

Review

# The Partnership of Citizen Science and Machine Learning: Benefits, Risks, and Future Challenges for Engagement, Data Collection, and Data Quality

Maryam Lotfian <sup>1,2,\*</sup> , Jens Ingensand <sup>1</sup> and Maria Antonia Brovelli <sup>2</sup> 

<sup>1</sup> Institute INSIT, School of Business and Engineering Vaud, University of Applied Sciences and Arts Western Switzerland, 1400 Yverdon-les-Bains, Switzerland; jens.ingensand@heig-vd.ch

<sup>2</sup> Department of Civil and Environmental Engineering, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy; maria.brovelli@polimi.it

\* Correspondence: maryam.lotfian@polimi.it or maryam.lotfian@heig-vd.ch

**Abstract:** Advances in artificial intelligence (AI) and the extension of citizen science to various scientific areas, as well as the generation of big citizen science data, are resulting in AI and citizen science being good partners, and their combination benefits both fields. The integration of AI and citizen science has mostly been used in biodiversity projects, with the primary focus on using citizen science data to train machine learning (ML) algorithms for automatic species identification. In this article, we will look at how ML techniques can be used in citizen science and how they can influence volunteer engagement, data collection, and data validation. We reviewed several use cases from various domains and categorized them according to the ML technique used and the impact of ML on citizen science in each project. Furthermore, the benefits and risks of integrating ML in citizen science are explored, and some recommendations are provided on how to enhance the benefits while mitigating the risks of this integration. Finally, because this integration is still in its early phases, we have proposed some potential ideas and challenges that can be implemented in the future to leverage the power of the combination of citizen science and AI, with the key emphasis being on citizen science in this article.

**Keywords:** citizen science; machine learning; big data; artificial intelligence; task automation; engagement; data validation



**Citation:** Lotfian, M.; Ingensand, J.; Brovelli, M.A. The Partnership of Citizen Science and Machine Learning: Benefits, Risks, and Future Challenges for Engagement, Data Collection, and Data Quality. *Sustainability* **2021**, *13*, 8087. <https://doi.org/10.3390/su13148087>

Academic Editor: Jakub Trojan

Received: 2 June 2021

Accepted: 7 July 2021

Published: 20 July 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The simulation of human intelligence in machines, known as artificial intelligence (AI), is widely applied in various domains, and the number of scientific publications in this area are significantly increasing [1]. AI is a term used when machines can perform tasks which simulate the human mind such as learning, reasoning, and solving problems [2]. Thus, machine learning (ML) is a sub field of AI, defined as the study of developing computer algorithms, which use data to learn patterns, make predictions, and improve their performance over time by more data [3]. The majority of ML algorithms require large amounts of labeled data, and this is resulting in a close partnership of ML with citizen science projects [4,5]. Citizen science—public participation in scientific research—has grown significantly in recent years as a result of technological advancements such as new smartphone features and fast Internet access in most parts of the world [6]. This growth in citizen science has resulted in large dataset collections in a variety of scientific domains [7], which can be a valuable input source for ML algorithms.

Although the combination of ML and citizen science is not new [8], until recently, these two fields have mostly been implemented separately [9]. The integration of ML and citizen science can result in producing a new learning paradigm for citizen scientists through human–computer interactions [10]. Moreover, it can result in increasing interdisciplinary

collaborations among researchers as well as members of the public in various fields such as computer science, ecology, astronomy, and medicine, to name a few [9]. This integration has been focused primarily on object detection in images and videos with the main focus on automatic species identification in biodiversity projects [11,12]. A well-known example is the iNaturalist project [13], which has included automated species identification suggestions since 2017 using images obtained from observers. The automatic identification has improved over the years as more images are used to train the model, and the latest model release was in March 2020 by the time of writing this article [14]. The automatic species identification in iNaturalist has provided citizen scientists the opportunity to learn about species and to minimize the contribution of erroneous observations [15].

The objective of combining citizen science and ML is not limited to providing data for the ML algorithms and automatizing the identification tasks. The aim is to combine human and machine intelligence to bring new adjustments to citizen science tasks, such as automated data collection, processing, and validation, as well as to increase public engagement. There are potential challenges and opportunities in the integration of ML and citizen science, which are essential to discuss. In this article, we aim at discussing the following research questions:

- What are some examples of successful citizen science projects where ML is integrated?
- What ML techniques have been used in these projects?
- What citizen science tasks have been affected by ML in such projects?
- What are the benefits and risks of integrating ML in citizen science for practitioners and citizen scientists?
- What are the possible future challenges that might arrive as a result of the combination of ML and citizen science?
- What are the gaps and limitations of including ML in citizen science?

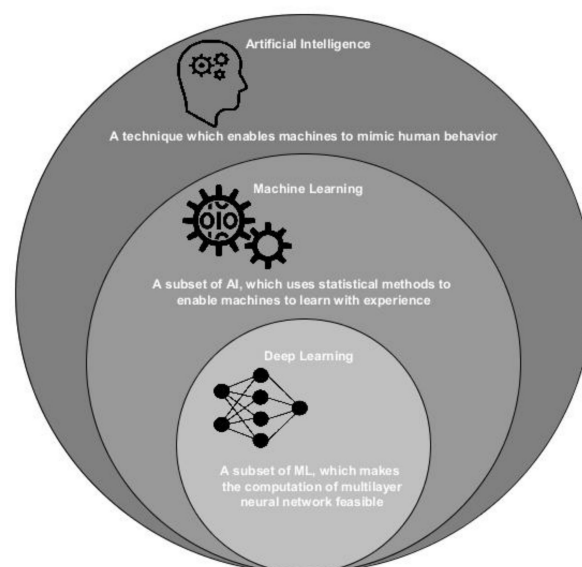
To answer these research questions, we explore use cases where ML and citizen science can be combined. We have reviewed successful citizen science projects, highlighting the typologies of techniques used in such projects and categorizing them in light of the effect of ML on citizen science tasks. Although the opportunities and challenges of merging ML and citizen science have been addressed in a few recent articles [8–10], the main emphasis has been on the transparency of using ML in citizen science in terms of how the ML algorithms use citizen science data [10], the effects of AI on human behavior and improving insights in citizen science [8], and the effects of this combination in ecological monitoring in terms of having cheaper or more efficient ways for data collection and data processing [9]. While these are key issues to explore, to the best of our knowledge, the integration of ML and CS has received less attention in terms of how this integration can affect the usual processes in a CS project, from volunteer involvement to influencing the quality of their contributions. Our primary objective is to explore how some CS tasks can be automated using ML and whether this automation is beneficial or detrimental for the project and its participants. Rather than being overly broad, we broke down the forms of ML combinations in various CS steps and discussed the benefits and risks of this integration in each step in this article. We outline how ML can be integrated in each step, including what has already been applied, what can be applied in the future, and what the current and potential risks and benefits of this integration are for each step.

The following is how the article is organized. In the following section, we will go through the ML paradigm, as well as the most popular ML applications, in greater detail. In Section 3, we explore the potential impacts of ML on citizen science projects, and in Section 4, we review successful use cases where ML and citizen science are combined. In Section 5, we discuss the benefits and risks of integrating citizen science and ML. Finally, in Section 6, we present the conclusions, with an emphasis on possible future transitions in citizen science projects in the age of AI.

## 2. Types of Machine Learning and Applications

As stated in the introduction, ML is a subset of AI, which was first introduced in 1955 by Arthur Samuel when he applied learning to his draughts (checkers) algorithm [2]. Samuel defined ML as a “field of study that gives computers the ability to learn without being explicitly programmed” [16]. ML algorithms build models which learn using the input data (known as training data) and are able to make predictions based on the learnt experience. There are three main machine learning types, known as supervised learning, unsupervised learning, and reinforcement learning [16].

- **Supervised learning:** In supervised learning, the training data are labeled, and the task is to map the input (independent variables) to the output (dependent variables). The two typical types of supervised learning are classification, where the output variable is categorized, and regression, where the output variable is continuous [16]. The most widely known algorithms of supervised learning are k-nearest neighbors (KNN), linear regression, logistic regression, support vector machines (SVMs), decision trees, random forest (RF), and neural networks (NN).
- **Unsupervised learning:** In unsupervised learning, the training data are not labeled, and the goal is to identify structures and patterns in the data [16]. The typical types of unsupervised learning include clustering (grouping similar input data), dimension reduction (extracting meaningful features from the data), and association (exploring the data to discover relationships between attributes) [16]. Some of the most known algorithms of unsupervised learning are k-means, one-class SVM, hierarchical cluster analysis (HCA), and principal component analysis (PCA).
- **Reinforcement learning:** In reinforcement learning, the learning algorithm, also called the agent, observes the environment and learns through a system of rewards and punishments. Reinforcement learning is commonly used in robotics, such as walking robots and self-driving vehicles, as well as in real-time decision making and game AI [16].
- **Deep learning,** a subset of ML (See Figure 1 for the relationship between AI, ML, and deep learning), is concerned with algorithms known as artificial neural networks that attempt to simulate the structure and functions of a biological brain [17]. Since there is a significant body of literature on AI and ML algorithms, we briefly discuss some of the common AI, ML, and deep learning techniques applied largely in scientific projects:



**Figure 1.** Relationship between artificial intelligence, machine learning, and deep learning.

- **Computer vision (CV):** CV is an interdisciplinary scientific field which aims at developing techniques so that computers can identify and understand the contents in

digital images and videos. In other words, CV aims at enabling computers to identify elements in images the same as humans would do. The advances in artificial neural networks and deep learning have had great impact on CV, which in some cases outperforms the human power to identify objects [18]. Some popular applications of CV include self-driving cars, face recognition, etc. [16]. Moreover, starting in the year 2020 and with the COVID-19 pandemic, CV has been applied in monitoring and detecting social distancing among people [19]. CV has also been commonly used in species identification, with Plant@net [20] and iNaturalist being two well-known citizen science examples. A class of deep learning which is commonly used in CV is the convolutional neural network (CNN).

- Natural language processing (NLP): NLP is a subfield of linguistics, computer science and AI that deals with human–computer interactions through the use of natural language, which means that NLP aims to enable computers to read and understand human language [21]. The mechanism involves the machine capturing the human’s words (text or audio), processing the words and preparing a response, and returning the produced response (in the form of audio or text) to the human. Language translation applications such as Google Translate or DeepL [22], as well as personal assistant applications (e.g., Siri or Alexa), are common uses of NLP in people’s daily lives.
- Acoustic identification: Acoustic identification is a technique based on pattern recognition and signal analysis, where the acoustic data are processed and features are extracted and classified. Main applications of acoustic identification are in species detection [23]. For example, BirdNet [24] is an application to identify bird species based on the bird song.
- Automated reasoning: Automated reasoning is a branch of AI that seeks to train machines to solve problems using logical reasoning [25]. In other words, in automated reasoning, the computer is given knowledge and can generate new knowledge from it, which it then uses to make rational decisions. Automated reasoning is mainly used to assess if something is true or false or whether an event will occur or not.

### 3. The Influence of ML on Citizen Science Steps

When it comes to the combination of ML and citizen science, the role of citizen science as a possible solution to the problem of a lack of training data in ML algorithms [7,26] has been discussed more intensively than the role of ML in addressing challenges in citizen science projects. Ceccaroni et al. [8] explored the AI technologies used in citizen science projects and the opportunities and risks that are expected to be encountered due to the increase in the use of AI in citizen science. The authors define three categories for the use of AI in citizen science including “assisting or replacing humans for completing tasks”, “influencing human behavior”, and “improving insights”. The first category describes the role of AI in fully or partially automating tasks that were previously performed by humans: for example, tasks related to automatically detecting and classifying data, such as classifying species based on images or sounds [27–29]. The second category discusses the aim of AI, data science, and citizen science to influence human behavior [30] and to extend the educational and social benefits of citizen science to the general public [31]. The third category discusses the impact of AI on identifying patterns in citizen science data for informing research and policies or on facilitating the understanding of citizen science concepts using ontologies. Another study by McClure et al. [9] discusses the integration of AI and citizen science in ecological monitoring. Rather than delving into the details of how AI and citizen science can be combined, the authors addressed the challenges and opportunities of performing ecological monitoring using only citizen science, only AI, or a combination of the two. The opportunities and challenges are discussed in the context of six categories, including efficiency, accuracy, discovery, engagement, resources, and ethics. Efficiency refers to the benefits that citizen science and ML can provide for scientific projects, such as facilitating data collection and automating laborious tasks, as well as the ability to perform extensive data processing when human and machine power are

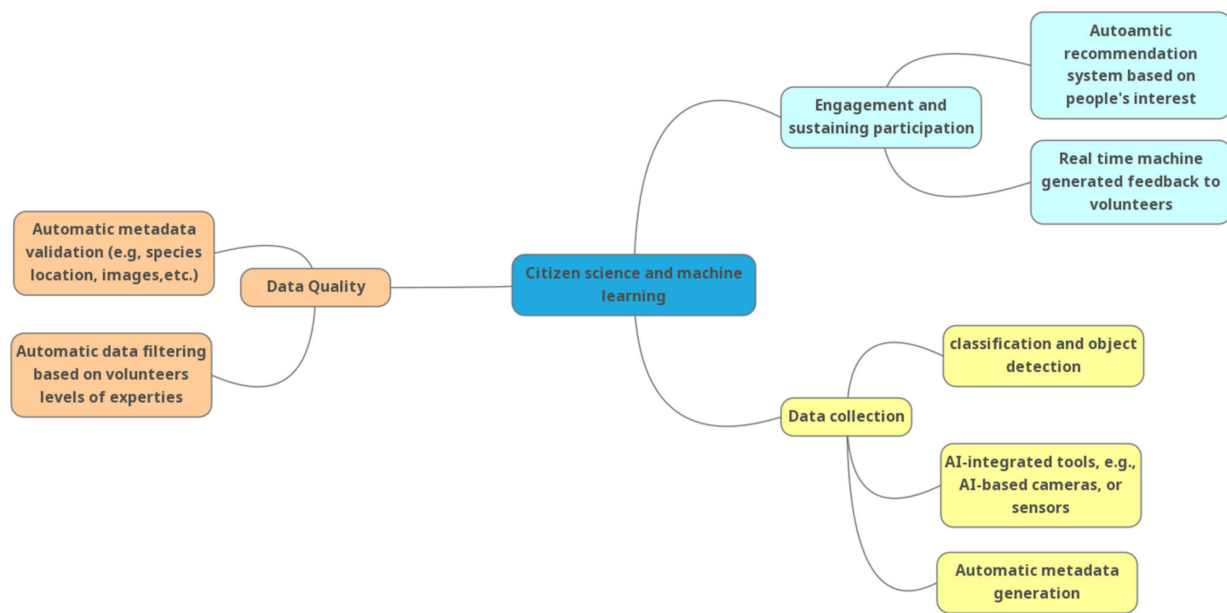
combined. Accuracy refers to the possibility of integrating human and machine intelligence to produce high-quality data or the challenge of providing incorrect and misleading information. Discovery explores the advantages of complex species identification and serendipitous discoveries made through the partnership of citizen scientists and deep learning. Engagement explores the impact of citizen science and AI on multidisciplinary engagement. Resources highlights the role of citizen scientists and machines in saving human and financial resources by, for example, freely contributing data and automating complex tasks, but it also covers the challenges of training citizen scientists, large data requirements, and the need for ML experts. Ethics highlights the challenges of potential information misuse when integrating AI. Another recent study by Franzen et al. [10] also discusses the opportunities and challenges of human–computer interaction in citizen science with a focus on the concept of transparency when integrating ML in citizen science projects, which means that information about data use, ML algorithms, and data processing must be transparent and communicated to participants.

In this article, we will look at the impact of ML and citizen science integration on citizen science steps, but first, it is important to understand the different types of citizen science projects, as well as the main steps and tasks in a project. Bonney et al. [32] described three types of citizen science projects: contributory projects, in which scientists design the project and members of the public contribute primarily to data collection; collaborative projects, in which scientists design the project and members of the public contribute not only to data collection but also to data analysis and/or interpretation of the findings; and finally, co-created projects are those in which the project is designed in collaboration with scientists and members of the public, and some members of the public are involved in most, if not all, of the project steps. Citizen science projects are comprised of five key steps, with participants engaging in all or some of the steps depending on the project type. The following are the primary steps for each citizen science project [33,34]:

- Defining the problem: Exploring the problem that needs to be solved by answering questions, such as why this issue is important, who the stakeholders are, and what will be achieved.
- Designing the project: Identifying the objectives, allocating the necessary resources (funding, team members, equipment, etc.), and defining the project planning.
- Building a community: Encouraging the general public to participate in the project and sustaining their engagement by establishing a trusting relationship with the volunteers.
- Data collection, quality assurance, and analysis: Designing data collection tools, training volunteers, determining how to store data, filtering and cleaning collected data, analyzing data to detect trends, and sharing data with participants or other practitioners.
- Sustain and improve the project: Maintaining project funding by searching for different sources of funding, and sustaining participation by communicating with volunteers and receiving/giving feedback from/to them.

Thus, our goal is to expand the existing literature on the integration of citizen science and ML by focusing not only on the scientific outcomes of citizen science projects, but also on the participants, who are at the heart of the projects. We therefore address the integration of ML into various components of a citizen science project, and focus on the impacts of ML on three categories: engaging people and sustaining their participation, data collection, and data validation (Figure 2).





**Figure 2.** A taxonomy showing the integration of machine learning and citizen science based on the three citizen science steps of engagement, data collection, and data quality.

### 3.1. ML for Engaging the Public and Sustaining Participation

A key aspect in a successful citizen science project is to understand how to motivate the public to participate in a project and how to sustain their participation [35]. Depending on the objectives and designs of the citizen science project, various approaches have been used to engage people [36]. We discuss two potential approaches in using ML towards engaging participants and sustaining participation:

- **Automatic community search:** The traditional approaches such as word-of-mouth, social media posts, direct emails, workshops, etc., while beneficial for building a community, can be time consuming or require financial resources (for instance, for organization of workshops or ads in newsletters). Antoniou et al. [37] have proposed a guidance tool to provide information to volunteers so that they can find the VGI (volunteered geographic information) project of their choice based on their motivations and interests. To automate what they have proposed, ML algorithms can be used to find and classify the potential target participants based on their interests and to introduce a project to them accordingly. Several studies have been conducted to apply ML algorithms to extract relevant information from social media (e.g., Twitter or Instagram) posts, such as where the images were taken, what type of content is contained in the image, or what topic is mostly discussed in the textual posts [38,39]. As a result, similar approaches can be adapted to citizen science projects by employing ML techniques such as CV and NLP to identify people's interests from social media posts and linking them to the relevant citizen science project. Furthermore, to the best of our knowledge, the use of ML in user profiling to create a recommendation system [40,41] where citizen science projects are recommended to people based on their sociodemographic details is not used as a way to engage people to contribute to citizen science projects. Moreover, the use of chat bots in citizen science projects can be a potential approach in engaging and sustaining participation, which has been applied in few studies [42,43]. Chat bots may also help as a real-time guide for participants.
- **Automatic feedback to participants:** As discussed in some studies, participants may become discouraged if they do not receive feedback on their contributions [44,45]. Moreover, due to massive amounts of data, it is time-consuming to provide feedback to all participants, or often, feedback from experts is provided after a long time has passed [45,46]. In order to inform participants regarding the quality of their

contributions and to update them regarding the project advancements, automatic informative and user-based feedback can be generated using ML algorithms [47]. The participants can be informed about the quality of their contribution and how they can enhance it and can learn from the feedback provided (e.g., learning about biodiversity through feedback regarding species habitat characteristics). Thus, human–computer interaction through machine-generated feedback can be a strategy for increasing and sustaining participation in citizen science projects.

### 3.2. ML for Data Collection

Data collection in citizen science projects usually can be categorized into two types. The first category is known as crowdsourcing [48], and it involves data collection that requires little or no cognition engagement, such as collecting biodiversity data (e.g., photographs of species), recording noise [49] or air pollution levels, or in volunteer computing projects, in which volunteers provide their computer’s unused resources for scientists to perform heavy computations [50]. The second type is when human cognition is employed to collect information, which primarily consists of labeling and identifying objects in images; in more complicated projects, training prior to data contribution is required to complete tasks, such as identifying protein structures in the Foldit project [51] or georeferencing historical images in the sMapShot project [52]. Thus, by incorporating ML techniques into citizen science, the data collection task can be partially or fully automated. As a result, considering the two key types of data collection, we define two possible approaches in which ML can be integrated in this step:

- **Machines as sensors (adapted from citizens as sensors):** The integration of ML in the first form of data collection, crowdsourcing, can be performed using AI-based tools, such as AI-based cameras. A well-known example in ecological studies is the use of camera traps to automatically capture images of species [53]. Moreover, sensors integrated with ML techniques can automatically record measurements such as noise recording [54] or air pollution [55].
- **Machine thinking (adapted from volunteer thinking):** For the second form of data collection, where cognition is involved, ML algorithms can learn to automate certain tasks, such as object detection in images/videos, which is the most common technique, or more complex tasks, such as automated prediction of protein structures using deep learning [56].

### 3.3. ML for Data Validation

Due to large amounts of data being contributed to citizen science projects, manual expert validation can be very time intensive. Thus, automatic or semiautomatic data validation can be applied by filtering potential erroneous data, considering both the contributed information and the ability and experience of participants in contributing data. Two types of potential automatic validation approaches can be the following:

- **Automatic data quality assurance:** The static comparison of the contributed data with reference datasets has been used in biodiversity citizen science projects to perform automated filtering of unusual observations [45]. However, rather than comparing the submitted data with the historical records, the ML algorithms could be used to perform real-time validation and confirmation of the newly contributed data. For example, species distribution models can be used to validate the spatial accuracy of biodiversity observations, or a CNN algorithm can be used to validate images labeled by the participants.
- **Classification of participant’s level of expertise:** The level of expertise and experience in contribution varies among participants in citizen science projects. For example, in biodiversity monitoring projects such as eBird [57] or iNaturalist, some participants contribute observations casually, while others are very involved and experienced and may even be considered as expert volunteers not only to contribute data but also to verify others’ observations [58]. Thus, the contributors’ previous records can be used

in ML algorithms to classify the participants (e.g., by assigning them scores based on their level of expertise), and the newly contributed data can be validated based on the classification of the participants' levels of expertise.

Figure 2 illustrates a taxonomy of possible combinations of ML and citizen science, which is classified according to the citizen science steps, including the three discussed categories of engagement, data collection, and data validation. Some of these ML integrations have already been applied in current citizen science projects, such as the automatic species identification or the classification of observers' levels of expertise in eBird, which will be explored in greater detail in the section on use cases. Nevertheless, there are some other potential impacts that, to the best of our knowledge, are not being applied in present projects, notably in terms of the role of ML in engaging participants through user profiles and recommendation systems. The following section presents and categorizes the use cases, taking into account the potential impacts of ML on citizen science stated in this section.

#### 4. Use Cases

In this section, we present some of the use cases in which ML and citizen science are combined, with the goal of developing a typology of such projects based on the AI and ML applications outlined in Section 2 and the impacts of ML on citizen science tasks outlined in Section 3. We begin by categorizing the use cases based on the field of science and then present the most commonly used approaches in each category. The categorization of the use cases is shown in Table 1.

Environmental science: The most common approach in environmental studies is training ML algorithms using the images/videos labeled by citizen scientists to automate species identification and/or classification. Some of the common applied methods are as follows:

- Camera trap projects: when it comes to the combination of ML and citizen science in biodiversity research, one of the most common approaches is the use of camera traps, where cameras are installed in nature to take photos of species, and the photos are then labeled by citizen scientists to feed and train ML algorithms [11,59]. Citizen scientists may, depending on the project, be involved in only one or all the activities of camera placement, submission of images, and labeling and classification of images/videos from camera traps [59]. MammalWeb [60], eMammal [61], and WildBook [62] are three examples of projects focused on camera traps data, and depending on the projects' goals, they invite volunteers to either collect or classify images (Table 1). The use of contributed images to train CNN algorithms for automatic wildlife identification can result in the implementation of software packages such as the R package MLWIC (Machine Learning for Wildlife Image Classification) [63], which can be useful for environmental studies, particularly for ecologists. Another approach of integrating human and machine intelligence in camera trap projects is to invite volunteers to observe species images and confirm machine predicted labels in each image [11]. This approach helps to balance the time required for labeling images while maintaining high quality classification, and human intelligence is used for verification and identifying more challenging species that are difficult for machines to classify.
- Species identification based on images and metadata: the majority of species identification projects use only images to train ML algorithms [64]. However, the identification of some species only with images and in the absence of other metadata is very complex both for humans and machines, and only human experts are able to distinguish among various images. Including metadata such as the spatial and temporal distribution or the ability of observers to identify species can increase ML predictive performance and provide more confidence in species identification. One example in this case is a study performed by Terry et al. [5] to identify ladybirds using both images and metadata such as location, date, and observer's expertise (Table 1). Another example is the eBird project [65], where a probabilistic model has been developed to classify observers as *experts* and *novices*, taking into account their experience in making contributions

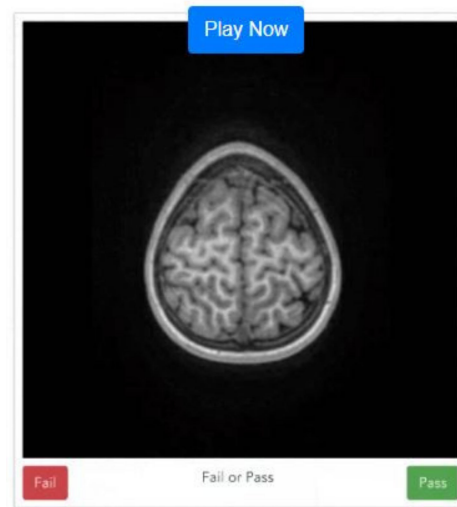


(Table 1). Another project, BeeWatch, invites citizen scientists to identify bumblebee species in images [66], and it employs natural language generation (NLG) to provide volunteers with real-time feedback (Table 1). Experiments conducted by the Bee-Watch researchers with project participants revealed that the automatically generated feedback improved the participants' learning and increased their engagement [66].

- Marine life identification: unlike other species, marine life identification by combining ML and citizen science has rarely been discussed [67]. In an article by Langenkämper et al. [67], the authors focused on combining ML and citizen science in annotation of marine life images. Citizen scientists are requested to annotate the images (digitize a bounding box around the species in the image); however, there is a possibility that volunteers may miss identifying the species (false negative), annotate a species which is not present in the image (false positive), or place the bounding box incorrectly. Despite all of the possible annotation errors, the authors conclude that merging citizen science with ML in marine life studies has considerable promise, providing that citizen scientists receive sufficient training prior to image annotation (Table 1).
- Automatic wildlife counts from aerial images: estimating wildlife abundance is an important aspect of biodiversity conservation studies. One approach is to count the species in aerial images. However, if done entirely manually, this is an extremely time consuming and labor-intensive process. A study focused on the counts of wildebeests in aerial images [68] has illustrated promising results in obtaining accurate counts by combining citizen science and deep learning (Table 1). In this study, the counting is done by both citizen scientists and machines (a trained CNN algorithm), and while the results indicate that the machine performance is faster and more accurate than the human, the authors state that the citizen scientists' contributions are essential in providing training data to feed the algorithm.

Neuroscience: similar to environmental studies and species identification tasks, citizen scientists' input can be very valuable in amplifying the gold standard data generated by neuroscience experts. In [26], an approach is proposed to amplify expert-labeled MRI (magnetic resonance imaging) images using citizen science and deep learning. This approach involves three main steps. First, the experts label a collection of MRI images. Second, to amplify the labels, a web application called BrainDr is implemented that presents a 2D brain slice to citizen scientists, and they are required to pass or fail the image taking into account its quality (check Figure 3). Finally, in the third step, a deep learning algorithm is used to verify the quality of the citizen science labels compared to the expert-labeled MRI images. Once the high-quality data are available, they are used to train a CNN algorithm to automate labeling the MRI images.

Astronomy: the involvement of the general public in online astronomy projects started in 2008 with the first release of the Galaxy Zoo project [69]. Traditionally, the classification of galaxy images in Galaxy Zoo was done by citizen scientists, but with advances in ML, the classification task was automated using amateurs and expert labels as input training data [70]. The Milky Way project is another well-known project in this field, with the goal of involving volunteers in identifying bubbles in images collected from space telescopes [71], and to automate the identification, the volunteers' labels were then used to train a random forest algorithm called Brut [72]. The authors mentioned that the combination of ML and citizen science in astrophysical image classification has opened a new path towards obtaining large scale classified datasets, which would have been more complex to achieve if each of these fields (citizen science and ML) were applied separately.



**Figure 3.** Screenshot from Braindr application [73] where citizen scientists are required to label the MRI images by selecting pass or fail.

**Table 1.** Example of use cases of combination of machine learning and citizen science.

Science Field	Use Case Example	Impact on Citizen Science Task	Machine Learning Technique	Brief Objective
	Wildlife species identification using camera traps [11,59–62]	Automate data collection: <ul style="list-style-type: none"> <li>• Automatic identification of species in images</li> <li>• Automatic photo capturing using AI-based cameras</li> </ul> Automate data validation: <ul style="list-style-type: none"> <li>• Real-time validation of new labeled images</li> </ul>	Supervised learning; computer vision and the use of CNN	<ul style="list-style-type: none"> <li>• The labeled images by citizen scientists are used to feed and train CNN algorithms to automate wildlife identification in images.</li> <li>• Volunteers can contribute other types of information besides images, such as species habitat, or they can focus on more challenging tasks, such as rare species identification</li> </ul>
Environmental science	Ladybird identification based on images and metadata [5]	Automate data validation: <ul style="list-style-type: none"> <li>• Auto-filtering of new observations</li> <li>• Auto-identification of species based on images and metadata</li> </ul>	Supervised learning: NN for metadata only, CNN (use of transfer learning) for image only and a combined model for metadata and image	Train ML algorithms to automatically identify ladybird species using images along with the structured metadata (date, location, and citizen scientists' experiences)
	eBird, use observers' expertise to verify the contribution [65,74]	Automating data validation: <ul style="list-style-type: none"> <li>• Screening of new observations based on observer's ability</li> </ul>	Using probabilistic models and automated reasoning based on observers' previous contributions	Classification of citizen scientists to experts and novices to improve identification of new species, and pass the rare species detection task to the expert observers
	BeeWatch, identification of bumblebees [66]	Generate automatic feedback to: <ul style="list-style-type: none"> <li>• Improve participants' learning to identify bumblebees</li> <li>• Increase participants' engagement</li> </ul>	Natural Language Generation (NLG)	Automatically generate feedback with the aim of improving participants' ability to identify bumblebees and increasing their engagement

Table 1. Cont.

Science Field	Use Case Example	Impact on Citizen Science Task	Machine Learning Technique	Brief Objective
	Marine life identification [67]	Automatic data collection and validation: <ul style="list-style-type: none"> <li>• Identification of marine life in images</li> <li>• Auto-detection of location (ROI) of marine species in the images</li> </ul>	Supervised learning: computer vision and CNN (use of transfer learning)	Improving marine species identification by combining citizen scientists and deep learning
	Automatic species count from aerial images [68]	Automatic data validation: <ul style="list-style-type: none"> <li>• Auto-filtering erroneous contributions caused by volunteer miscount</li> <li>• Automatic validation of species reports based on the expected density</li> </ul>	Supervised learning: computer vision and use of CNN	Combination of citizen scientists and deep learning to improve wildlife counting in aerial images for conservation purposes
Neuroscience	Braindr [26]	Automatic data collection: <ul style="list-style-type: none"> <li>• Automatic labeling of MRI images</li> </ul> Automatic data validation: <ul style="list-style-type: none"> <li>• Validation of new added labels by citizen scientists</li> </ul>	Supervised learning: computer vision and use of CNN	Amplification of expert-labeled MRI images with the help of citizen scientists, followed by the use of the amplified labels to train an algorithm to automatically replicate the labeling task of experts
	Galaxy Zoo [70]	Automatic data collection: <ul style="list-style-type: none"> <li>• Automatic classification of galaxy images</li> </ul>	Supervised learning: computer vision and use of CNN	Classify galaxy images by training an ML algorithm based on citizen scientists' input
Astronomy	Milky Way [72]	Automatic data collection: <ul style="list-style-type: none"> <li>• Auto-detection of bubbles in space telescope images</li> </ul> Automatic data validation: <ul style="list-style-type: none"> <li>• Auto-filtering of amateurs' contributions</li> </ul>	Supervised learning: Random Forest algorithm	<ul style="list-style-type: none"> <li>• Detect bubbles in space telescope images by feeding an ML algorithm using labels provided by citizen scientists and experts</li> <li>• Let the participants spend time on labeling more challenging images</li> </ul>

Table 1 illustrates that the majority of projects that combine citizen science and ML are in environmental science, which is also true for citizen science projects in general, where the number of biodiversity citizen science projects far outnumbers projects in other domains [75]. Furthermore, the table shows that, regardless of the area of science, the integration of citizen science and ML comprises primarily the use of labeled data from citizen scientists to feed ML algorithms. Typically, trained models are used to automate data collection (mostly labeling and object detection tasks in online citizen science projects) and data validation (automatic filtering and flagging the erroneous contributions). In contrast, the use of ML in citizen science to increase and sustain participation has received far less attention, with the BeeWatch project being the only one (among the studied use cases) that has directly evaluated the effects of automatic feedback on engagement.

Furthermore, while in most projects, once the model is trained, the identification/labeling tasks can be completely automated, the majority of authors argue that the role of citizen scientists does not fade away and that human cognition can be used to perform more challenging tasks, such as verifying machine predictions or identifying rare species. Given these current projects and the prospect of further possible ML and citizen science integrations, the next section discusses the benefits and risks that may arise as a result of this combination.

## 5. Benefits and Risks

Although it is discussed that the combination of ML and citizen science offers more benefits than when they are implemented in isolation [9], there are several points that need to be considered prior to the integration of ML and citizen science. In this section, we discuss the benefits of combining citizen science and ML, as well as the potential risks that can arise if ML is not used cautiously in citizen science projects. The benefits and risks of ML and citizen science integration are discussed in the scope of engagement, data quality assurance, and ethics (check Figure 4). Data collection is not listed as a separate category in the section of benefits and risks since the impacts of ML on this step are integrated into the categories of engagement and data quality.

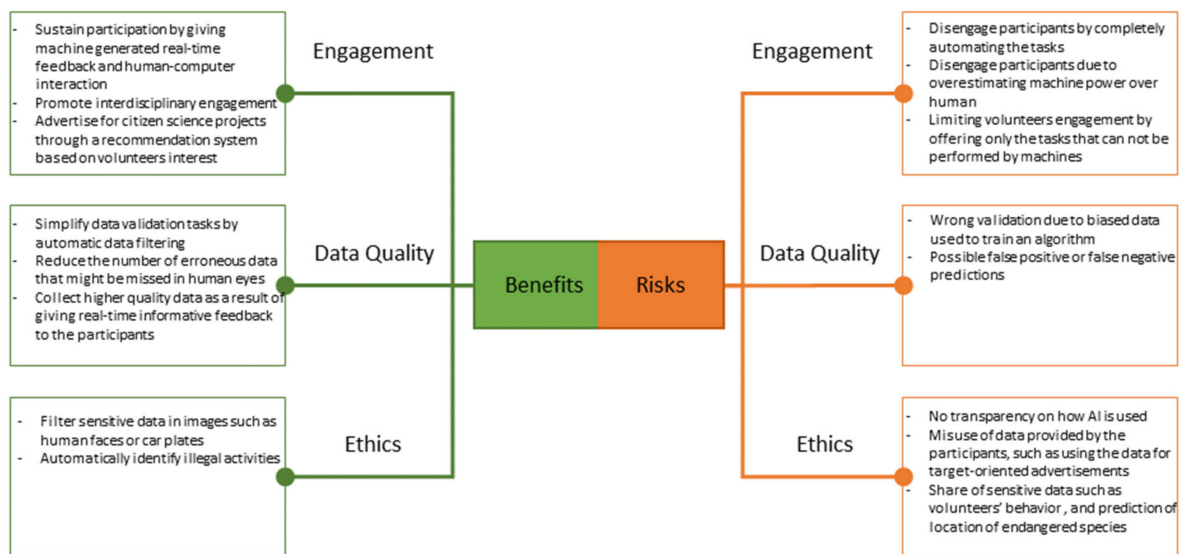


Figure 4. Benefits and risks of combining citizen science and machine learning.

### 5.1. Engagement

- Benefits:** As mentioned earlier, one of the benefits of AI for community building in citizen science projects is to encourage engagement by targeting the potential volunteers through social media. Another important factor in citizen science is the impact of the interaction with and feedback to the participants on the basis of their contributions [76,77]. Thus, the use of ML in citizen science in providing automated feedback to the participants might promote engagement through human–computer interaction and result in sustaining participation. Furthermore, the intelligently generated feedback can provide participants with useful knowledge about the research subject, allowing them to learn while contributing, which can be another factor in increasing participation (e.g., BeeWatch project). Another potential benefit of combining ML and citizen science is that it encourages interdisciplinary engagement among volunteers and researchers, which can lead to collaborations from several scientific fields [9]. Finally, automating certain simple tasks allows volunteers to concentrate on more complicated ones, such as identifying common species from camera trap images using CNN and leaving the identification of the unusual species to volunteers. However, there is another side to the task automation, which is discussed in the risk section.
- Risks:** The use of ML in citizen science could result in the automation of most tasks, which may demotivate participants because they are fully or partially being replaced by machines. As previously mentioned in the use cases, in most projects, citizen science data is used to train ML algorithms, and then the tasks can be performed entirely by machines, effectively replacing humans. While it has been mentioned that in the case of task automation, citizen scientists would then concentrate on more challenging tasks, some participants would like to contribute to citizen science projects to fill

their spare time with activities that make them feel good, such as helping science or spending time in nature (see [36]), which are not inherently challenging. For example, in the sMapShot project [52] (a citizen science project for georeferencing historical images), there is strong competition among participants of higher age groups, and the incentive system plays an important role in motivating them; therefore, if the computer performs the task more efficiently, motivation is expected to drop, and thus participation will decline. One solution is that, considering all activity levels among participants, participants are allowed to contribute with their task of interest even if the task can be fully automated by machines, and thus the contributions can be helpful in retraining the algorithms to have a better performance. Another recommendation is to incorporate new forms of contributions to fill in the gap caused by automated tasks. Furthermore, another potential risk is the overestimation of AI power in citizen science projects, such as trusting model predictions over expert volunteers, which could result in disengaging the participants [8].

### 5.2. Data Quality

- **Benefits:** The use of ML in citizen science will speed up the process of big data validation, reducing the workload of manual data quality assurance for experts [46,47]. Prescreening and filtering data (for example, removing empty images or low-quality images in camera trap projects), flagging erroneous observations, and submitting only flagged observations for expert verification will save a lot of time and allow the experts to concentrate on the scientific aspects of the project rather than the manual filtering of all data. Furthermore, the generation of real-time informative and user-centered feedback for participants with information about their contributions will improve the participants' knowledge on the subject, their proficiency, and, as a result, the quality of data they contribute over time. Another finding from the BeeWatch project concerning the impact of feedback on volunteers was that NLG feedback resulted in increased learning, and the identification accuracy was higher for those who received informative feedback than for those who only received confirmation of correct identification [66].
- **Risks:** Although the benefits of automatic filtering and validation have been discussed, the efficiency and reliability of automated validation and feedback are highly dependent on the data used to train the ML algorithms. For example, if the training data are biased in some way, such as spatially or temporally, the automated data validation based on the trained model is also biased and could provide participants and experts with misleading information [9]. In addition to bias in the data, it is critical that the data used to train the model are of a gold standard and validated by experts, since the trained model will be used to verify new data, and if the input data are uncertain, the model will predict false detections [9], such as failing to identify a species, in the case of a false negative, or incorrectly detecting an abnormal shape in an MRI image, in the case of a false positive. It is important to keep in mind that machine intelligence should not be overestimated in comparison to human intelligence. In other words, when participants receive machine-generated feedback on their contribution, the decision to either modify or retain the contribution should be made by the participants, and human experts will make the final confirmation in such cases. It is also necessary to note that when a model is trained on data from a specific region, it cannot necessarily be applicable in other areas, and doing so can result in misevaluation and the generation of misleading information. Furthermore, training algorithms for small datasets (such as rare species, see [12]) or multitype datasets (such as a mix of images and metadata, see [5]) and learning how to tune the parameters of the algorithms to achieve the desired performance are hard challenges that must be considered prior to performing automated data validation in citizen science projects.



### 5.3. Ethics

- **Benefits:** The use of machine learning (ML) can be advantageous in filtering sensitive information from citizen science data, such as human faces or license plates in images. Furthermore, ML can be used to detect illegal actions, such as illegal animal trades, by sentiment analysis using information posted on social media platforms such as Twitter [78].
- **Risks:** One major concern of integrating ML in citizen science is the use of data collected from participants for other commercial reasons, which may go against the participants wishes and result in their disengagement from the project. Thus, it is critical to be transparent and communicate effectively with participants on how their inputs are being used in the algorithms, rather than simply creating a black box project in which the participants function is limited to producing data and feeding the algorithms [8–10]. As discussed in [8], technology giants like Google and Facebook offer target-oriented advertisement services by selling personal information, which can be a danger for the future of AI-based services used in citizen science projects, as it may lead to a lack of confidence on the part of participants to freely share their contributions and personal information. Another ethical issue that may emerge from ML-based citizen science projects is the sharing of sensitive data that may be deceptive or result in geoprivacy violations, such as predicting the position of endangered species or predicting participant activity based on the history of their contributions.

## 6. Future Challenges and Conclusions

Despite the existing projects and articles on the integration of ML and citizen science, this topic is still at its initial steps and requires further research discussing other benefits and risks and even proposing other use cases that are different from those that have been applied. In addition to this, there are potential challenges to and ideas about this subject that can be seen as future extensions of this integration, some of which can be performed in the near future of citizen science, and others requiring more time and investigation before being implemented in practice. The following are some potential challenges and future ideas:

- (1) One potential challenge is to explore the integration of ML in biodiversity citizen science projects for rare species identification, for instance, by using approaches such as few-shot learning [79]. In contrast to common ML algorithms, few-shot learning requires a very minimal amount of data to train the model, and it is primarily utilized in computer vision [80], of which a particular case is one-shot learning for face recognition [81].
- (2) The focus of the use of ML in citizen science is currently more on automatic identification and less on user engagement; thus, exploring the use of ML in increasing engagement and sustaining participation remains an area for future investigation. For instance, one potential approach to be explored is the use of gamified AI in citizen science towards attracting more volunteers as well as sustaining participation [10,82].
- (3) While the impact of machine-generated feedback on sustaining participation is discussed, one possible future challenge is to determine whether the generation of feedback that simulates more human responses, rather than repetitive generated feedback, can have an impact on increasing engagement.
- (4) Training participants has been shown in studies to improve data quality; however, providing training is not always simple and requires both human and financial resources. A possible suggestion will be to use AI to provide training prior to data collection; although this has been achieved in the case of feedback (for example, in the BeeWatch project [66]), AI can be used to provide training in a variety of ways, such as through interactive courses entirely managed by AI.
- (5) Participants are more motivated to contribute to a project if there have been prior contributions or if there are other participants for the sake of competition; however, large numbers of contributions will make participants feel less motivated and assume

they have little to contribute to the project. One theory is that people in older age groups can become demotivated if there are too many contributions. One role of AI may be to consider user demographics and, as a result, balance how much data each user can visualize.

Furthermore, citizen science data are primarily based on the collection of images/videos or textual data, as seen in the use cases, but with emerging technology, the types of data collected can be extended. For example, some of the most recent smartphones support sensors that acquire LiDAR (light detection and ranging) data, and while this is currently a device-specific feature, given the rapid pace of technological development, we would expect it to be included in many future smartphones. Thus, LiDAR data can be a potential data type obtained in citizen science projects, and although some studies have been performed to identify objects from point clouds using deep learning [83,84], applying such techniques to LiDAR data collected by citizen scientists is a very interesting challenge towards the combination of ML and citizen science.

This review and other recent articles on the integration of AI and citizen science indicate that this combination demonstrates considerable potential for both fields. However, there are some consequences to this as well, as advancements in AI and the superior power of computers, in some cases better than humans, raise the possibility of completely replacing humans in citizen science projects. Nevertheless, there are certain tasks that cannot be performed without human input, such as activities that involve imagination, critical thinking, and communication skills. Furthermore, when combining ML and citizen science, it is critical that the primary goal of citizen science, engaging the general public in scientific projects and knowledge sharing with the public, does not fade away as a result of giving machines too much control. Furthermore, it is critical to apply transparency to the project and effectively communicate with volunteers about how the ML is being integrated and how the ML algorithms are using participants' input. Finally, prior to integrating ML in citizen science, the possible risks and benefits must be thoroughly investigated to determine which one has more weight, as well as to understand how to mitigate risks and maximize benefits from ML integration in the project at all levels, from user engagement to data quality assurance. Aside from the aforementioned concerns, a general aspect to consider is that, while incorporating AI into scientific research can be highly beneficial, it is essential to consider the context in which it is employed. For example, if AI is integrated into education, it is important to keep in mind that it does not prevent students from thinking by providing auto-responses to questions, such as the automatic identification of a vegetation type for an environmental student, which may result in preventing the student from learning the various landcover characteristics.

A potential extension of this review article will be to look for future AI-based citizen science projects and investigate their effect on each step of citizen science, as well as to elaborate on how the above listed challenges can be successfully implemented. Another potential extension would be to conduct analyses to quantify the risks and benefits discussed here. For example, one approach could be to evaluate the impact of real-time validation and feedback to participants by using indices to measure their engagement with the project, as well as by evaluating the quality of their contribution as a result of learning from the real-time feedback. We have developed a biodiversity citizen science project with the goal of collecting bird observations and using ML techniques to perform automatic data validation based on the location and time of observations. In this project, we provide real-time feedback to volunteers, for example, on bird species habitat characteristics [47]. As a follow-up to this review, we intend to analyze volunteers' behavior and explain the findings in the context of the risks and benefits addressed in this article.

**Author Contributions:** Conceptualization, M.L. and J.I.; literature review and investigation, M.L.; writing—original draft preparation, M.L.; writing—review and editing, M.L, J.I., and M.A.B.; supervision, J.I. and M.A.B.; All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Shoham, Y.; Perrault, R.; Brynjolfsson, E.; Openai, J.C.; Manyika, J.; Niebles, J.C.; Lyons, T.; Etchemendy, J.; Grosz, B.; Bauer, Z. *The AI Index 2018 Annual Report*; AI Index Steering Committee, Human-Centered AI Initiative, Stanford University: Stanford, CA, USA, 2018.
- Shinde, P.P.; Shah, S. A Review of Machine Learning and Deep Learning Applications. In Proceedings of the 2018 4th International Conference on Computing, Communication Control and Automation, ICCUBEA, Pune, India, 16–18 August 2018; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2018.
- Popenici, S.A.D.; Kerr, S. Exploring the impact of artificial intelligence on teaching and learning in higher education. *Res. Pract. Technol. Enhanc. Learn.* **2017**, *12*, 22. [[CrossRef](#)]
- Rzanny, M.; Seeland, M.; Wäldchen, J.; Mäder, P. Acquiring and preprocessing leaf images for automated plant identification: Understanding the tradeoff between effort and information gain. *Plant Methods* **2017**, *13*, 97. [[CrossRef](#)]
- Terry, J.C.D.; Roy, H.E.; August, T.A. Thinking like a naturalist: Enhancing computer vision of citizen science images by harnessing contextual data. *Methods Ecol. Evol.* **2020**, *11*, 303–315. [[CrossRef](#)]
- Hecker, S.; Bonney, R.; Haklay, M.; Hölker, F.; Hofer, H.; Goebel, C.; Gold, M.; Makuch, Z.; Ponti, M.; Richter, A.; et al. Innovation in Citizen Science—Perspectives on Science-Policy Advances. *Citiz. Sci. Theory Pract.* **2018**, *3*, 4. [[CrossRef](#)]
- Wright, D.E.; Fortson, L.; Lintott, C.; Laraia, M.; Walmsley, M. Help Me to Help You. *ACM Trans. Soc. Comput.* **2019**, *2*, 1–20. [[CrossRef](#)]
- Ceccaroni, L.; Bibby, J.; Roger, E.; Flemons, P.; Michael, K.; Fagan, L.; Oliver, J.L. Opportunities and Risks for Citizen Science in the Age of Artificial Intelligence. *Citiz. Sci. Theory Pract.* **2019**, *4*, 29. [[CrossRef](#)]
- McClure, E.C.; Sievers, M.; Brown, C.J.; Buelow, C.A.; Ditria, E.M.; Hayes, M.A.; Pearson, R.M.; Tulloch, V.J.D.; Unsworth, R.K.F.; Connolly, R.M. Artificial Intelligence Meets Citizen Science to Supercharge Ecological Monitoring. *Patterns* **2020**, *1*, 100109. [[CrossRef](#)] [[PubMed](#)]
- Franzen, M.; Kloetzer, L.; Ponti, M.; Trojan, J.; Vicens, J. Machine Learning in Citizen Science: Promises and Implications. In *The Science of Citizen Science*; Springer: Cham, Switzerland, 2021.
- Willi, M.; Pitman, R.T.; Cardoso, A.W.; Locke, C.; Swanson, A.; Boyer, A.; Veldhuis, M.; Fortson, L. Identifying animal species in camera trap images using deep learning and citizen science. *Methods Ecol. Evol.* **2019**, *10*, 80–91. [[CrossRef](#)]
- Norouzzadeh, M.S.; Nguyen, A.; Kosmala, M.; Swanson, A.; Palmer, M.S.; Packer, C.; Clune, J. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E5716–E5725. [[CrossRef](#)] [[PubMed](#)]
- iNaturalist. Available online: <https://www.inaturalist.org/> (accessed on 20 May 2021).
- Ueda, K. iNaturalist. Available online: <https://www.inaturalist.org/blog/31806-a-new-vision-model> (accessed on 26 May 2021).
- Horn, G.; Van Aodha, O.; Mac Song, Y.; Cui, Y.; Sun, C.; Shepard, A.; Adam, H.; Perona, P.; Belongie, S. The iNaturalist Species Classification and Detection Dataset. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8769–8778. [[CrossRef](#)]
- Géron, A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*; O'Reilly Media: Sebastopol, CA, USA, 2019.
- Paeglis, A.; Strumfs, B.; Mezale, D.; Fridrihsone, I. A Review on Machine Learning and Deep Learning Techniques Applied to Liquid Biopsy. In *Liquid Biopsy*; IntechOpen: London, UK, 2019.
- Borji, A.; Itti, L. Human vs. computer in scene and object recognition. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 113–120. [[CrossRef](#)]
- Saponara, S.; Elhanashi, A.; Gagliardi, A. Implementing a real-time, AI-based, people detection and social distancing measuring system for Covid-19. *J. Real-Time Image Process.* **2021**, 1–11. [[CrossRef](#)]
- Pl@ntNet. Available online: <https://identify.plantnet.org/> (accessed on 28 May 2021).
- Chowdhury, G.G. Natural language processing. *Annu. Rev. Inf. Sci. Technol.* **2003**, *37*, 51–89. [[CrossRef](#)]
- DeepL. Available online: <https://www.deepl.com/translator> (accessed on 5 July 2021).
- Stowell, D.; Petrusková, T.; Šálek, M.; Linhart, P. Automatic acoustic identification of individuals in multiple species: Improving identification across recording conditions. *J. R. Soc. Interface* **2019**, *16*, 20180940. [[CrossRef](#)] [[PubMed](#)]
- BirdNet. Available online: <https://birdnet.cornell.edu/> (accessed on 26 May 2021).
- Robinson, A.J.A.; Voronkov, A. *Handbook of Automated Reasoning*; Elsevier: Amsterdam, The Netherlands, 2001.
- Keshavan, A.; Yeatman, J.D.; Rokem, A. Combining citizen science and deep learning to amplify expertise in neuroimaging. *Front. Neuroinform.* **2019**, *13*, 29. [[CrossRef](#)] [[PubMed](#)]
- Joppa, L.N. The Case for Technology Investments in the Environment. *Nature* **2017**, *552*, 325–328. [[CrossRef](#)] [[PubMed](#)]

28. Mac Aodha, O.; Gibb, R.; Barlow, K.E.; Browning, E.; Firman, M.; Freeman, R.; Harder, B.; Kinsey, L.; Mead, G.R.; Newson, S.E.; et al. Bat detective—Deep learning tools for bat acoustic signal detection. *PLoS Comput. Biol.* **2018**, *14*, e1005995. [[CrossRef](#)]
29. Parham, J.; Stewart, C.; Crall, J.; Rubenstein, D.; Holmberg, J.; Berger-Wolf, T. An Animal Detection Pipeline for Identification. In Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision, WACV, Lake Tahoe, NV, USA, 12–15 March 2018; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2018; pp. 1075–1083.
30. Deng, D.P.; Chuang, T.R.; Shao, K.T.; Mai, G.S.; Lin, T.E.; Lemmens, R.; Hsu, C.H.; Lin, H.H.; Kraak, M.J. Using social media for collaborative species identification and occurrence: Issues, methods, and tools. In Proceedings of the 1st ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information, Redondo Beach, CA, USA, 6 November 2012; ACM Press: New York, NY, USA, 2012; pp. 22–29.
31. Joshi, S.; Randall, N.; Chiplunkar, S.; Wattimena, T.; Stavrianakis, K. ‘We’—A Robotic System to Extend Social Impact of Community Gardens. In Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction, Chicago, IL, USA, 5–8 March 2018; IEEE Computer Society: Washington, DC, USA, 2018; pp. 349–350.
32. Bonney, R.; Ballard, H.; Jordan, R.; McCallie, E.; Phillips, T.; Shirk, J.; Wilderman, C. *Public Participation in Scientific Research: Defining the Field and Assessing Its Potential for Informal Science Education*; A CAISE Inquiry Group Report; Center for Advancement of Informal Science Education (CAISE): Washington, DC, USA, 2009.
33. Bonney, R.; Cooper, C.B.; Dickinson, J.; Kelling, S.; Phillips, T.; Rosenberg, K.V.; Shirk, J. Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy. *Bioscience* **2009**, *59*, 977–984. [[CrossRef](#)]
34. CitizenScience.gov. Basic Steps for Your Project Planning. Available online: <https://www.citizen-science.gov/toolkit/howto> (accessed on 20 May 2021).
35. Rotman, D.; Hammock, J.; Preece, J.; Hansen, D.; Boston, C. Motivations Affecting Initial and Long-Term Participation in Citizen Science Projects in Three Countries. In *iConference 2014 Proceedings*; iSchools: Grandville, MI, USA, 2014.
36. Lotfian, M.; Ingensand, J.; Brovelli, M.A. A Framework for Classifying Participant Motivation that Considers the Typology of Citizen Science Projects. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 704. [[CrossRef](#)]
37. Antoniou, V.; Fonte, C.; Minghini, M.; See, L.; Skopeliti, A. Developing a Guidance Tool for VGI Contributors. 2016. Available online: <https://core.ac.uk/download/pdf/80335283.pdf> (accessed on 30 May 2021).
38. Devaraj, A.; Murthy, D.; Dontula, A. Machine-learning methods for identifying social media-based requests for urgent help during hurricanes. *Int. J. Disaster Risk Reduct.* **2020**, *51*, 101757. [[CrossRef](#)]
39. Park, J.; Krishna, R.; Khadpe, P.; Fei-Fei, L.; Bernstein, M. AI-Based Request Augmentation to Increase Crowdsourcing Participation. In Proceedings of the AAAI Conference on Human Computation and Crowdsourcing, Stevenson, WA, USA, 28–30 October 2019.
40. Kanoje, S.; Mukhopadhyay, D.; Girase, S. User Profiling for University Recommender System Using Automatic Information Retrieval. *Phys. Procedia* **2016**, *78*, 5–12. [[CrossRef](#)]
41. Barnard, T.C. User Profiling Using Machine Learning. Ph.D. Thesis, University of Southampton, Southampton, UK, 2012.
42. Schade, S.; Manzoni, M.; Fullerton, K.T. *Activity Report on Citizen Science—Discoveries from a Five Year Journey*; Publications Office of the European Union: Luxembourg, 2020.
43. Tinati, R.; Simperl, E.; Luczak-Roesch, M. To help or hinder: Real-time chat in citizen science. In Proceedings of the Eleventh International AAAI Conference on Web and Social Media; The AAAI Press: Palo Alto, CA, USA, 2017; pp. 270–279.
44. Ingensand, J.; Nappes, M.; Joost, S.; Widmer, I.; Ertz, O.; Rappo, D. The urbangene project experience from a crowdsourced mapping campaign. In Proceedings of the 2015 1st International Conference on Geographical Information Systems Theory, Applications and Management (GISTAM), Barcelona, Spain, 28–30 April 2015; pp. 178–184. [[CrossRef](#)]
45. Kelling, S.; Yu, J.; Gerbracht, J.; Wong, W.K. Emergent filters: Automated data verification in a large-scale citizen science project. In Proceedings of the 2011 IEEE Seventh International Conference on e-Science Workshops, Stockholm, Sweden, 5–8 December 2011; pp. 20–27. [[CrossRef](#)]
46. Bonter, D.N.; Cooper, C.B. Data validation in citizen science: A case study from Project FeederWatch. *Front. Ecol. Environ.* **2012**, *10*, 305–307. [[CrossRef](#)]
47. Lotfian, M.; Ingensand, J.; Ertz, O.; Oulevay, S.; Chassin, T. Auto-filtering validation in citizen science biodiversity monitoring: A case study. *Proc. Int. Cartogr. Assoc.* **2019**, *2*, 78. [[CrossRef](#)]
48. Haklay, M. Citizen Science and Volunteered Geographic Information—Overview and typology of participation. Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice. *Crowdsour. Geogr. Knowl.* **2013**, *9789400745*, 1–396. [[CrossRef](#)]
49. Guillaume, G.; Can, A.; Petit, G.; Fortin, N.; Palominos, S.; Gauvreau, B.; Bocher, E.; Picaut, J. Noise mapping based on participative measurements. *Noise Mapp.* **2016**, *3*, 140–156. [[CrossRef](#)]
50. Yadav, P.; Charalampidis, I.; Cohen, J.; Darlington, J.; Grey, F. A Collaborative Citizen Science Platform for Real-Time Volunteer Computing and Games. *IEEE Trans. Comput. Soc. Syst.* **2018**, *5*, 9–19. [[CrossRef](#)]
51. Cooper, S.; Khatib, F.; Treuille, A.; Barbero, J.; Lee, J.; Beenen, M.; Leaver-Fay, A.; Baker, D.; Popović, Z. Foldit players: Predicting protein structures with a multiplayer online game. *Nature* **2010**, *466*, 756–760. [[CrossRef](#)]
52. Produit, T.; Ingensand, J. 3D Georeferencing of historical photos by volunteers. *Lect. Notes Geoinf. Cartogr.* **2018**, 113–128. [[CrossRef](#)]



53. Wiggers, K. Google's AI Can Identify Wildlife from Trap-Camera Footage with Up to 98.6% Accuracy. Available online: <https://venturebeat.com/2019/12/17/googles-ai-can-identify-wildlife-from-trap-camera-footage-with-up-to-98-6-accuracy/> (accessed on 30 May 2021).
54. Monti, L.; Vincenzi, M.; Mirri, S.; Pau, G.; Salomoni, P. RaveGuard: A Noise Monitoring Platform Using Low-End Microphones and Machine Learning. *Sensors* **2020**, *20*, 5583. [[CrossRef](#)] [[PubMed](#)]
55. Le, D.; Van Tham, C.K. Machine learning (ML)-based air quality monitoring using vehicular sensor networks. In Proceedings of the 2017 IEEE 23rd International Conference on Parallel and Distributed Systems (ICPADS), Shenzhen, China, 15–17 December 2017; IEEE Computer Society: Washington, DC, USA, 2018; pp. 65–72.
56. Panou, D.; Reczko, M. DeepFoldit-A Deep Reinforcement Learning Neural Network Folding Proteins. *arXiv* **2020**, arXiv:2011.03442.
57. EBird. Available online: <https://ebird.org/home> (accessed on 28 May 2021).
58. Kelling, S.; Johnston, A.; Hochachka, W.M.; Iliff, M.; Fink, D.; Gerbracht, J.; Lagoze, C.; La Sorte, F.A.; Moore, T.; Wiggins, A.; et al. Can observation skills of citizen scientists be estimated using species accumulation curves? *PLoS ONE* **2015**, *10*, e139600. [[CrossRef](#)]
59. Green, S.E.; Rees, J.P.; Stephens, P.A.; Hill, R.A.; Giordano, A.J. Innovations in Camera Trapping Technology and Approaches: The Integration of Citizen Science and Artificial Intelligence. *Animals* **2020**, *10*, 132. [[CrossRef](#)] [[PubMed](#)]
60. Hsing, P.Y.; Bradley, S.; Kent, V.T.; Hill, R.A.; Smith, G.C.; Whittingham, M.J.; Cokill, J.; Crawley, D.; Stephens, P.A. Economical crowdsourcing for camera trap image classification. *Remote Sens. Ecol. Conserv.* **2018**, *4*, 361–374. [[CrossRef](#)]
61. McShea, W.J.; Forrester, T.; Costello, R.; He, Z.; Kays, R. Volunteer-run cameras as distributed sensors for macrosystem mammal research. *Landsc. Ecol.* **2016**, *31*, 55–66. [[CrossRef](#)]
62. Berger-Wolf, T.Y.; Rubenstein, D.I.; Stewart, C.V.; Holmberg, J.A.; Parham, J.; Menon, S. Wildbook: Crowdsourcing, computer vision, and data science for conservation. *arXiv* **2017**, arXiv:1710.08880.
63. Tabak, M.A.; Norouzzadeh, M.S.; Wolfson, D.W.; Sweeney, S.J.; Vercauteren, K.C.; Snow, N.P.; Halseth, J.M.; Di Salvo, P.A.; Lewis, J.S.; White, M.D.; et al. Machine learning to classify animal species in camera trap images: Applications in ecology. *Methods Ecol. Evol.* **2019**, *10*, 585–590. [[CrossRef](#)]
64. Weinstein, B.G. A computer vision for animal ecology. *J. Anim. Ecol.* **2018**, *87*, 533–545. [[CrossRef](#)]
65. Yu, J.; Wong, W.K.; Hutchinson, R.A. Modeling experts and novices in citizen science data for species distribution modeling. In Proceedings of the 2010 IEEE International Conference on Data Mining, Sydney, Australia, 13–17 December 2010; pp. 1157–1162. [[CrossRef](#)]
66. Van der Wal, R.; Sharma, N.; Mellish, C.; Robinson, A.; Siddharthan, A. The role of automated feedback in training and retaining biological recorders for citizen science. *Conserv. Biol.* **2016**, *30*, 550–561. [[CrossRef](#)] [[PubMed](#)]
67. Langenkämper, D.; Simon-Lledó, E.; Hosking, B.; Jones, D.O.B.; Nattkemper, T.W. On the impact of Citizen Science-derived data quality on deep learning based classification in marine images. *PLoS ONE* **2019**, *14*, e0218086. [[CrossRef](#)]
68. Torney, C.J.; Lloyd-Jones, D.J.; Chevallier, M.; Moyer, D.C.; Maliti, H.T.; Mwita, M.; Kohi, E.M.; Hopcraft, G.C. A comparison of deep learning and citizen science techniques for counting wildlife in aerial survey images. *Methods Ecol. Evol.* **2019**, *10*, 779–787. [[CrossRef](#)]
69. Lintott, C.J.; Schawinski, K.; Slosar, A.; Land, K.; Bamford, S.; Thomas, D.; Raddick, M.J.; Nichol, R.C.; Szalay, A.; Andreescu, D.; et al. Galaxy Zoo: Morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. *Mon. Not. R. Astron. Soc.* **2008**, *389*, 1179–1189. [[CrossRef](#)]
70. Jimenez, M.; Torres, M.T.; John, R.; Triguero, I. Galaxy image classification based on citizen science data: A comparative study. *IEEE Access* **2020**, *8*, 47232–47246. [[CrossRef](#)]
71. Kendrew, S.; Simpson, R.; Bressert, E.; Povich, M.S.; Sherman, R.; Lintott, C.J.; Robitaille, T.P.; Schawinski, K.; Wolf-Chase, G. The milky way project: A statistical study of massive star formation associated with infrared bubbles. *Astrophys. J.* **2012**, *755*, 71. [[CrossRef](#)]
72. Beaumont, C.N.; Goodman, A.A.; Kendrew, S.; Williams, J.P.; Simpson, R. The milky way project: Leveraging citizen science and machine learning to detect interstellar bubbles. *Astrophys. J. Suppl. Ser.* **2014**, *214*, 3. [[CrossRef](#)]
73. Braindr. Available online: <https://braindr.us/> (accessed on 20 May 2021).
74. Johnston, A.; Fink, D.; Hochachka, W.M.; Kelling, S. Estimates of observer expertise improve species distributions from citizen science data. *Methods Ecol. Evol.* **2018**, *9*, 88–97. [[CrossRef](#)]
75. Pettibone, L.; Vohland, K.; Ziegler, D. Understanding the (inter)disciplinary and institutional diversity of citizen science: A survey of current practice in Germany and Austria. *PLoS ONE* **2017**, *12*, e178778. [[CrossRef](#)]
76. Tang, J.; Zhou, X.; Yu, M. Designing feedback information to encourage users' participation performances in citizen science projects. *Proc. Assoc. Inf. Sci. Technol.* **2019**, *56*, 486–490. [[CrossRef](#)]
77. Zhou, X.; Tang, J.; Zhao, Y.; Wang, T. Effects of feedback design and dispositional goal orientations on volunteer performance in citizen science projects. *Comput. Hum. Behav.* **2020**, *107*, 106266. [[CrossRef](#)]
78. Di Minin, E.; Fink, C.; Hiippala, T.; Tenkanen, H. A framework for investigating illegal wildlife trade on social media with machine learning. *Conserv. Biol.* **2019**, *33*, 210–213. [[CrossRef](#)] [[PubMed](#)]
79. Wang, Y.X.; Girshick, R.; Hebert, M.; Hariharan, B. Low-Shot Learning from Imaginary Data. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7278–7286. [[CrossRef](#)]



80. Wang, Y.; Yao, Q.; Kwok, J.T.; Ni, L.M. Generalizing from a Few Examples: A Survey on Few-shot Learning. *ACM Comput. Surv.* **2020**, *53*, 1–34. [[CrossRef](#)]
81. Chanda, S.; Gv, A.C.; Brun, A.; Hast, A.; Pal, U.; Doermann, D. Face recognition—A one-shot learning perspective. In Proceedings of the 2019 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Sorrento, Italy, 26–29 November 2019; pp. 113–119. [[CrossRef](#)]
82. Bowser, A.; Hansen, D.; He, Y.; Boston, C.; Reid, M.; Gunnell, L.; Preece, J. Using gamification to inspire new citizen science volunteers. In *Proceedings of the First International Conference on Gameful Design, Research, and Applications*; Association for Computing Machinery: New York, NY, USA, 2013; pp. 18–25. [[CrossRef](#)]
83. Wu, Y.; Wang, Y.; Zhang, S.; Ogai, H. Deep 3D Object Detection Networks Using LiDAR Data: A Review. *IEEE Sens. J.* **2021**, *21*, 1152–1171. [[CrossRef](#)]
84. Engels, G.; Aranjuelo, N.; Arganda-Carreras, I.; Nieto, M.; Otaegui, O. 3D object detection from LiDAR data using distance dependent feature extraction. In Proceedings of the 6th International Conference on Vehicle Technology and Intelligent Transport Systems, Online, 2–4 May 2020; pp. 289–300. [[CrossRef](#)]