorithm was trained to detect potential archaeological objects in remotely sensed data, then citizens and experts validated findings through three validation steps. The results were fed to the algorithm to improve object detection accuracy.

To summarize, the discussed workflows represent continuous interdependent relationships between citizens and algorithms, where citizens develop expertise, make discoveries, and help validate ML results, while ML algorithms train citizens, help them, and continuously improve as well, due to the ongoing feedback from citizens.

Discussion

The results of this review show that several projects have already integrated human efforts and ML to perform data-centered tasks. In citizen science, the integration of human efforts and algorithmic processing has taken different directions but, currently, it is represented by ML, especially applied to computer vision, which includes diverse methods of automatically identifying objects from digital photographs (Ceccaroni et al. 2019).

Our review has revealed which problems are commonly addressed by combining citizen science and ML efforts, which tasks are allocated to humans and machines, and how those tasks are integrated into one workflow. Currently, several citizen science projects using ML are centered around analyzing, coding/classifying, and clustering data provided, for example, by cameras and telescope images. Collecting, coding/classifying, and interpreting data are some of the most common activities that participants carry out, depending on their level of engagement in the research process (Shirk et al. 2012).

Similarly, ML is used at various stages of the data–science life cycle, through algorithms that perform tasks like classification, regression, clustering, and association, especially when dealing with huge amounts of data.

Unlike automated systems that tend to remove or reduce the need for humans, the reported projects can be considered heteromated systems that do not function without the necessary mediation of engaged citizens (Ekbia and Nardi 2017). Both citizens and experts constitute the "human infrastructure", without which citizen science projects cannot exist. According to Ekbia and Nardi, citizen science projects are not heteromated in terms of experts' work, but through the participation of volunteers who act as indispensable mediators. Their heteromated systems push critical tasks to citizens, as they seek them to provide, classify, and annotate data, or solve problems. In heteromated systems, efficiency is not opposed to engagement (Ekbia and Nardi 2017), as they cater to people's motivations to participate in projects that offer opportunities for meaningful tasks. In citizen science projects, researchers are involved in recruiting, training, and supporting volunteers.

Leveraging the complementarity of strengths is one of the main arguments to combine humans and machines. For example, machines cannot yet match the human ability to identify certain objects and it is unclear to what extent they will ever succeed. Conversely, manual classification or identification of objects from images in a large data set can be made more efficient in combination with ML. Even so, the participation of citizens remains critical to perform certain tasks, such as creating data sets with correctly tagged data to feed algorithms (Torney et al. 2019). Galaxy Zoo and the classification and identification of galaxy morphological shapes is a good case in point (Walmsley et al. 2019).

While the complementarity of humans and machine is commonly acknowledged to enhance their respective capabilities, essentializing the attributes of humans and machines, as in Fitt's MABA-MABA – Men Are Better At -vs- Machines Are Better At – list (1951), should be avoided. Treating his list of attributes as stable and natural does not take into account that cognitive work will be shifting between humans and machines (Ekbia and Nardi 2017), as the range of research tasks that machines can do is growing, although algorithms are still second to humans on recognizing patterns and they have longer learning curves.

Conclusions

In this paper, we have summarized the results of a systematic review of human-machine integration in citizen science projects. Although we conducted this review to include all relevant articles and collect and synthesize the data from them in a meaningful way, there can be articles we did not include as we left out preprints, reports, and other types of non-peer-reviewed literature. Most of the reviewed articles are methodology papers aimed at designing or comparing various ML algorithms. Only a limited number of articles in this review focus on the development of integrated systems, where both humans and machines would benefit from collaboration. The questions of using AI-generated feedback to train citizen scientists, increase their engagement, or inform them about the quality of their contributions – while insightfully addressed in few articles – largely remain under-researched. Finally, most of the reviewed research used citizen contributions in the form of data collection or classification, while only a few investigated how ML can facilitate serendipitous discovery (Trouille, Lintott, and Fortson, 2019).

We discussed the “character” of human-machine systems in citizen science as heteromated systems, wherein citizen participation remains crucial and volunteer and ML efforts are often positioned as complementary rather than mutually exclusive. Future research should examine how algorithms taking over some human tasks will affect the balance between humans and machines in citizen science. Further exploration of the kind of partnership and delegation of skills that are created when integrating humans and machines should also be considered. A close examination of some specific cases could reveal what and how ML contributes to the creation of knowledge and whether it will result in skill-biased technological change. The results of such studies may achieve designs that have an impact on citizen science and the field of collective intelligence.

Competing Interests

The authors have no competing interests to declare.

Data Accessibility Statement

The authors declare that all the data supporting the findings of this study are in the following supplementary files:

Supplementary file 1: Appendix 1 – Summary table with the terms found in the literature

Supplementary file 2: Appendix 2 – Spreadsheet with the list of articles included in the review

Supplementary file 3: Appendix 3 – Summary table of the tasks as reported in the literature

Funding Information

This work has been supported by…

Author Contributions

Author 1 contributed to the acquisition, analysis, and interpretation of the resources, and wrote the results section. Author 2 also contributed to the literature search, while Author 3 wrote other sections of the paper and revised critically the content and the integrity of the work.

References

Adams MD, Massey F, Chastko K, and **Cupini C.** 2020. Spatial modelling of particulate matter air pollution sensor measurements collected by community scientists while cycling, land use regression with spatial cross-validation, and applications of machine learning for data correction. *Atmospheric Environment*, 230:117479. DOI: https://doi.org/10.1016/j.atmosenv.2020.117479

Baert C. 2019. Citizen science game. Available at https://citizensciencegames.com/games/ (Last accessed 21 October 2020).

Bahaadini S, Noroozi V, Rohani N, Coughlin S, Zevin M, Smith JR, Kalogera V and **Katsaggelos A.** 2018. Machine learning for Gravity Spy: glitch classification and dataset. *Information Sciences*, 444:172–86. DOI: 10.1016/j.ins.2018.02.068

Beaumont C N, Goodman AA, Kendrew S, Williams JP, and **Simpson R.** 2014. The Milky Way Project: leveraging citizen science and machine learning to detect interstellar bubbles. *The Astrophysical Journal Supplement Series*, *214*(1): 3. Copy at http://www.tinyurl.com/yymgqpye

Blickhan S, Trouille L and **Lintott CJ.** 2018. Transforming research (and public engagement) through citizen science. *International Astronomical Union. Proceedings of* *the International Astronomical Union.* 14(A30): 518-523. (Section 4). DOI: https://doi.org/10.1017/S174392131900526X

Botella C, Joly A, Bonnet P, Monestiez P and **Munoz F.** 2018. Species distribution modeling based on the automated identification of citizen observations. *Applications in Plant Species*, 6(2):e1029. DOI: 10.1002/aps3.1029.

Bowley C, Mattingly M, Barnas A, Ellis-Felege S and **Desell T.** 2019. An analysis of altitude, citizen science and a convolutional neural network feedback loop on object detection in Unmanned Aerial Systems. *Journal of Computational Science*, 34:102–116. DOI: 10.1016/J.JOCS.2019.04.010

Capinha C. 2019. Predicting the timing of ecological phenomena using dates of species occurrence records: a methodological approach and test case with mushrooms. *International Journal of Biometeorology* 63(8):1015–24. DOI: https://doi.org/10.1007/s00484-019-01714-0

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

Coughlin S, Bahaadini S, Rohani N, Zevin M, Patane O, Harandi M, ... and **Coughlin M.** 2019. Classifying the unknown: discovering novel gravitational-wave detector glitches using similarity learning. *Physical Review D*, 99(8), 082002. DOI: https://doi.org/10.1103/PhysRevD.99.082002

Crowston K, Osterlund C, Lee TK, Jackson C, Harandi M, Allen S, Bahaadini S, Coughlin S, Katsaggelos AK, Larson SL, Rohani N, Smith JR, Trouille L and **Zevin**

M. 2020. Knowledge Tracing to Model Learning in Online Citizen Science Projects. *IEEE Transactions on Learning Technologies,* 13(1):123–34. DOI: 10.1109/TLT.2019.2936480

Curry CM, Ross JD, Contina AJ and **Bridge ES**. 2018. Varying dataset resolution alters predictive accuracy of spatially explicit ensemble models for avian species distribution. *Ecology and Evolution*, 8(24):12867-12878. DOI: 10.1002/ece3.4725.

Das R, Keep B, Washington P and **Riedel-Kruse IH.** (2019). Scientific discovery games for biomedical research. *Annual Review of Biomedical Data Science*, 2: 253-279. (section 6.4). DOI: https://doi.org/10.1146/annurev-biodatasci-072018-021139

Derville S, Torres LG, Iovan C, and **Garrigue C. 2018.** Finding the right fit: comparative cetacean distribution models using multiple data sources and statistical approaches. *Diversity and Distributions* 24(11):1657–73. DOI: https://doi.org/10.1111/ddi.12782

de Winter JCF and **Dodou D**. 2014. Why the Fitts list has persisted throughout the history of function allocation. *Cognition, Technology and Work*, 16:1–11. DOI 10.1007/s10111-011-0188-1

Duo X and **Offner Stella SR.** 2017. Assessing the performance of a machine learning algorithm in identifying bubbles in dust emission. *The Astrophysical Journal* 851(2):149. DOI: https://doi.org/10.3847/1538-4357/aa9a42

Ekbia HR and **Nardi, BA**. 2017. *Heteromation, and Other Stories of Computing and Capitalism*. Cambridge, MA: MIT Press.

Everaars J, Strohbach MW, Gruber B, and **Dormann CF.** 2011. Microsite conditions dominate habitat selection of the red mason bee (*Osmia Bicornis, Hymenoptera: Megachilidae*) in an urban environment: a case study from Leipzig, Germany. *Landscape and Urban Planning*, 103(1):15–23.

Fitts PM. 1951. *Human Engineering for an Effective Air-Navigation and Traffic-Control System*. Division of National Research Council: Oxford, England.

Hackman JR. (1969). Toward understanding the role of tasks in behavioral research. *Acta Psychologica*, 31: 97-128. DOI: https://doi.org/10.1016/0001-6918(69)90073-0

Hardison DR, Holland WC, Currier RD, Kirkpatrick B, Stumpf R, Fanara T, Burris D, Reich A, Kirkpatrick GJ and **Litaker RW.** 2019. HABscope: A tool for use by citizen scientists to facilitate early warning of respiratory irritation caused by toxic blooms of Karenia brevis. *PLoS One*. 14(6):e0218489. DOI: 10.1371/journal.pone.0218489.

Jackson MM, Gergel SE and **Martin K.** 2015. Citizen science and field survey observations provide comparable results for mapping Vancouver Island White-tailed Ptarmigan (Lagopus leucura saxatilis) distributions. *Biological Conservation*, 181: 162-172. DOI: https://doi.org/10.1016/j.biocon.2014.11.010

Jiménez M, Torres MT, John R and **Triguero I.** 2020. Galaxy image classification based on citizen science data: a comparative study. *IEEE Access* 8:47232–46. DOI:10.1109/ACCESS.2020.2978804

Kelling S, Gerbracht J, Fink D, Lagoze C, Wong WK, Yu J … and **Gomes C**. 2013. A human/computer learning network to improve biodiversity conservation and research. *AI magazine*. 34(1): 10-10. DOI: https://doi.org/10.1609/aimag.v34i1.2431

Kerkow A, Wieland R, Früh L, Hölker F, Jeschke JM, Werner D and **Kampen H**. 2020. Can data from native mosquitoes support determining invasive species habitats? Modelling the climatic niche of Aedes japonicus japonicus (Diptera, Culicidae) in Germany. *Parasitology Research*, 119(1):31-42. DOI: 10.1007/s00436-019-06513-5.

Keshavan A, Yeatman JD and **Rokem A**. Combining citizen science and deep learning to amplify expertise in neuroimaging. *Frontiers in Neuroinformatics*, 13:29. DOI: 10.3389/fninf.2019.00029.

Kim JS, Greene MJ, Zlateski A, Lee K, Richardson M, Turaga SC, Purcaro M, Balkam M, Robinson A, Behabadi BF, Campos M, Denk W, Seung HS and **EyeWirers.** 2014. Space-time wiring specificity supports direction selectivity in the retina. *Nature*, 509(7500):331-336. DOI: 10.1038/nature13240.

Koodli RV, Keep B, Coppess KR, Portela F, Eterna participants and **Das R.** 2019. PLoS EternaBrain: Automated RNA design through move sets and strategies from an Internet-scale RNA videogame. *PLOS Computational Biology*, 15(6):e1007059. DOI: 10.1371/journal.pcbi.1007059.

Kress WJ, Garcia-Robledo Carlos, Soares, João VB, Jacobs D, Wilson K, Lopez IC and Belhumeur PN. 2018. Citizen science and climate change: mapping the range expansions of native and exotic plants with the mobile app Leafsnap. *BioScience*, 68(5):348–358. DOI: https://doi.org/10.1093/biosci/biy019

Kumagai NH, Yamano H and **Committee Sango-Map-Project.** 2018. High-resolution modeling of thermal thresholds and environmental influences on coral bleaching for local and regional reef management. PeerJ. 6:e4382. DOI: 10.7717/peerj.4382.

Kuminski E, George J, Wallin J and **Shamir L.** 2014. Combining human and machine learning for morphological analysis of galaxy images. *Publications of the Astronomical Society of the Pacific*, 126(944):959–67. DOI: 10.1086/678977

Lambers K, Verschoof-van der Vaart W B and **Bourgeois QP.** 2019. Integrating remote sensing, machine learning, and citizen science in Dutch archaeological prospection. *Remote Sensing*, 11(7): 794. DOI: https://doi.org/10.3390/rs11070794

Lee J, Kladwang W, Lee M, Cantu D, Azizyan M, Kim H, Limpaecher A, Gaikwad S, Yoon S, Treuille A, Das R and **EteRNA Participants.** 2014. RNA design rules from a massive open laboratory. *Proceedings of the National Academy of Sciences of the United States of America,* 111(6):2122-7. DOI: 10.1073/pnas.1313039111

Lim CC, Kim H, Vilcassim MJR, Thurston GD, Gordon T, Chen LC, Lee K, Heimbinder M and **Kim SY.** 2019. Mapping urban air quality using mobile sampling with low-cost sensors and machine learning in Seoul, South Korea. *Environment International*, 131:105022. DOI: 10.1016/j.envint.2019.105022.

Lintott C and **Reed J.** 2013. Human computation in citizen science. In: Michelucci P (ed.), *Handbook of Human Computation,* 153-162. New York: Springer. DOI: 10.1007/978-1-4614-8806-4

Lukic V, Brüggen M, Banfield JK, Wong OI, Rudnick L, Norris RP and **Simmons B.** 2018. Radio Galaxy Zoo: compact and extended radio source classification with deep learning. *Monthly Notices of the Royal Astronomical Society*, 476(1): 246-260. DOI: https://doi.org/10.1093/mnras/sty163

Mac Aodha O, Gibb R, Barlow KE, Browning E, Firman M, Freeman R, Harder B, Kinsey L, Mead GR, Newson SE, Pandourski I, Parsons S, Russ J, Szodoray-Paradi A, Szodoray-Paradi F, Tilova E, Girolami M, Brostow G and **Jones KE**. 2018. Bat detective-Deep learning tools for bat acoustic signal detection. *PLoS Computational Biology,* 14(3):e1005995. DOI: 10.1371/journal.pcbi.1005995.

Mulgan G. 2018. Artificial intelligence and collective intelligence: the emergence of a new field. *AI&Society*, 33:631–632. DOI: https://doi.org/10.1007/s00146-018-0861-5**.**

Nguyen T, Pankratius V, Eckman L and **Seager S.** 2018. Computer-aided discovery of debris disk candidates: A case study using the Wide-Field Infrared Survey Explorer (WISE) catalog. *Astronomy and computing*, 23: 72-82. DOI: 10.1016/j.ascom.2018.02.004

Pearse WD, Morales-Castilla I, James LS, Farrell M, Boivin F and **Davies TJ**. 2018. Global macroevolution and macroecology of passerine song. *Evolution*, 72(4):944-960. DOI: 10.1111/evo.13450.

Robinson OJ, Ruiz-Gutierrez V and **Fink D** 2018. Correcting for bias in distribution modelling for rare species using citizen science data. *Diversity and Distributions*, 24(4): 460-472. DOI: https://doi.org/10.1111/ddi.12698

Shamir L, Diamond D and **Wallin J.** 2016. Leveraging pattern recognition consistency estimation for crowdsourcing data analysis. *IEEE Transactions on Human-Machine Systems*, 46(3): 474-480. DOI: 10.1109/THMS.2015.2463082

Shirk, JL, Ballard, HL, Wilderman, CC, Phillips, T, Wiggins, A, Jordan, R, McCallie, E, Minarchek, M, Lewenstein, BV, Krasny, ME and **Bonney, R.** 2012. Public participation in scientific research: A framework for deliberate design. *Ecology and Society*, 17(2): 29. DOI: https://doi.org/10.5751/ES-04705-170229

Sullivan DP, Winsnes CF, Åkesson L, Hjelmare M, Wiking M, Schutten R, Campbell L, … and **Lundberg E.** 2018. Deep learning is combined with massive-scale citizen science to improve large-scale image classification. *Nature Biotechnology*, *36*(9): 820–832. DOI: https://doi.org/10.1038/nbt.4225.

Tar PD, Bugiolacchi R, Thacker NA, Gilmour JD and **MoonZoo Team.** 2017. Estimating false positive contamination in crater annotations from citizen science data. *Earth Moon Planets*, 119(2):47-63. DOI: 10.1007/s11038-016-9499-9.

Terry J, Christopher D, Roy HE and **Tom A.** 2020. Thinking like a naturalist: Enhancing computer vision of citizen science images by harnessing contextual data. *Methods in Ecology and Evolution*, 11(2):303–15. DOI: 10.1111/2041-210X.13335

Thornhill I, Ho JG, Zhang Y, Li H, Ho KC, Miguel-Chinchilla L and **Loiselle SA**. 2017. Prioritising local action for water quality improvement using citizen science; a study across three major metropolitan areas of China. *Science Total Environment*, 584-585:1268-1281. DOI: 10.1016/j.scitotenv.2017.01.200.

Torney CJ, Lloyd-Jones DJ, Chevallier M, Moyer DC, Maliti HT, Mwita M, Kohi EM and **Hopcraft GC. 2019.** A comparison of deep learning and citizen science techniques for counting wildlife in aerial survey images. *Methods in Ecology and Evolution*, 10(6):779–87. DOI: 10.1111/2041-210x.13165

Trouille L, Lintott CJ and **Fortson LF.** (2019). Citizen science frontiers: Efficiency, engagement, and serendipitous discovery with human–machine systems. *Proceedings of the National Academy of Sciences*, 116(6): 1902-1909. DOI: 10.1073/pnas.1807190116

Walmsley, M., Smith, L., Lintott, C., Gal, Y., Bamford, S., Dickinson, H., …**Wright, D.** 2020). Galaxy Zoo: probabilistic morphology through Bayesian CNNs and active learning. *Monthly Notices of the Royal Astronomical Society*, 491(2): 1554-1574. DOI: https://doi.org/10.1093/mnras/stz2816

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

Winter M, Bourbeau J, Campos F, Gallay B, Meehan M, Morgan R, Peacock J, Pizzuto A, Schneider C, Simons AL and **Vandenbroucke J.** 2019. Particle identification in smartphone camera images using the distributed electronic cosmic-ray observatory. *Astroparticle Physics*, 104:42–5. DOI: https://doi.org/10.1016/j.astropartphys.2018.08.009

Wright DE, Fortson L, Lintott C, Laraia M and **Walmsley M.** 2019. Help me to help you: machine augmented citizen science. *ACM Transactions on Social Computing*, *2*(3):1-20. DOI: https://doi.org/10.1145/3362741

Wright DE, Lintott CJ, Smartt SJ, Smith KW, Fortson L, Trouille L, Allen CR … and **Chambers KC.** 2017. A transient search using combined human and machine classifications. *Monthly Notices of the Royal Astronomical Society*. 472(2): 1315-1323. DOI: https://doi.org/10.1093/mnras/stx1812

Zevin M, Coughlin S, Bahaadini S, Besler E, Rohani N, Allen S ... and **Lee TK.** 2017. Gravity Spy: Integrating advanced LIGO detector characterization, machine learning, and citizen science. *Classical and Quantum Gravity*, 34(6), 064003. DOI: https://doi.org/10.1088/1361-6382/aa5cea

Zilli D, Parson O, Merrett GV, and **Rogers A.** 2014. A hidden Markov model-based acoustic cicada detector for crowdsourced smartphone biodiversity monitoring. *Journal of Artificial Intelligence Research*, 51:805–27. DOI: https://doi.org/10.1613/jair.4434