



Gravity Spy: lessons learned and a path forward

Michael Zevin^{1,2,7,a} , Corey B. Jackson³, Zoheyr Doctor⁴, Yunan Wu⁵, Carsten Østerlund⁶, L. Clifton Johnson^{4,7}, Christopher P. L. Berry⁸, Kevin Crowston⁶, Scott B. Coughlin⁴, Vicky Kalogera^{4,9}, Sharan Banagiri⁴, Derek Davis¹⁰, Jane Glanzer¹¹, Renzhi Hao⁵, Aggelos K. Katsaggelos^{4,5}, Oli Patane¹², Jennifer Sanchez⁴, Joshua Smith¹³, Siddharth Soni¹⁴, Laura Trouille⁷, Marissa Walker¹⁵, Irina Aerith¹⁸, Wilfried Domainko¹⁹, Victor-Georges Baranowski¹⁶, Gerhard Niklasch¹⁷, Barbara Téglás²⁰

¹ Kavli Institute for Cosmological Physics, The University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA

² Enrico Fermi Institute, The University of Chicago, 933 East 56th Street, Chicago, IL 60637, USA

³ Information School, University of Wisconsin–Madison, 600 N Park Street, Madison, WI 53706, USA

⁴ Center for Interdisciplinary Exploration and Research in Astrophysics (CIERA), Northwestern University, 1800 Sherman Ave, Evanston, IL 60201, USA

⁵ Department of Electrical and Computer Engineering, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA

⁶ School of Information Studies, Syracuse University, 343 Hinds Hall, Syracuse, NY 13210, USA

⁷ Zooniverse, The Adler Planetarium, 1300 South DuSable Lake Shore Drive, Chicago, IL 60605, USA

⁸ SUPA, School of Physics and Astronomy, University of Glasgow, Kelvin Building, University Ave, Glasgow G12 8QQ, UK

⁹ Department of Physics and Astronomy, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA

¹⁰ LIGO Laboratory, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, USA

¹¹ Department of Physics and Astronomy, Louisiana State University, 202 Nicholson Hall, Baton Rouge, LA 70803, USA

¹² LIGO Hanford Observatory, 127124 N Route 10, Hanford, WA 99354, USA

¹³ The Nicholas and Lee Begovich Center for Gravitational-Wave Physics and Astronomy, California State University Fullerton, 800 N. State College Blvd, Fullerton, CA 92831, USA

¹⁴ LIGO Laboratory, Massachusetts Institute of Technology, 185 Albany Street, Cambridge, MA 02139, USA

¹⁵ Department of Physics, Computer Science and Engineering, Christopher Newport University, One Avenue of the Arts, Newport News, VA 23606, USA

¹⁶ Sorbonne Paris Nord University, 99 Avenue Jean Baptiste Clément, 93430 Villetaneuse, France

¹⁷ ConSol Software GmbH, St.-Cajetan-Strasse 43, 81669 Munich, Germany

¹⁸ Cedar Falls, Iowa, USA

¹⁹ Hermagor, Austria

²⁰ Budapest, Hungary

Received: 24 August 2023 / Accepted: 13 December 2023

© The Author(s) 2024

Abstract The Gravity Spy project aims to uncover the origins of glitches, transient bursts of noise that hamper analysis of gravitational-wave data. By using both the work of citizen-science volunteers and machine learning algorithms, the Gravity Spy project enables reliable classification of glitches. Citizen science and machine learning are intrinsically coupled within the Gravity Spy framework, with machine learning classifications providing a rapid first-pass classification of the dataset and enabling tiered volunteer training, and volunteer-based classifications verifying the machine classifications, bolstering the machine learning training set and identifying new morphological classes of glitches. These classifications are now routinely used in studies characterizing the performance of the LIGO gravitational-wave detectors. Providing the volunteers with a training framework that teaches them to classify a wide range of glitches, as well as additional tools to aid their investigations of interesting glitches, empowers them to make discoveries of new classes of glitches. This demonstrates that, when giving suitable support, volunteers can go beyond simple classification tasks to identify new features in data at a level comparable to domain experts. The Gravity Spy project is now providing volunteers with more complicated data that includes auxiliary monitors of the detector to identify the root cause of glitches.

1 Introduction

Gravitational waves (GWs), ripples in the fabric of space and time that are a key prediction of Albert Einstein's century-old theory of general relativity [1, 2], were directly observed for the first time by the Advanced Laser Interferometer Gravitational-wave Observatory (LIGO) instruments in September 2015 [3]. Since this groundbreaking discovery, the two LIGO detectors [4] and their

Irina Aerith, Wilfried Domainko, Victor-Georges Baranowski, Gerhard Niklasch, Barbara Téglás: Gravity Spy Moderator.

Michael Zevin: NASA Hubble Fellow.

^a e-mail: mzevin@adlerplanetarium.org (corresponding author)

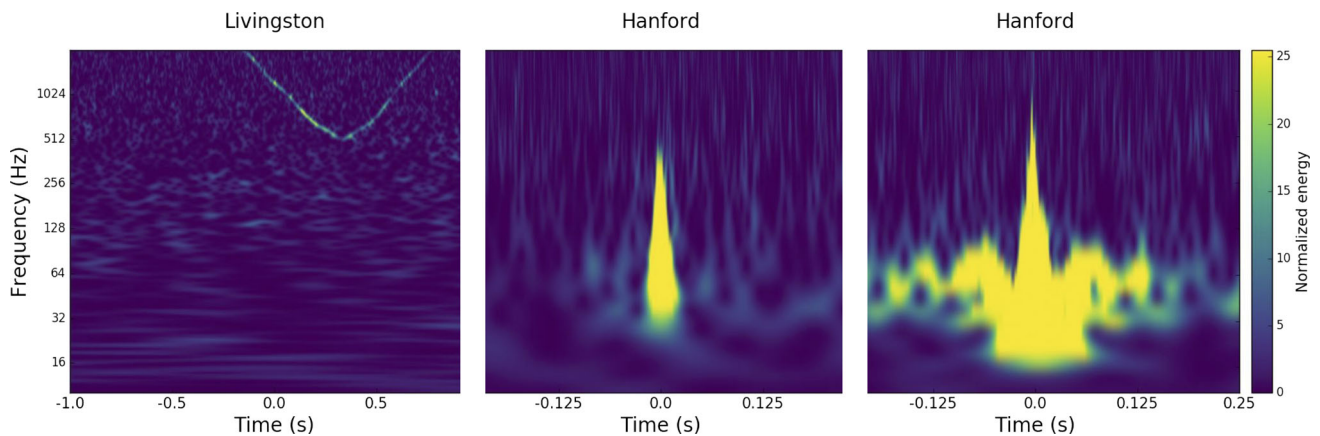


Fig. 1 Spectrograms of example glitches seen in LIGO detectors. Time is along the horizontal axis, and frequency along the vertical axis. The color denotes the energy in each time–frequency bin. These glitches exemplify a subset of glitch classes. The left panel shows a Whistle glitch, middle panel a Blip glitch, and right panel a Koi Fish glitch

partner, the Virgo detector [5] have measured nearly 100 GW candidates from coalescing black holes and neutron stars during their first three observing runs [6]. The fourth observing run (O4) of the LIGO, Virgo, and now KAGRA [7] GW detector network is under way and expected to result in hundreds more GW detections over the next 2 years [8].

To observe GWs from ground-based detectors such as LIGO and Virgo, the instruments need to be sensitive to changes in the length of the detector arms on the order of 10^{-18} m [9]. State-of-the-art instrumentation and data analysis techniques have been developed to achieve these sensitivity requirements. However, non-astrophysical noise sources, originating from both instrumental and environmental factors, affect the detectors and lessen their sensitivity to the GW universe. Of particular concern are transient, non-Gaussian noise sources known as *glitches* that appear in the detectors at a high rate [10–12]. Though some glitches have known causes, others show no obvious correlation with instrumental and environmental noise monitors, making their root causes difficult to diagnose [13, 14]. The classification and characterization of glitches are paramount to minimizing their effect on GW measurements. Classification is the crucial first step because many glitches are morphologically similar, as can be identified through their spectrograms. Figure 1 shows spectrograms of three glitches that exemplify common glitch classes seen in the LIGO detectors: Whistles, Blips and Koi Fish [15, 16]. Although these glitches can be differentiated by eye in their spectrograms (and by their sound, if played through a speaker), classification is a daunting task because of the sheer volume of glitch data that is accumulated throughout an observing run.

Such large-scale data analysis challenges are not unique to GW astronomy. Fundamental challenges and opportunities for twenty-first-century science lie in developing machinery and techniques to characterize and use expansive data. The paradigm in which individual researchers or even teams of researchers can analyze data has shifted to relying on novel methods for analysis of large-scale scientific datasets.

Machine learning (ML) techniques are heavily embedded in scientific data analysis, becoming a standard approach for analyzing large-scale data (e.g., [17, 18]); however, there are many instances where analysis by humans is still critical. One example is identifying new data classes not initially accounted for in training sets for ML algorithms. This novelty is indeed the case for GW detector characterization, as new glitches regularly appear due to environmental and instrumental changes as the detectors evolve [14, 19]. To this end, the Gravity Spy project was created to characterize glitches in GW detectors by combining computer- and human-based classification schemes [15].

At its core, Gravity Spy is a citizen-science project.¹ As a citizen-science project, it involves members of the general public in the scientific process, including formulating research questions, collecting or analyzing data, observing and recording natural phenomena, and disseminating results [20]. As Internet-enabled devices have become increasingly ubiquitous, citizen-science projects have become a feasible approach to providing access to human insights at a large scale. For example, Galaxy Zoo [21] invites volunteers to engage in the morphological classification of images of galaxies produced by the Sloan Digital Sky Survey: the project has successfully increased engagement with the public and led to novel discoveries in astronomy.²

Gravity Spy has a web-based interface through which anyone can provide analysis of LIGO glitches and interact with LIGO scientists about the state of the detectors. It is hosted on the Zooniverse citizen-science platform.³ The Zooniverse platform has fielded a workable crowdsourcing model (involving over 2.5 million people in more than four hundred projects since its inception) through which volunteers provide large-scale scientific data analysis. Zooniverse provides tools for systematically performing data analysis tasks on data collections, making it an ideal platform for Gravity Spy.

¹ Gravity Spy Project www.gravityspy.org.

² Galaxy Zoo Results www.zooniverse.org/projects/zookeeper/galaxy-zoo/about/results.

³ Zooniverse www.zooniverse.org.

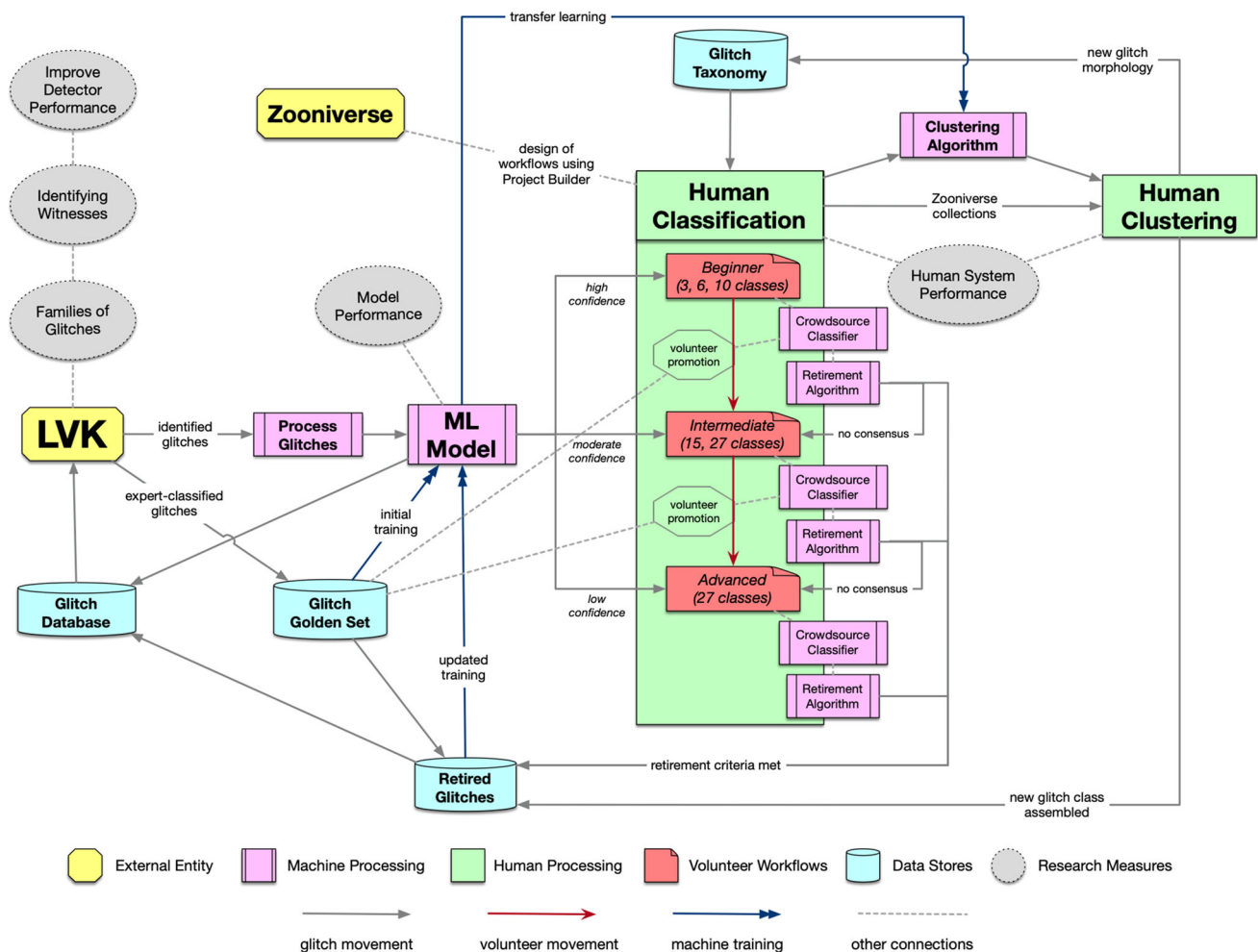


Fig. 2 Components of the interconnected Gravity Spy system. Gray arrows show the movement of glitches throughout the project, red arrows the progression of volunteers, and blue arrows the training of ML models. We note that there are multiple levels that volunteers progress through in the Beginner and Intermediate workflows with differing number of glitch classes, though the ML confidence of glitches are consistent within all Beginner workflows and within all Intermediate workflows

Building a citizen-science project using Zooniverse would not solve the specific challenges GW detectors face alone. The glitches in the detectors can change over the course of a run as GW scientists adjust the detectors or from other transient terrestrial events. Understanding and mitigating glitches requires innovative solutions, testing, and meticulous analysis of the data that only an interdisciplinary team of citizen-science volunteers, LIGO detector-characterization experts, computer scientists trained in machine learning, and social scientists could provide.

Since its inception in October 2016, Gravity Spy has accumulated over 7 million classifications of glitches by tens of thousands of Zooniverse users, significantly contributing to the characterization and identification of glitches in LIGO data. Gravity Spy also provides a testbed for fostering a symbiotic relationship between human and machine classification approaches. This connection goes far deeper than the classification task itself; the interplay between the human and machine components of the project allows for a customized training regimen for project volunteers, a more efficient and individually-tailored classification task, and the ability of volunteers to actively improve the underlying ML algorithms.

This article details activities undertaken as part of the 3-year multi-institution National Science Foundation grant (INSPIRE 1547880) to develop Gravity Spy as a prototype for the next generation of citizen-science projects. We describe the Gravity Spy project, its novel approaches to citizen science, its impact on GW detector characterization over the past half-decade, and the citizen-science techniques exclusively developed for this project. We will argue why such techniques are necessary for citizen science to remain highly relevant even as datasets grow exponentially and machine-based classification algorithms advance. Following an overview of the project, results, and lessons learned, we will look ahead to future advancements and adaptations to the Gravity Spy project, particularly focusing on the latest extension known as Gravity Spy 2.0.

2 Building Gravity Spy

The Gravity Spy project relies on an intricate interconnection between GW data analysis, citizen science, and machine learning. An overview of the main Gravity Spy system is shown in Fig. 2 and summarized as follows:

1. Glitches are identified using the Omicron pipeline [22], which identifies moments with excess power in the data stream of each LIGO detector individually, referred to as *Omicron triggers*. Omicron triggers that exceed a signal-to-noise threshold and are from times of a suitable detector state are selected for study.
2. Spectrograms like those in Fig. 1 are created for four different time durations around each trigger. Each particular trigger's collection of spectrograms in the project is referred to as a *subject*.
3. Each new subject is sent through a trained ML algorithm and assigned a probability of being an instance of one of the predetermined morphological classes in the Gravity Spy project.
4. New subjects are distributed to the Gravity Spy volunteer workflows based on the confidence score from the initial ML classification, with glitches likely to be of a known glitch class being provided to new volunteers and less certain glitches to more advanced volunteers.
5. Volunteers morphologically classify subjects until the combined machine and human classification score reaches a predetermined threshold, at which point that particular subject is retired from the project (i.e., removed from the classification workflows). If no consensus is reached after a set number of classifications, the glitch is migrated to a more advanced workflow.
6. Volunteers advance through the project and gain the ability to access workflows with additional classes of glitches as they correctly classify *golden* subjects (i.e., subjects that experts have already classified).⁴
7. Volunteers in the most advanced workflows examine glitches that neither human nor machine classification has been able to confidently identify, in order to look for possible new classes of glitches that can be proposed for follow-up by LIGO scientists.
8. All machine learning, volunteer, and combined classification results for active and retired subjects are provided to LIGO–Virgo–KAGRA (LVK) scientists to aid glitch studies. Retired images are actively added to the ML training set, which can be retrained at a predetermined latency to improve its initial classification ability.⁵

The following subsections provide more details about the building and performance of each component of the Gravity Spy system.

2.1 The dataset: transient noise in GW detectors

As summarized above, Gravity Spy classifications operate on spectrograms of GW data around the times of Omicron triggers [22]. The Omicron software relies on the Q -transform, an analog of the short Fourier transform [23]. Times of excess power are identified through the Q -transform, and the Q -transformed data are represented as spectrogram images (also known as Q -scans, see Fig. 1).

A glitch in one of the LIGO detectors will trigger Omicron to produce spectrograms for four different time windows: 0.5 s, 1.0 s, 2.0 s, and 4.0 s [15]. The reason multiple time windows are used is so that both humans and ML models can examine glitch morphologies that occur on different characteristic timescales. These images are then exported to an ML model and for human vetting, as described in later sections of this article. The classifications from the ML and humans are then added to the Gravity Spy dataset for use by analysts.

Since the start of Advanced LIGO observations in the fall of 2015 through the end of the third observing run (O3) in spring 2020, the Gravity Spy dataset has accumulated $\gtrsim 1.4 \times 10^6$ glitch triggers from the Omicron pipeline with a signal-to-noise ratio > 7.5 , corresponding on average to a new glitch being added to the project per every minute of observing time. Compared to the $\sim 10^2$ astrophysical events observed over this timespan [6], the number of glitches that have been uploaded to Gravity Spy is $> 10^4$ times larger. As of the start of the fourth observing run of the Advanced LIGO, Advanced Virgo and KAGRA detectors, 27 morphological classes [16] including the catch-all None-of-the-Above (NOTA) and No Glitch classes are available for volunteer classification of LIGO glitches, with 23 incorporated into the ML classifier [19].

2.2 The Gravity Spy classification task

The Gravity Spy classification task on the Zooniverse user interface involves citizen-science volunteers reviewing images of glitches and classifying them into morphological categories. The different levels at which volunteers can classify glitches are called *workflows*. As volunteers learn the classification task and classify glitches in accordance with glitches pre-classified by experts, they unlock and advance through workflows that introduce more glitch classes, more options in the classification interface, and glitches that are classified with less confidence by the ML algorithm.

The classification interface for the most advanced workflow is shown in Fig. 3. The left side of the interface presents the glitch to be classified, and the right side, a list of possible classes, along with NOTA for glitches that do not match any listed classes.

⁴ The original plan for Gravity Spy to use the unclassified subjects with high-confidence ML scores in advancing volunteers through the project as to not waste volunteer classifications on glitches with already-known labels; however, this proved to be suboptimal due to the ML algorithm spuriously identifying certain glitches incorrectly with high confidence.

⁵ The active addition of retired subjects to the ML training set and model retraining has yet to be fully automated.

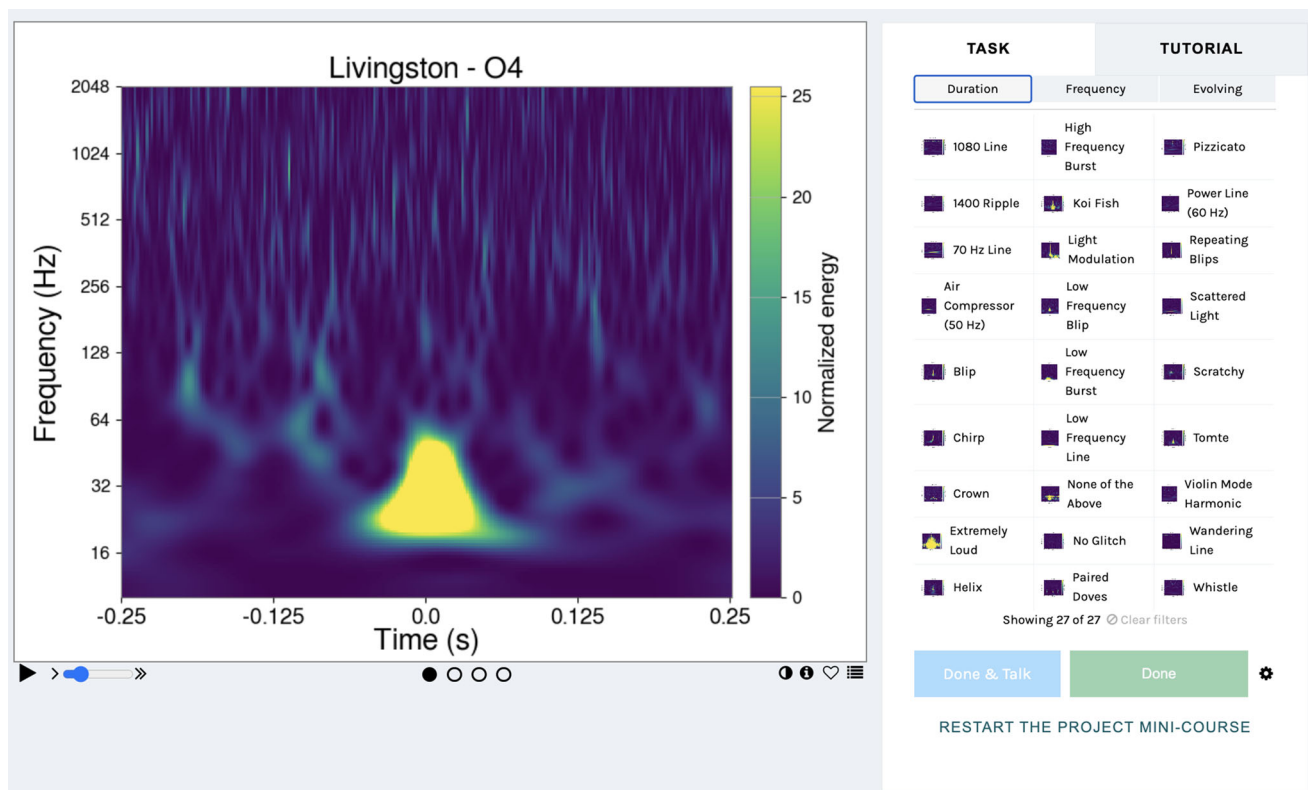


Fig. 3 The Gravity Spy classification interface for advanced workflows, named (Binary Black Hole Merger (Level 5) and Inflationary Gravitational Waves (Level 6)). These workflows allow for volunteers to pick from all of the 24 morphological classes that are currently accounted for in the Gravity Spy project

To help volunteers pick a class, a Field Guide with exemplary images and text descriptions is available. Metadata (e.g., the date on which the glitch occurred) and image filtering options are present in the interface. Volunteers can also save favorite subjects or create collections of subjects that they can view after the glitch has been classified.

Upon selecting a class, volunteers are presented with two options: Done or Done & Talk. Both options submit the response to the response database. If the subject was used to evaluate promotion (gold-standard data), feedback is provided that the volunteer agreed or disagreed with the expert evaluation of the glitch. Selecting Done & Talk directs the volunteer to a thread on the Gravity Spy Talk discussion board where they can review comments posted about the image by other volunteers or post a new comment.⁶ The Talk feature is extensively used to comment on NOTA glitches that might represent a new glitch class.

2.3 Machine learning for Gravity Spy: classification and volunteer training

ML processing is an integral part of the Gravity Spy architecture and is used for classification and volunteer training. The ML model used for initial classification is a convolutional neural network (CNN), a class of deep learning algorithms that shows exceptional performance in image recognition endeavors [24]. The detailed architecture of the CNN used in the Gravity Spy classification task and how glitch spectrograms with differing temporal durations are combined can be found in [15, 25].

A training set of $\simeq 7700$ labeled glitches across the 19 initial morphological classes was synthesized for the initial Gravity Spy launch, and this training set has been supplemented over time to include nearly 10^4 labeled glitches over 23 classes after O3 [16]. This dataset continues to be expanded as new data are being taken. The initial training set was constructed by experts: first, ~ 1000 labeled glitches were used to train a preliminary ML model (with relatively poor accuracy), and then these ML labels were vetted by experts to expand the initial training set.

Gravity Spy uses a novel approach for volunteer training and advancement that relies on ML to provide task scaffolding. The design is informed by learning theories, namely the zone of proximal development [26–28]. Specifically, new volunteers to the project progress through a series of increasingly challenging and complex workflows to develop their knowledge of the classification task and the zoo of glitch morphologies, with more advanced workflows and more difficult-to-classify glitches opening to volunteers as they proceed through the project. The ML classifications and confidence scores determine the workflow to which a particular subject is assigned. Subjects that the ML confidently classified are assigned to beginner workflows, while subjects that the ML

⁶ Gravity Spy Talk www.zooniverse.org/projects/zooniverse/gravity-spy/talk.

is uncertain about, which may thus represent new categories of glitches, are assigned to the more advanced workflows. There are multiple beginner workflows containing an increasing number of glitch classes. Glitches are assigned to the appropriate workflows so that new volunteers see glitches that are likely (but not certain) to be examples of one of the included classes. In the initial level, volunteers begin with only two options and, upon meeting performance thresholds as assessed by classification of gold-standard data, are promoted to subsequent levels in which additional glitch classes are presented. Also, instances of the glitch classes introduced to the volunteer in the previous level may be more challenging to classify, as indicated by lower ML confidence scores. At each level, volunteers have the option of NOTA to handle the case where the ML has confidently misclassified a glitch. In the most advanced workflow, most glitches have low ML confidence and may be representative of new morphological classes. The task of the volunteers thus shifts from classifying glitches into known classes to searching for similarities in the images that indicate a possible novel class of glitch.

As volunteers classify images, their classifications (which are coupled to a unique *confusion matrix* for each user that determines which classes a particular user commonly confuses with another) are combined with the initial ML confidence score to determine whether a particular subject has been classified with a high enough pre-set accuracy to be retired from the project and added to the ML training set; these glitches that are added to the ML training set improve its morphological coverage. Theoretically, this iterative retraining of the ML model could be automated. However, several variables govern this retirement procedure, and investigation into the optimal settings for these retirement criteria is ongoing.

2.4 Volunteers of Gravity Spy: supporting the machine and making discoveries

Hand-labeling glitches with the predefined glitch classes represents the lion's share of work in Gravity Spy, without which the ML could not be trained nor benchmarked. Gravity Spy volunteers are essential in providing these by-eye classifications. As discussed earlier, the input provided by volunteers goes beyond the initial classification task; volunteers help identify and characterize glitches that do not fit into previously known glitch classes. When glitches do not fit one of the known glitch classes in the primary labeling, volunteers can begin by labeling them as NOTA, and proceed to create collections of such glitches that exhibit similar morphologies.

Large numbers in a new glitch class indicate that this morphology may be particularly detrimental to GW detector sensitivity. Collecting large samples of such glitches can allow for LIGO scientists to identify trends (e.g., in the times that they occur or in the auxiliary sensors that are triggered at the time of the glitch) [14, 29]. In addition to supplementing glitch collections by continuing through the classification task and identifying similar images, which can be cumbersome and time consuming, we enable volunteers to perform additional data analysis tasks using additional tools side by side with Zooniverse, collectively known as *Gravity Spy Tools*.⁷ One such tool that has proved to be highly useful is the *Similarity Search*, in which volunteers (and project scientists) are able to query similar-looking glitches within the full dataset.⁸ The search utilizes transfer learning, first modeling the properties of existing glitch classes in a high-dimensional feature space and then relying on a clustering algorithm to identify images that are morphologically similar in the database [30, 31]. To use the tool, project scientists or volunteers input a particular glitch subject (each glitch has a unique subject label), and query other glitches in the dataset that most closely resemble that subject morphologically. After running the Similarity Search (a screenshot of the output is shown in Fig. 4), users can evaluate the metadata of resulting glitches, decide which images to include or exclude, and export the results of the search to a new collection. The Similarity Search plays a crucial role by effectively filtering out the majority of the non-matching glitches, significantly enhancing the purity of the set that the volunteer will examine. The ability of volunteers to search the entire dataset for similar subjects is a methodology unique to citizen-science projects that proved highly useful; rather than proceeding through the classification task until similar-looking subjects appeared and could be added to a collection, this tool allowed project users to quickly build morphologically similar classes of glitches that could be vetted and added to the ML training. Details of the clustering algorithm and similarity search can be found in [30–32].

Following the identification and curation of a new glitch class, volunteers can submit a New Glitch Proposal (Fig. 5, bottom), including a proposed name for the glitch, a description of its typical morphological features, a single exemplar or reference glitch, and their collections of similar images. The proposal is evaluated by a LIGO member who assesses the robustness and usefulness of the proposed new class and then communicates whether it should be included in the list of glitch options in the classification interface.

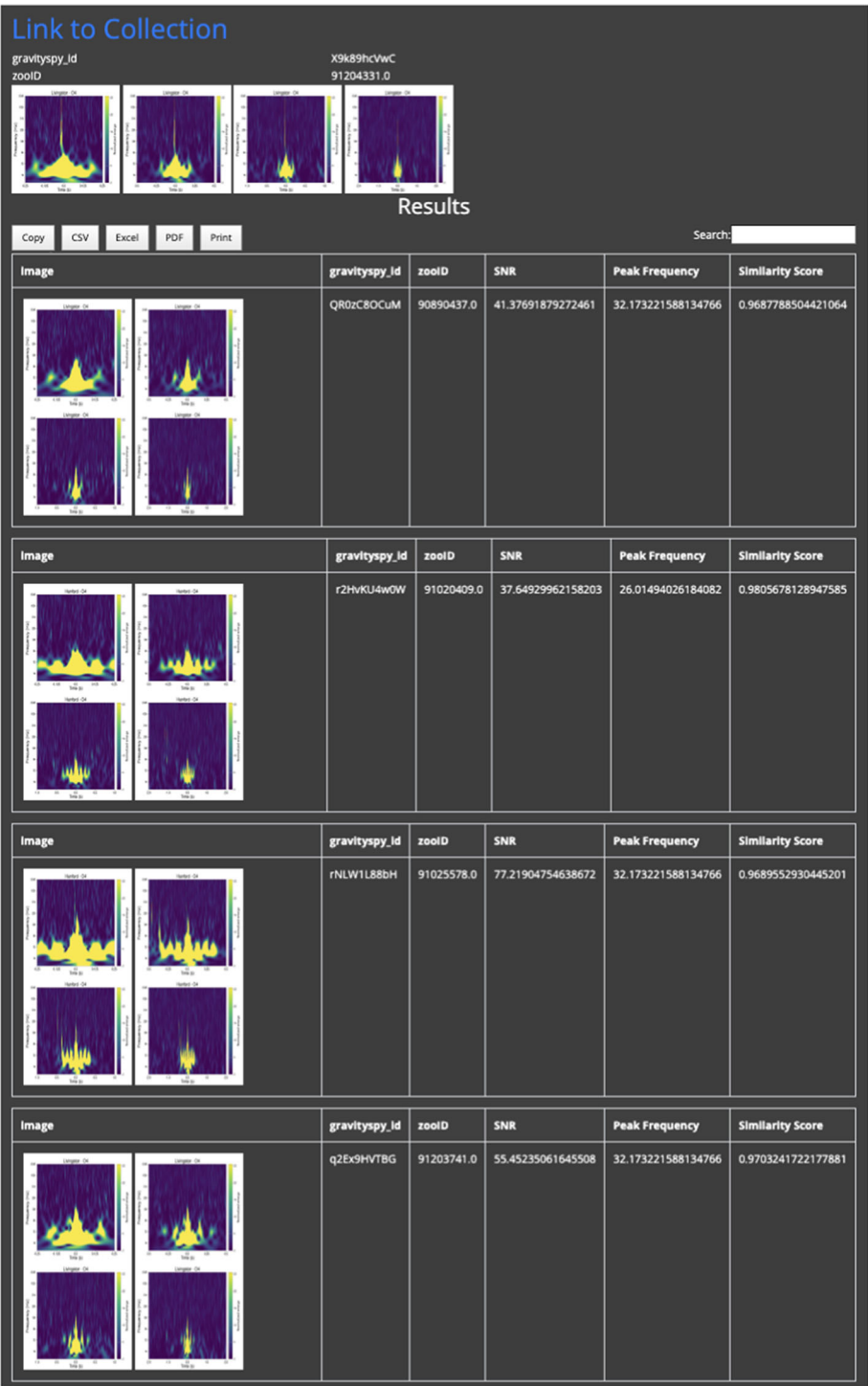
3 Results from Gravity Spy

Gravity Spy's impact has been three-fold: it has improved glitch mitigation in GW detectors and, thereby, the quality of GW observations; it has explored new approaches to ML and human-computer interaction, and it has increased scientific engagement from community members. In the following subsections, we describe these impacts in detail.

⁷ Gravity Spy Tools gravityspytools.ciera.northwestern.edu.

⁸ Gravity Spy Similarity Search gravityspytools.ciera.northwestern.edu/search/.

Fig. 4 Results from a query of the Similarity Search tool. The glitch at the top of the image is was glitch that was queried on, and the four glitches below are some of the most similar glitches in the database. Similarity scores are in the rightmost column of the table



3.1 Improving the sensitivity of GW detectors and analyses

One key investigation that Gravity Spy has enabled is monitoring the rate of glitches by glitch type. The disappearance of a specific glitch class can, for example, indicate that a specific mitigation strategy undertaken by the detector specialists is working [12, 33]. Monitoring different glitch types over time provides both a high-level assessment of the detectors’ state and some indication of whether a specific glitch type is responsible for increased detector noise. For example, Fig. 5 of [16] shows the hourly glitch rates for four types of glitches at LIGO Hanford and LIGO Livingston during O3.

Beyond a high-level assessment of detector noise, detector-characterization experts also use Gravity Spy classifications to search for correlations between individual glitch classes and features in auxiliary channels that monitor the state of the detectors and related

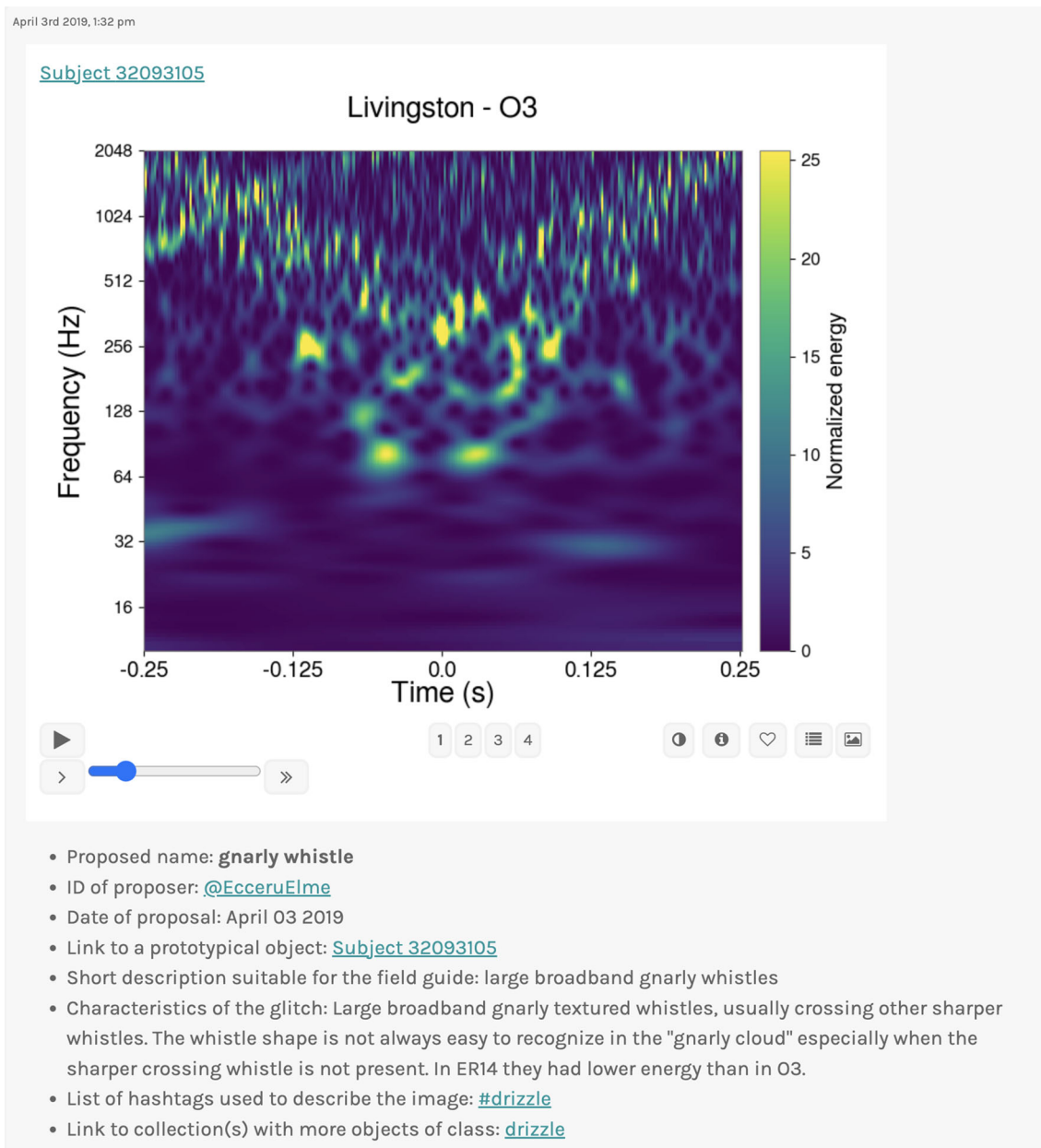


Fig. 5 New glitch proposal by Gravity Spy user EcceruElme

instruments. Gravity Spy classifications are easily accessible to LIGO scientists through the internal database *LigoDV-Web* [34]. Investigations by detector experts, which could involve Gravity Spy classifications, are typically described in *aLogs*, LIGO's records on detector functioning and data quality. One such example can be seen in [35], where potential correlations are identified between a sub-class of the Low-Frequency Blip class and noise in an auxiliary channel monitoring the suspension systems in LIGO Hanford. Similarly, one can investigate the effects of adjusting the instruments on the rate of different classes of glitches, as exemplified in another *aLog* [36].

Gravity Spy has also had a major impact on the understanding of LIGO noise through the identification of new glitch classes. For example, LIGO scientists identified the Low-Frequency Blip class through investigations of many glitches classified by Gravity Spy as regular Blips; the sub-class of Blips at lower frequencies prompted the creation of a new class in the Gravity Spy project [19]. Gravity Spy volunteers play a key role in this process; they can propose new glitch classes, some of which are eventually incorporated into the ML and classifications. One of the first instances of this was the identification of the Paired Doves glitch class by Gravity Spy volunteers during beta testing [15, 37]. Paired Doves are glitches that are morphologically similar to the GW signals expected from merging compact objects, so identifying this new class was critical for mitigating false positive GW detections.

Another notable example is when Gravity Spy volunteers identified the Crown glitch class, which was contemporaneously identified by LIGO scientists as Fast Scattering [19, 38]. These glitches were originally classified into the NOTA class or the Scattered Light class, but further investigation noted they should be their own class, separate from another sub-class of scattered light now known as Slow Scattering. The independent work of the volunteers and LIGO scientists enabled us to verify that the ML algorithm trained using either the volunteers' Crown training set or the experts' Fast Scattering training set produced comparable results [19], demonstrating the volunteers' capability in identifying new classes. This Fast Scattering/Crown class was found to be associated with ground motion and accounted for a major fraction of glitches in LIGO Livingston during O3 [6, 16]. The combination of quick classification by ML and by-eye investigations by both LIGO scientists and community members was essential to identifying and understanding this new glitch class.

Beyond direct application to characterizing glitches, Gravity Spy classifications have also been used to understand the effects of detector noise on GW measurements and as inputs to other GW software pipelines. Gravity Spy data products, including both ML and volunteer classifications of the entire set of glitches up through O3, are available for public use [39, 40], with proprietary data from ongoing observing runs similarly available to LVK scientists. Gravity Spy classifications have been used to examine how different glitch classes affect data analysis [41–44], for example, how they may contaminate GW searches [45]; the development of algorithms to distinguish GW signals from glitches [46–63]; the development of techniques to remove glitches from the data [64–66]; investigations of potential environmental or instrumental origins of noise [12, 67–69], such as what triggers the appearance of light-scattering noise [19, 33]; simulating synthetic (noisy) GW data [70–72]; and training or testing alternative ML glitch classification algorithms [57, 73–79]. Overall, Gravity Spy classifications have been instrumental in GW detector characterization, spanning from informal investigations to a key part of GW data analysis software [14].

3.2 Development of machine learning

As described earlier, Gravity Spy incorporates CNNs for machine learning-based glitch classification. CNNs can automatically learn hierarchical patterns and features from images to effectively distinguish between different glitch classes. The CNN architecture presented in Fig. 6 of [15] serves as the foundation, comprising two convolutional (Conv) and max-pooling layers, followed by two fully-connected (FC) layers. The final FC layer employs a softmax activation function, generating scores for each class based on a given set of input images. To create an input, spectrograms with four durations generated with different time windows (0.5 s, 1.0 s, 2.0 s, and 4.0 s, the same that are shown to the volunteers of the project) are combined to form a square image. This merging process ensures the convolutional kernels effectively slide over all four durations, enabling the model to learn distinct features from both long- and short-duration glitches. This architecture has been trained multiple times in different contexts as shown in Table 1, as we describe below.

The initial model was trained using a dataset of 7718 glitches from 20 classes observed during the first observing run (O1) of Advanced LIGO [15, 25]. This dataset was curated and labeled via a collaboration between detector-characterization experts and citizen-science volunteers from Gravity Spy. The initial model achieved an average accuracy of 97.1% on the testing set. Later, two new glitch classes, the 1080 Line and 1400 Ripple classes, were discovered by citizen-science volunteers during the second observing run (O2) of Advanced LIGO. Therefore, the model was retrained using an expanded dataset of 7932 glitches from 22 classes, encompassing glitches from both of the first two observing runs [32].

The model was later improved with two additional Conv layers and max-pooling layers before the existing two FC layers. These deeper layers allowed the model to capture more intricate patterns in the glitch data. This retrained model achieved an accuracy of 98.2% on the testing set. During O3, LIGO detector-characterization experts and the Gravity Spy volunteers identified two new glitch classes: Fast Scattering/Crown and Low-frequency Blips [19]. In addition, the NOTA class was removed from the glitch classes defined in the ML model. This particular class is useful for volunteers to flag potential new glitch types but held limited significance for the model's classification process due to the large variety of morphological features it encompassed. Therefore, the model was retrained again on a dataset of 9631 glitches from 23 classes in O3 [16, 19]. This most recent model reported training and validation accuracies of 99.9% and 99.8%, respectively.

While each of the ML classifiers mentioned above demonstrated strong performance, they have some common challenges and limitations. First is the imbalanced distribution of glitches across classes. For instance, certain classes like Chirp and Wandering Line suffer from a scarcity of training samples, which can introduce bias during model training. To address this issue, it is recommended to incorporate class-specific evaluation metrics, such as precision and recall, when assessing the model. Second is the presence of noisy glitch labels. The training labels are derived from a combination of ML and volunteer classification, making it likely that the model relies on partially incorrect labels during the training phase. Furthermore, the model architecture remains relatively simple, relying on the direct combination of a few convolutional layers. This simplicity imposes constraints on the model's ability to perform effective feature extraction, particularly when dealing with morphologically similar glitch classes. Future studies could potentially explore more advanced model architectures to further enhance the performance of the glitch classifiers. Finally, due to its fully supervised nature, the ML method faces challenges in recognizing new glitch classes that fall outside the predefined classes during training, which highlights the importance of involving volunteers in the process of identifying new glitch classes.

Table 1 The development of ML models for glitch classification in the Gravity Spy system

Papers	Number of glitch classes	Dataset size	Input size	Model architecture
Bahaadini et al. [25] and Zevin et al. [15]	20	7718	94×114 (47×57 for each duration)	(Conv + MaxPool) $\times 2$ + FC $\times 2$
Bahaadini et al. [32]	22	7932	280×340 (140×170 for each duration)	(Conv + MaxPool) $\times 4$ + FC $\times 2$
Soni et al. [19]	23	9631	280×340 (140×170 for each duration)	(Conv + MaxPool) $\times 4$ + FC $\times 2$

3.3 Engagement with project volunteers

While a significant amount of work has studied both machine- and human-based classification schemes, we know little about how to use machine-coded data to improve human learning and performance. To this end, we devoted substantial effort to researching how to recruit volunteers, train volunteers with the support of ML scaffolding, and assess volunteer engagement.

3.3.1 Volunteer recruitment

Zooniverse hosts hundreds of projects, allowing participants to be recruited from the existing user base. We engaged in supplemental recruitment through blog posts and announcements about GW discoveries and empirical research on attracting and retaining participants (i.e., motivation). A persistent challenge in citizen-science projects is that most volunteers participate infrequently, oftentimes only once. Research in citizen science has focused on strategies to motivate contributions, with studies on volunteer motivation in citizen-science finding that user motivation differs person-to-person and changes over time [80, 81].

To test theories of motivation, we conducted one study focused on increasing classification volume using novelty theory [82]. In this experiment, we showed participants a unique message when they were the first person to see a particular glitch. While the theory was tested in other projects, the findings revealed novelty messages had no effect on volunteer behaviors. We hypothesize that other aspects of the project (e.g., gamified elements through leveling not found in other projects) might be sufficiently motivating. We also conducted an experiment to evaluate the efficacy of recruitment messages appealing to four types of motivations known to be salient in citizen-science projects: science learning, contribution to science, joining a community of practice, and helping scientists [83]. Messages about contributing to science resulted in more classifications; however, statements about helping scientists generated more initial interest from participants.

3.3.2 Volunteer training

Participants of the Gravity Spy project may have little classification task experience, so training volunteers is crucial for ensuring data quality. The Zooniverse platform supports training through text- and image-based tutorials. We conducted several investigations to gauge the efficacy of our training and learning regimens. One such study was conducting an online A/B field experiment to evaluate the training procedure described in Sect. 2.2 [28]: Group A received all glitch classes (with a wide range ML confidence) without training when they started the project, whereas Group B received the default scaffolded training through the workflow levels that increase in complexity and difficulty as they proceed through the project. As anticipated, we found that volunteers who received the training were more accurate in their classifications (as indicated by agreement on gold-standard data and with the eventual consensus decision on the glitch) and contributed significantly more classifications to the project.

Another study used digital trace data (produced as a by-product of interactions with computer systems) to evaluate how volunteers use learning resources when given feedback about their classification (e.g., “You answered Blip, but our experts classified this image as Whistle”). The results demonstrated that authoritative knowledge provided by the project scientists improved learning during early workflows. However, as the challenge increases in advanced workflows, volunteers rely on resources constructed by the community (e.g., discussion boards) to learn to identify glitches more accurately [84]. The results of our investigation into the learning process have informed design decisions for integrating scaffolded tasks in subsequent citizen-science initiatives. Our findings outline optimal strategies to blend human and ML schemes to train volunteers and identify the temporal importance of learning resources. Additionally, this research underscores the significance of learning resources generated by volunteers in the absence of expert guidance.

3.3.3 Volunteer engagement

Much research has been conducted about people engaging in virtual citizen-science projects. It tends to demonstrate that many volunteers contribute once or infrequently while a handful of volunteers perform the majority of work and limit their engagement to image classification [85, 86]. This observation is corroborated by a Gini coefficient of 0.85, indicative of a high level of inequality in participation as found in other citizen-science projects [87]. In Gravity Spy, on average, a volunteer makes 235 classifications

(median = 2) before dropping out, and less than 12% of volunteers have contributed to the project's discussion boards. While engagement is skewed, a small cadre of highly motivated and engaged volunteers is active on the site. Our analysis of digital trace data shows how volunteers engage beyond submitting classifications, engaging with discussion boards and other project infrastructure. Our research describes how volunteers share new knowledge by linking external resources (e.g., arXiv preprints) and sharing results of independent investigations speculating about the causes of glitches in the data [88, 89]. As noted above, many glitch class proposals are the product of intense data curation and evaluation on the discussion boards. We find both independent and collaborative interactions [90]. Overall, our research on engagement demonstrates the ability of volunteers to engage in more complex work when provided the means to do so.

3.3.4 Volunteer contributions

Volunteer efforts have been crucial to Gravity Spy. At present, over 7.4 million classifications have been performed by project volunteers, with more than 32,000 registered users (and many more users that did not register with a Zooniverse account) contributing to these classifications. Volunteers have drafted over 30 new glitch proposals: in O3, new glitch classes from these proposals were included in the glitch classification interface and used in retraining the ML model [16, 19].

4 Challenges and future considerations

Gravity Spy has faced several challenges that can be traced back to the intersection between the science team's and volunteers' work: (1) science team–volunteer engagement; (2) mismatching temporal rhythms and divided priorities across the science team and volunteers; (3) retraining of ML models, and (4) the constant evolution of the GW detectors.

First, the science team often lacked the capacity for sustained volunteer interactions, a challenge for many citizen-science projects [91]. In Gravity Spy, this was evidenced through bottle necks in approving new glitch class proposals and questions posted to discussion boards. Glitch proposals required extensive efforts by the science team to evaluate the veracity of the proposed glitch in the data stream, whether proposed glitch classes are still impacting GW detectors and relevant to detector-characterization science, and, when proposals were rejected, justifying the decision to the volunteers. For some proposals, many months can pass between an initial glitch proposal by a volunteer and a resolution to the proposal by the science team.⁹ In evaluating glitch proposals, we trained physics undergraduate and graduate students to review new glitch class proposals. However, student schedules quickly fill up, or they move on to other projects, so solely relying on students for glitch proposal review is not a long-term solution. Meanwhile, some moderators know as much or more about the detectors as students and even members of the science team, but lack formal training, easy access to experienced science team members, and password-protected LIGO resources. In such instances, we have asked moderators to tag science team members who they believe have expertise in the question. Another challenge is that volunteers do not always realize that seemingly simple questions posted to discussion boards may take hours to address satisfactorily.

We have experimented with different methods to triage questions to the relevant experts and enrolled detector-characterization group members (not already a part of the Gravity Spy research team) to be on call to respond to questions on the discussion boards. We quickly discovered that many questions pertained to the project infrastructure (e.g., debugging the promotion algorithm) or similar issues that could only be handled and addressed by a small number of people or a single person on the project team.

In the latest version of the project (see Sect. 5), we strive to make volunteers more self-reliant by giving them access to more background knowledge about the detectors when needed (the same knowledge science teams use to evaluate glitch class proposals). One element of this solution involves building a wiki that gradually expands volunteers' access to expert knowledge about the detectors [92, 93].

Second, the temporal rhythms and priorities guiding science team members' and volunteers' work are not always consistent. For instance, the investigations by the detector-characterization group and volunteers do not always align. While the detector-characterization team works on priorities for an entire observing run, which may involve time-sensitive investigations and fixes to the instruments, these may occur before volunteers can identify a new glitch, given the time required to build collections and develop solid proposals about a new glitch class. In contrast, volunteers may submit new glitch class proposals involving glitches unknown to the detector-characterization group (e.g., the Pizzicato glitch). To solve this issue, we are considering a real-time citizen-science format where volunteers would work on urgent problems and new data that call for quick turnarounds.

Similarly, the science team often needs nearly real-time classification of glitches. As a result, they have come to rely on the automated ML classifications, rather than waiting for the theoretically better human-classified results. Our realization of this dynamic has led to a refocusing of the project: more quickly retiring images rather than expending volunteer effort to refine them, using the volunteer results for model retraining rather than as immediate feedback to LIGO scientists, and emphasizing the role of volunteers in the discovery and characterization of novel glitch classes.

The third major challenge in the Gravity Spy project has been that retraining the ML model based on volunteer classifications has taken longer than initially planned. The issue has multiple elements. The dataset containing volunteer classifications has impurities,

⁹ Gravity Spy Talk: Gnarly Whistle Glitch Proposal www.zooniverse.org/projects/zooniverse/gravity-spy/talk/762/951832, and Pizzicato Glitch Proposal www.zooniverse.org/projects/zooniverse/gravity-spy/talk/762/935664.

and we have had to experiment with different ways to assess its accuracy. Rare glitch classes have few examples, making the integration of them into a retrained model difficult. From a ML perspective, dealing with four versions of the same glitch images in different time durations can also be challenging. While it is helpful for the volunteers to move between different time durations, the ML can sometimes regard some time durations as more important than others, leading to potential classification problems.

Fourth, GW detectors continually evolve [94–97]; with it, new glitch classes emerge, and known classes disappear or become less prevalent. One cannot leave the ML model unattended for long. At the beginning of O4, the volunteers and, soon after, the science team found that the ML model was confidently mislabeling instances of a novel glitch as a Whistle glitch, which is one of the two glitch classes presented to volunteers in their first workflow. During this initial portion of O4, the detectors did not produce many Whistles, even though they used to be common enough that Gravity Spy used them for training purposes in the initial workflow. As such, we are considering a replacement for the Whistle at the introductory volunteer level as we retrain the ML model to distinguish Whistles from the new glitch.

The constant evolution of the detectors emphasizes the importance of clearly distinguishing the architecture of the ML model from the different retrained versions of a specific model. We have had to carefully track how the model changes as one experiments with new, more effective architectures. Equally important, we have had to verify the provenance of the different versions of a model as it gets retrained on new classification data, some of which might include new glitches. This is not always easy in a distributed science team with many stakeholders interested in the continuous improvement of the model.

Finally, making all the improvements deemed necessary is not always possible. For instance, we have yet to complete a volunteer performance assessment system that moves beyond relying on gold-standard data. The current ML model would guide the presentation of image classification tasks to newcomers to help them learn how to do the task more quickly while contributing to the project's work. We designed but never implemented a Bayesian model that would both estimate volunteers' ability and decide the classification of an image. The reason for not implementing this scheme was due to occasional high-confidence misclassifications from the ML model, which, when the volunteers made an alternative selection, would hinder their ability to progress through the project workflows. A simulation of the model applied to volunteer promotion and image retirement suggests that the model would require fewer classifications than the current system and, in the process, save volunteers' time [98].

The challenges that Gravity Spy faces need to be understood in the context of its successes. The Gravity Spy project has led to substantial innovations in combining machine learning and volunteer-based classification schemes, and demonstrated how the facilitation of a symbiotic relationship between these methodologies leads to significant improvement in classification and characterization tasks. The Gravity Spy classifications and associated models have benefited GW detector characterization, and the project has engaged more than 30 thousand volunteers in the scientific process. This success was tied to the substantial user base and promotion that Zooniverse has developed over the past 15 years, as well as the public interest in the growing field of GW science since the start of O1 in 2015. The project also benefits from an active group of advanced volunteers. Led by a dedicated team of volunteer moderators, Gravity Spy continues to discover, collect, characterize, and name new glitch classes. The volunteer moderators have also played a critical role in responding to the many questions of other users in the project, a feat that would not have been possible by the science team alone. The scaffolded training where volunteers gradually level up has greatly improved volunteer engagement and retention [28], and is being implemented across other Zooniverse projects. To a certain extent, the success of the Gravity Spy classification model and the high volunteer engagement has removed the urgency of addressing some the outstanding issues in the project itself.

5 Gravity Spy 2.0

The success and challenges of Gravity Spy have spurred us to explore how volunteers can engage in more complicated analysis of LIGO data that would help detector-characterization scientists identify the causes of glitches and isolate these signals in the data stream. Along with the main strain channel that is sensitive to GWs, the LIGO detectors record more than 200,000 auxiliary channels of data per detector from a diverse set of sensors that continuously measure every aspect of the detectors and their environment (e.g., equipment functioning, activation of components, seismic activity, or weather) [11, 99]. To explore the cause of glitches (i.e., what is happening in the detector or the environment that causes particular glitches), detector-characterization scientists conduct investigations using the temporal correlation of these auxiliary channels and the main GW channel. Using these data, researchers further isolate noise signals in the data stream, often analyzing output data using statistical software and conducting visual inspections of GW and auxiliary channel glitch images.

In 2020, the Gravity Spy collaboration began work funded by a new multi-year National Science Foundation grant (HCC 2106865) focused on developing the capacity for amateur volunteers to conduct similar activities and develop causal insights about glitches. This project, known as *Gravity Spy 2.0*, is currently live as part of the broader Gravity Spy infrastructure. It is an extension of our efforts in the initial Gravity Spy and is designed to help volunteers develop their knowledge of the data over time by progressing through three stages (also Phases of the grant).

Stage 1 Volunteers will compare glitches in the main GW channel to glitches in auxiliary channels. At first, we provide users with 3 auxiliary channels per glitch subject, totaling in ~ 20 distinct auxiliary channels being used across the glitch classes

considered in the first phase of the project. Human input is valuable because there could be morphological similarities between the glitches in the channels that point to some association. However, there is no simple rule to determine which channels are relevant or how the channels are coupled. Engagement with the data and the channel ontology in this phase will also support volunteers in learning about the auxiliary channels and their potential relationship to a glitch in the GW channel. A similar approach has been tried using data from the Virgo detector in the GWitchHunters project [100].¹⁰

- Stage 2 The primary task will shift from individual glitches to collections, building on the filtered dataset of glitches and auxiliary channels created in Stage 1. Volunteers are expected to (1) create collections of glitches that might represent a novel glitch class with hypothesized common causality, not just similar appearance; (2) search for relations between those glitches and groups of auxiliary channels over time; and (3) note possible causes behind the correlated glitches and auxiliary channels.
- Stage 3 Advanced volunteers will analyze the network compiled at Stage 2 to identify the root causes of glitches. A supporting activity is adding paths among auxiliary channels, e.g., by examining patterns that appear across multiple classes of glitches. For instance, if several channels are involved simultaneously in many classes of glitches, they are likely interrelated.

In each stage, we will use novel interfaces and tools to support volunteer tasks. Our research adopts a *human-centered design* approach involving volunteers and detector-characterization scientists in developing Gravity Spy 2.0.

Development of Gravity Spy 2.0 and research into how best to use its output is ongoing. Discussions with detector-characterization scientists are needed to understand and translate their work making causal inferences. To that end, we need to understand what background knowledge about the glitches and the detector is necessary and how to structure the tasks. On the volunteer side, we need to know how best to align the system support with volunteers' current knowledge and capabilities. To that end, we recently completed interviews and design critiques with long-time Gravity Spy volunteers. During the user studies, we gained insights into volunteers' glitch classification practices, which will inform how we design the three stages.

Much like the original Gravity Spy, the output of this investigation will contribute new knowledge about cutting-edge citizen-science techniques that involve amateur volunteers in more advanced scientific tasks.

6 Citizen science in physics and beyond

Gravity Spy's use and advancement of strategies that combine human and machine effort have produced an important blueprint for future citizen-science projects in physics and a broad range of other research domains. The capabilities of human crowdsourcing alone cannot keep up with the demands of ever-growing data volumes, even when assuming continued access to large communities of volunteers. However, the value of human input is unquestionably high, particularly in applications such as out-of-sample classification, serendipitous discovery, and other advanced analyses. Therefore, the continued development and application of hybrid machine–human workflows is essential for unlocking the full potential of many datasets.

Gravity Spy has demonstrated paths forward for capitalizing on the capabilities and strengths of human crowdsourcing. First, the ML-informed assignment of human effort in Gravity Spy, where subjects with high machine confidence are tasked to beginners and retired quickly and subjects with uncertain or unidentified machine classifications are assigned to experienced volunteers, is a model applicable to many projects. If ML models can expedite the classification of easy and routine data, that leaves greater effort for humans to contribute in more meaningful and advanced ways. Moreover, if projects successfully combine machine and human classifications for use in improving and refining the machine model, the resulting classification quality and system efficiency can be improved and optimized. The training set augmentation and model retraining strategy of Gravity Spy or the active learning strategy employed by a recent iteration of the Galaxy Zoo project [101], show the benefits of this human–machine interplay and its applicability in multiple research domains.

Second, Gravity Spy affirms a broader trend across Zooniverse citizen-science projects that a subset of volunteers is willing to contribute in advanced ways that go well beyond the project's primary classification task. The investigation and proposal of new glitch classes and the use of the project's Similarity Search tool show the willingness of certain volunteers to perform information synthesis activities, which augment the typical, more narrow scope of machine classifiers. This behavior mirrors that seen in other projects, such as volunteers performing follow-up analysis of previously unknown galaxy sub-types via the Galaxy Zoo project (e.g., Green Pea galaxies [102]) or the use of publicly available external data and software to characterize exoplanets as part of the Planet Hunters TESS project [103]. A common thread among these examples is the ability for volunteers to access data and tools that allow additional contributions toward the project science goals, and the resulting reward in the form of new findings to project teams that facilitate and support these volunteer activities.

Third, the patterns of interaction associated with Gravity Spy's glitch proposals demonstrate the importance of setting up effective processes for interaction and feedback between research teams and volunteers, especially those carrying out advanced work. While ambitious volunteers may take it upon themselves to reach out and communicate with the project team using basic means (email, Zooniverse Talk boards, etc.), both the research team and volunteers often benefit from creating structures and processes to facilitate useful and beneficial interactions. This conclusion motivates a recommendation that citizen-science projects craft a plan for volunteer interaction that includes tools for both communication and analysis, especially to encourage more independent forms of volunteer

¹⁰ GWitchHunters Project www.zooniverse.org/projects/reinforce/gwitchhunters.

contribution. We highlight that the Gravity Spy team's interdisciplinary composition was an important asset in assembling a successful plan for volunteer engagement and advocate that research teams that include members spanning a broad range of project-enabling skills will better position their projects for potential success.

In summary, Gravity Spy has pioneered how ML approaches and citizen science can be synthesized across various domains to produce useful results. GW science has been supported in the process through large-scale classification and characterization of glitch classes and identification of new glitch morphologies, thereby enabling the identification of glitch origins and the removal of glitches from the GW datastream or eliminating the root cause of glitches from the detectors entirely. The next phase of the Gravity Spy project and the continued flow of new observations from LIGO, Virgo and KAGRA show that this project will continue to meaningfully contribute to scientific discovery for years to come.

Acknowledgements The authors first and foremost are indebted to the citizen-science volunteers of the Gravity Spy project, whose dedicated work allowed for this ambitious project to become a reality. The authors also thank Laura Nuttall, Lynn Cominsky, and the anonymous referees for comments that improved this manuscript. This publication uses data generated via the Zooniverse.org platform, development of which is funded by generous support, including a Global Impact Award from Google, and by a Grant from the Alfred P. Sloan Foundation. Gravity Spy is partly supported by the National Science Foundation (NSF) awards INSPIRE 1547880, IIS-2107334, 2106882, 2106896, 2106865, 2107334 and PHY-1912648. Support for MZ was provided by NASA through the NASA Hubble Fellowship Grant HST-HF2–51474.001-A awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Incorporated, under NASA contract NAS5-26555. ZD is supported by the CIERA Board of Visitors Research Professorship. ZD, VK, SB, and JS acknowledge support from PHY-2207945. YW, VK, SB, and RH were also supported by IIS-2107334. VK is grateful for support from a Guggenheim Fellowship, from CIFAR as a Senior Fellow, and from Northwestern University, including the Daniel I. Linzer Distinguished University Professorship fund. CPLB acknowledges past support from the CIERA Board of Visitors Research Professorship, and current support from Science and Technology Facilities Council (STFC) Grant ST/V005634/1. DD is supported by the NSF as a part of the LIGO Laboratory. SS is supported by NSF Award PHY-1764464 to LIGO Laboratory. This material is based upon work supported by NSF's LIGO Laboratory, a major facility fully funded by the National Science Foundation. This work used computing resources at CIERA funded by NSF Grant PHY-1726951, and the computational resources and staff contributions provided for the Quest high performance computing facility at Northwestern University, which is jointly supported by the Office of the Provost, the Office for Research, and Northwestern University Information Technology. The authors are grateful for computational resources provided by the LIGO Laboratory and supported by National Science Foundation Grants PHY-0757058 and PHY-0823459. This document has been assigned LIGO document number LIGO-P2300256.

Data Availability Statement This manuscript has associated data in a data repository. [Authors' comment: The Gravity Spy datasets generated and analyzed during the current study, including machine learning and volunteer classification information, are available in Zenodo repositories [39, 40]. Datasets generated for Gravity Spy 2.0 are available from the corresponding author on reasonable request.]

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. A. Einstein, Approximative integration of the field equations of gravitation. *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)* **1916**, 688–696 (1916)
2. A. Einstein, Über Gravitationswellen. *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)* **1918**, 154–167 (1918)
3. B.P. Abbott et al., Observation of gravitational waves from a binary black hole merger. *Phys. Rev. Lett.* **116**(6), 061102 (2016). <https://doi.org/10.1103/PhysRevLett.116.061102>. [arXiv:1602.03837](https://arxiv.org/abs/1602.03837) [gr-qc]
4. J. Aasi et al., Advanced LIGO. *Class. Quant. Grav.* **32**, 074001 (2015). <https://doi.org/10.1088/0264-9381/32/7/074001>. [arXiv:1411.4547](https://arxiv.org/abs/1411.4547) [gr-qc]
5. F. Acernese et al., Advanced Virgo: a second-generation interferometric gravitational wave detector. *Class. Quant. Grav.* **32**(2), 024001 (2015). <https://doi.org/10.1088/0264-9381/32/2/024001>. [arXiv:1408.3978](https://arxiv.org/abs/1408.3978) [gr-qc]
6. R. Abbott, et al.: GWTC-3: compact binary coalescences observed by LIGO and virgo during the second part of the third observing run (2021). [arXiv:2111.03606](https://arxiv.org/abs/2111.03606) [gr-qc]
7. T. Akutsu et al., KAGRA: 2.5 generation interferometric gravitational wave detector. *Nat. Astron.* **3**(1), 35–40 (2019). <https://doi.org/10.1038/s41550-018-0658-y>. [arXiv:1811.08079](https://arxiv.org/abs/1811.08079) [gr-qc]
8. B.P. Abbott et al., Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA. *Living Rev. Rel.* **23**(1), 3 (2020). <https://doi.org/10.1007/s41114-020-00026-9>. [arXiv:1304.0670](https://arxiv.org/abs/1304.0670) [gr-qc]
9. B.P. Abbott et al., A guide to LIGO-Virgo detector noise and extraction of transient gravitational-wave signals. *Class. Quant. Grav.* **37**(5), 055002 (2020). <https://doi.org/10.1088/1361-6382/ab685e>. [arXiv:1908.11170](https://arxiv.org/abs/1908.11170) [gr-qc]
10. L. Blackburn et al., The LSC glitch group: monitoring noise transients during the fifth LIGO science run. *Class. Quant. Grav.* **25**, 184004 (2008). <https://doi.org/10.1088/0264-9381/25/18/184004>. [arXiv:0804.0800](https://arxiv.org/abs/0804.0800) [gr-qc]
11. B.P. Abbott et al., Characterization of transient noise in Advanced LIGO relevant to gravitational wave signal GW150914. *Class. Quant. Grav.* **33**(13), 134001 (2016). <https://doi.org/10.1088/0264-9381/33/13/134001>. [arXiv:1602.03844](https://arxiv.org/abs/1602.03844) [gr-qc]
12. D. Davis et al., LIGO detector characterization in the second and third observing runs. *Class. Quant. Grav.* **38**(13), 135014 (2021). <https://doi.org/10.1088/1361-6382/abfd85>. [arXiv:2101.11673](https://arxiv.org/abs/2101.11673) [astro-ph.IM]
13. L.K. Nuttall, Characterizing transient noise in the LIGO detectors. *Philos. Trans. R. Soc. Lond. A* **376**(2120), 20170286 (2018). <https://doi.org/10.1098/rsta.2017.0286>. [arXiv:1804.07592](https://arxiv.org/abs/1804.07592) [astro-ph.IM]
14. D. Davis, M. Walker, Detector characterization and mitigation of noise in ground-based gravitational-wave interferometers. *Galaxies* **10**(1), 12 (2022). <https://doi.org/10.3390/galaxies10010012>

15. M. Zevin et al., Gravity Spy: integrating Advanced LIGO detector characterization, machine learning, and citizen science. *Class. Quant. Grav.* **34**(6), 064003 (2017). <https://doi.org/10.1088/1361-6382/aa5cea>. arXiv:1611.04596 [gr-qc]
16. J. Glanzer et al., Data quality up to the third observing run of Advanced LIGO: Gravity Spy glitch classifications. *Class. Quant. Grav.* **40**(6), 065004 (2023). <https://doi.org/10.1088/1361-6382/acb633>. arXiv:2208.12849 [gr-qc]
17. J. Schmidt, M.R.G. Marques, S. Botti, M.A.L. Marques, Recent advances and applications of machine learning in solid-state materials science. *npj Comput. Math.* **5**, 83 (2019). <https://doi.org/10.1038/s41524-019-0221-0>
18. N.M. Ball, R.J. Brunner, Data mining and machine learning in astronomy. *Int. J. Mod. Phys. D* **19**, 1049–1106 (2010). <https://doi.org/10.1142/S0218271810017160>. arXiv:0906.2173 [astro-ph.IM]
19. S. Soni et al., Discovering features in gravitational-wave data through detector characterization, citizen science and machine learning. *Class. Quant. Grav.* **38**(16), 195016 (2021). <https://doi.org/10.1088/1361-6382/ac1ccb>. arXiv:2103.12104 [gr-qc]
20. R. Bonney, C.B. Cooper, J. Dickinson, S. Kelling, T. Phillips, K.V. Rosenberg, J. Shirk, Citizen science: a developing tool for expanding science knowledge and scientific. *Literacy* **59**(11), 977–984 (2009). <https://doi.org/10.1525/bio.2009.59.11.9>
21. C.J. Lintott et al., Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. *Mon. Not. R. Astron. Soc.* **389**, 1179–1189 (2008). <https://doi.org/10.1111/j.1365-2966.2008.13689.x>. arXiv:0804.4483 [astro-ph]
22. F. Robinet, N. Arnaud, N. Leroy, A. Lundgren, D. Macleod, J. McIver, Omicron: a tool to characterize transient noise in gravitational-wave detectors. *SoftwareX* **12**, 100620 (2020). <https://doi.org/10.1016/j.softx.2020.100620>. arXiv:2007.11374 [astro-ph.IM]
23. S. Chatterji, L. Blackburn, G. Martin, E. Katsavounidis, Multiresolution techniques for the detection of gravitational-wave bursts. *Class. Quant. Grav.* **21**, 1809–1818 (2004). <https://doi.org/10.1088/0264-9381/21/20/024>. arXiv:gr-qc/0412119
24. A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Processing Syst.* **25** (2012)
25. S. Bahaadini, N. Rohani, S. Coughlin, M. Zevin, V. Kalogera, A.K. Katsaggelos, Deep multi-view models for glitch classification, in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2017), pp. 2931–2935. <https://doi.org/10.1109/ICASSP.2017.7952693>
26. L.S. Vygotsky, *Mind in Society: Development of Higher Psychological Processes* (Harvard University Press, 1978)
27. Y. Engeström, *Learning by Expanding* (Cambridge University Press, 2014)
28. C. Jackson, C. Østerlund, K. Crowston, M. Harandi, S. Allen, S. Bahaadini, S. Coughlin, V. Kalogera, A. Katsaggelos, S. Larson et al., Teaching citizen scientists to categorize glitches using machine learning guided training. *Comput. Hum. Behav.* **105**, 106198 (2020)
29. M. Cabero et al., Blip glitches in Advanced LIGO data. *Class. Quant. Grav.* **36**(15), 15 (2019). <https://doi.org/10.1088/1361-6382/ab2e14>. arXiv:1901.05093 [physics.ins-det]
30. S. Bahaadini, V. Noroozi, N. Rohani, S. Coughlin, M. Zevin, A.K. Katsaggelos, DIRECT: deep discriminative embedding for clustering of LIGO data, in *2018 25th IEEE International Conference on Image Processing (ICIP)* (2018), pp. 748–752. <https://doi.org/10.1109/ICIP.2018.8451708>
31. S.B. Coughlin et al., Classifying the unknown: discovering novel gravitational-wave detector glitches using similarity learning. *Phys. Rev. D* **99**(8), 082002 (2019). <https://doi.org/10.1103/PhysRevD.99.082002>. arXiv:1903.04058 [astro-ph.IM]
32. S. Bahaadini, V. Noroozi, N. Rohani, S. Coughlin, M. Zevin, J.R. Smith, V. Kalogera, A. Katsaggelos, Machine learning for Gravity Spy: glitch classification and dataset. *Inf. Sci.* **444**, 172–186 (2018). <https://doi.org/10.1016/j.ins.2018.02.068>
33. S. Soni et al., Reducing scattered light in LIGO's third observing run. *Class. Quant. Grav.* **38**(2), 025016 (2020). <https://doi.org/10.1088/1361-6382/abc906>. arXiv:2007.14876 [astro-ph.IM]
34. J.S. Areeda, J.R. Smith, A.P. Lundgren, E. Maros, D.M. Macleod, J. Zweig, LigoDV-web: providing easy, secure and universal access to a large distributed scientific data store for the LIGO Scientific Collaboration. *Astron. Comput.* **18**, 27–34 (2017). <https://doi.org/10.1016/j.ascom.2017.01.003>. arXiv:1611.01089 [astro-ph.IM]
35. G. Vajente, aLIGO LHO Logbook. <https://alog.ligo-wa.caltech.edu/aLOG/index.php?callRep=35073> (2017)
36. O. Patane, aLIGO LHO Logbook. <https://alog.ligo-wa.caltech.edu/aLOG/index.php?callRep=43177> (2018)
37. A. Lundgren, aLIGO LHO Logbook. <https://alog.ligo-wa.caltech.edu/aLOG/index.php?callRep=27138> (2016)
38. J. Smith, aLIGO LLO Logbook. <https://alog.ligo-la.caltech.edu/aLOG/index.php?callRep=44803> (2019)
39. M. Zevin, S. Coughlin, E. Chase, S. Allen, S. Bahaadini, C. Berry, K. Crowston, M. Harandi, C. Jackson, V. Kalogera, A. Katsaggelos, C. Osterlund, O. Patane, N. Rohani, J. Smith, S. Soni, L. Trouille, Gravity Spy volunteer classifications of LIGO glitches from observing runs O1, O2, O3a, and O3b. <https://doi.org/10.5281/zenodo.5911227>
40. J. Glanzer, S. Banagari, S. Coughlin, M. Zevin, S. Bahaadini, N. Rohani, S. Allen, C. Berry, K. Crowston, M. Harandi, C. Jackson, V. Kalogera, A. Katsaggelos, V. Noroozi, C. Osterlund, O. Patane, J. Smith, S. Soni, L. Trouille, Gravity Spy machine learning classifications of LIGO glitches from observing runs O1, O2, O3a, and O3b. <https://doi.org/10.5281/zenodo.5649212>
41. G. Ashton, S. Thiele, Y. Lecoecuche, J. McIver, L.K. Nuttall, Parameterised population models of transient non-Gaussian noise in the LIGO gravitational-wave detectors. *Class. Quant. Grav.* **39**(17), 175004 (2022). <https://doi.org/10.1088/1361-6382/ac8094>. arXiv:2110.02689 [gr-qc]
42. R. Macas, J. Pooley, L.K. Nuttall, D. Davis, M.J. Dyer, Y. Lecoecuche, J.D. Lyman, J. McIver, K. Rink, Impact of noise transients on low latency gravitational-wave event localization. *Phys. Rev. D* **105**(10), 103021 (2022). <https://doi.org/10.1103/PhysRevD.105.103021>. arXiv:2202.00344 [astro-ph.HE]
43. S. Hourihane, K. Chatziioannou, M. Wijngaarden, D. Davis, T. Littenberg, N. Cornish, Accurate modeling and mitigation of overlapping signals and glitches in gravitational-wave data. *Phys. Rev. D* **106**(4), 042006 (2022). <https://doi.org/10.1103/PhysRevD.106.042006>. arXiv:2205.13580 [gr-qc]
44. J. Heinzl, C. Talbot, G. Ashton, S. Vitale, Inferring the astrophysical population of gravitational wave sources in the presence of noise transients. *Mon. Not. R. Astron. Soc.* **523**(4), 5972–5984 (2023). <https://doi.org/10.1093/mnras/stad1823>. arXiv:2304.02665 [astro-ph.HE]
45. D. Davis, L.V. White, P.R. Saulson, Utilizing aLIGO glitch classifications to validate gravitational-wave candidates. *Class. Quant. Grav.* **37**(14), 145001 (2020). <https://doi.org/10.1088/1361-6382/ab91e6>. arXiv:2002.09429 [gr-qc]
46. Z. Benkő, T. Bábel, Z. Somogyvári, Model-free detection of unique events in time series. *Sci. Rep.* **12**(1), 227 (2022). <https://doi.org/10.1038/s41598-021-03526-y>. arXiv:2004.11468 [cs.LG]
47. T. Marianer, D. Poznanski, J.X. Prochaska, A semisupervised machine learning search for never-seen gravitational-wave sources. *Mon. Not. R. Astron. Soc.* **500**(4), 5408–5419 (2020). <https://doi.org/10.1093/mnras/staa3550>. arXiv:2010.11949 [astro-ph.IM]
48. M. Cabero, A. Mahabal, J. McIver, GWSkyNet: a real-time classifier for public gravitational-wave candidates. *Astrophys. J. Lett.* **904**(1), 9 (2020). <https://doi.org/10.3847/2041-8213/abc5b5>. arXiv:2010.11829 [gr-qc]
49. S. Jadhav, N. Mukund, B. Gadre, S. Mitra, S. Abraham, Improving significance of binary black hole mergers in Advanced LIGO data using deep learning: confirmation of GW151216. *Phys. Rev. D* **104**(6), 064051 (2021). <https://doi.org/10.1103/PhysRevD.104.064051>. arXiv:2010.08584 [gr-qc]
50. S. Singh, A. Singh, A. Prajapati, K.N. Pathak, Deep learning for estimating parameters of gravitational waves. *Mon. Not. R. Astron. Soc.* **508**(1), 1358–1370 (2021). <https://doi.org/10.1093/mnras/stab2417>. arXiv:2008.06550 [astro-ph.HE]
51. T.C. Abbott, E. Buffaz, N. Vieira, M. Cabero, D. Haggard, A. Mahabal, J. McIver, GWSkyNet-multi: a machine-learning multiclass classifier for LIGO-Virgo public alerts. *Astrophys. J.* **927**(2), 232 (2022). <https://doi.org/10.3847/1538-4357/ac5019>. arXiv:2111.04015 [astro-ph.IM]

52. P. Chaturvedi, A. Khan, M. Tian, E.A. Huerta, H. Zheng, Inference-optimized AI and high performance computing for gravitational wave detection at scale. *Front. Artif. Intell.* **5**, 828672 (2022). <https://doi.org/10.3389/frai.2022.828672>. arXiv:2201.11133 [gr-qc]
53. S. Choudhary, A. More, S. Suyamprakasam, S. Bose, SiGMA-Net: deep learning network to distinguish binary black hole signals from short-duration noise transients (2022). arXiv:2202.08671 [gr-qc]
54. S. Choudhary, S. Bose, P. Joshi, S. Dhurandhar, Improved binary black hole searches through better discrimination against noise transients (2022). arXiv:2212.02026 [gr-qc]
55. V. Boudart, Convolutional neural network to distinguish glitches from minute-long gravitational wave transients. *Phys. Rev. D* **107**(2), 024007 (2023). <https://doi.org/10.1103/PhysRevD.107.024007>. arXiv:2210.04588 [gr-qc]
56. S. Bini, G. Vedovato, M. Drago, F. Salemi, G.A. Prodi, An autoencoder neural network integrated into gravitational-wave burst searches to improve the rejection of noise transients. *Class. Quant. Grav.* **40**(13), 135008 (2023). <https://doi.org/10.1088/1361-6382/acd981>. arXiv:2303.05986 [gr-qc]
57. T.S. Fernandes, S.J. Vieira, A. Onofre, J. Calderón Bustillo, A. Torres-Forné, J.A. Font, Convolutional Neural Networks for the classification of glitches in gravitational-wave data streams (2023). arXiv:2303.13917 [gr-qc]
58. S. Jadhav, M. Shrivastava, S. Mitra, Towards a robust and reliable deep learning approach for detection of compact binary mergers in gravitational wave data (2023). arXiv:2306.11797 [gr-qc]
59. N. Shah, A.M. Kneer, J. McIver, D. Stenning, Waves in a forest: a random forest classifier to distinguish between gravitational waves and detector glitches (2023). arXiv:2306.13787 [gr-qc]
60. A. Trovato, E. Chassande-Mottin, M. Beijger, R. Flamary, N. Courty, Neural network time-series classifiers for gravitational-wave searches in single-detector periods (2023). arXiv:2307.09268 [gr-qc]
61. S. Jarov, S. Thiele, S. Soni, J. Ding, J. McIver, R. Ng, R. Hatoya, D. Davis, A new method to distinguish gravitational-wave signals from detector noise transients with Gravity Spy (2023). arXiv:2307.15867 [gr-qc]
62. S. Alvarez-Lopez, A. Liyanage, J. Ding, R. Ng, J. McIver, GSPyNetTree: a signal-vs-glitch classifier for gravitational-wave event candidates (2023). arXiv:2304.09977 [gr-qc]
63. S. Jarov, S. Thiele, S. Soni, J. Ding, J. McIver, R. Ng, R. Hatoya, D. Davis, A new method to distinguish gravitational-wave signals from detector noise transients with Gravity Spy. arXiv e-prints (2023). arXiv:2307.15867 [gr-qc]
64. A. Torres-Forné, E. Cuoco, J.A. Font, A. Marquina, Application of dictionary learning to denoise LIGO's blip noise transients. *Phys. Rev. D* **102**(2), 023011 (2020). <https://doi.org/10.1103/PhysRevD.102.023011>. arXiv:2002.11668 [gr-qc]
65. J. Merritt, B. Farr, R. Hur, B. Edelman, Z. Doctor, Transient glitch mitigation in Advanced LIGO data. *Phys. Rev. D* **104**(10), 102004 (2021). <https://doi.org/10.1103/PhysRevD.104.102004>. arXiv:2108.12044 [gr-qc]
66. G. Ashton, Gaussian processes for glitch-robust gravitational-wave astronomy. *Mon. Not. R. Astron. Soc.* **520**(2), 2983–2994 (2023). <https://doi.org/10.1093/mnras/stad341>. arXiv:2209.15547 [gr-qc]
67. A. Longo, S. Bianchi, G. Valdes, N. Arnaud, W. Plastino, Daily monitoring of scattered light noise due to microseismic variability at the Virgo interferometer. *Class. Quant. Grav.* **39**(3), 035001 (2022). <https://doi.org/10.1088/1361-6382/ac4117>. arXiv:2112.06046 [astro-ph.IM]
68. R.E. Colgan, Z. Márka, J. Yan, I. Bartos, J.N. Wright, S. Márka, Detecting and diagnosing terrestrial gravitational-wave mimics through feature learning (2022). arXiv:2203.05086 [astro-ph.IM]
69. J. Glanzer, S. Soni, J. Spoon, A. Effler, G. González, Noise in the LIGO Livingston Gravitational Wave Observatory due to Trains (2023). arXiv:2304.07477 [astro-ph.IM]
70. M. Lopez, V. Boudart, K. Buijsman, A. Reza, S. Caudill, Simulating transient noise bursts in LIGO with generative adversarial networks. *Phys. Rev. D* **106**(2), 023027 (2022). <https://doi.org/10.1103/PhysRevD.106.023027>. arXiv:2203.06494 [astro-ph.IM]
71. J. Powell, L. Sun, K. Gereb, P.D. Lasky, M. Dollmann, Generating transient noise artifacts in gravitational-wave detector data with generative adversarial networks (2022). arXiv:2207.00207 [astro-ph.IM]
72. T. Dooney, S. Bromuri, L. Curier, DVGAN: stabilize Wasserstein GAN training for time-domain gravitational wave physics (2022). arXiv:2209.13592 [astro-ph.IM]
73. D. George, H. Shen, E.A. Huerta, Classification and unsupervised clustering of LIGO data with Deep Transfer Learning. *Phys. Rev. D* **97**(10), 101501 (2018). <https://doi.org/10.1103/PhysRevD.97.101501>. arXiv:1706.07446 [gr-qc]
74. M. Cavaglia, K. Staats, T. Gill, Finding the origin of noise transients in LIGO data with machine learning. *Commun. Comput. Phys.* **25**(4), 963–987 (2019). <https://doi.org/10.4208/cicp.OA-2018-0092>. arXiv:1812.05225 [physics.data-an]
75. S. Sankarapandian, B. Kulis, β -Annealed Variational Autoencoder for glitches (2021). arXiv:2107.10667 [cs.LG]
76. Y. Sakai et al., Unsupervised learning architecture for classifying the transient noise of interferometric gravitational-wave detectors. *Sci. Rep.* **12**(1), 9935 (2022). <https://doi.org/10.1038/s41598-022-13329-4>. arXiv:2111.10053 [gr-qc]
77. J. Yan, A.P. Leung, D.C.Y. Hui, On improving the performance of glitch classification for gravitational wave detection by using Generative Adversarial Networks. *Mon. Not. R. Astron. Soc.* **515**(3), 4606–4621 (2022). <https://doi.org/10.1093/mnras/stac1996>. arXiv:2207.04001 [astro-ph.HE]
78. Y. Sakai et al., Training Process of Unsupervised Learning Architecture for Gravity Spy Dataset (2022). <https://doi.org/10.1002/andp.202200140>. arXiv:2208.03623 [gr-qc]
79. A.E. Tolley, G.S. Cabourn Davies, I.W. Harry, A.P. Lundgren, ArchEnemy: removing scattered-light glitches from gravitational wave data. *Class. Quant. Grav.* **40**(16), 165005 (2023). <https://doi.org/10.1088/1361-6382/ace22f>. arXiv:2301.10491 [gr-qc]
80. M.J. Raddick, G. Bracey, P.L. Gay, C.J. Lintott, P. Murray, K. Schawinski, A.S. Szalay, J. Vandenberg, Galaxy Zoo: exploring the motivations of citizen science volunteers. *Astron. Educ. Rev.* **9**, 010103 (2010). <https://doi.org/10.3847/AER2009036>. arXiv:0909.2925 [astro-ph.IM]
81. D. Rotman, J. Hammock, J. Preece, C.L. Boston, D.L. Hansen, A. Bowser, Y. He, Does motivation in citizen science change with time and culture? in *The Companion Publication of the 17th ACM Conference*, pp. 229–232 (2014). <https://doi.org/10.1145/2556420.2556492>
82. C. Jackson, Characterizing novelty as a motivator in online citizen science. PhD thesis, Syracuse University (2019)
83. T.K. Lee, K. Crowston, M. Harandi, C. Østerlund, G. Miller, Appealing to different motivations in a message to recruit citizen scientists: results of a field experiment. *J. Sci. Commun.* **17**(1), 02 (2018)
84. C.B. Jackson, C. Østerlund, K. Crowston, M. Harandi, L. Trouille, Shifting forms of engagement: volunteer learning in online citizen science. *Proc. ACM Hum. Comput. Interact.* **4**(CSCW1), 1–19 (2020)
85. H. Sauermann, C. Franzoni, Crowd science user contribution patterns and their implications. *Proc. Natl. Acad. Sci. U. S. A.* **112**(3), 679–684 (2015). <https://doi.org/10.1073/pnas.1408907112>
86. F. Rohden, C. Kullenberg, N. Hagen, D. Kasperowski, Tagging, pinging and linking—user roles in virtual citizen science forums. *Citiz. Sci. Theory Pract.* **4**(1), 10008–13 (2019). <https://doi.org/10.5334/cstp.181>
87. H. Spiers, A. Swanson, L. Fortson, B. Simmons, L. Trouille, S. Blickhan, C. Lintott, Everyone counts? Design considerations in online citizen science. *J. Sci. Commun.* **18**(1) (2019)

88. C. Jackson, K. Crowston, C. Østerlund, M. Harandi, Folksonomies to support coordination and coordination of folksonomies. *Comput. Supported Coop. Work (CSCW)* **27**, 647–678 (2018)
89. B. Ekström, C. Jackson, C. Østerlund, Tracing hyperlinks: How to support different forms of presence and knowledge production in an online citizen science community? in *Good Relations: Practices and Methods in Unequal and Uncertain Worlds. 4S Annual Meeting*. Toronto, Canada (2021)
90. M. Harandi, Occasional groups in crowdsourcing platforms. PhD thesis, Syracuse University (2021)
91. C.D. Styliński, K. Peterman, T. Phillips, J. Linhart, R. Becker-Klein, Assessing science inquiry skills of citizen science volunteers: a snapshot of the field. *Int. J. Sci. Educ. Part B* **10**(1), 77–92 (2020). <https://doi.org/10.1080/21548455.2020.1719288>
92. K. Crowston, C. Jackson, I. Corieri, C. Østerlund, Design principles for background knowledge to enhance learning in citizen science, in *International Conference on Information* (Springer, 2023), pp. 563–580
93. I. Corieri, C. Østerlund, K. Crowston, C.B. Jackson, Advanced work on user-generated content systems: theory-driven method development, in *iConference 2023 Proceedings* (2023)
94. B.P. Abbott et al., GW150914: the Advanced LIGO detectors in the era of first discoveries. *Phys. Rev. Lett.* **116**(13), 131103 (2016). <https://doi.org/10.1103/PhysRevLett.116.131103>. arXiv:1602.03838 [gr-qc]
95. D.V. Martynov et al., Sensitivity of the Advanced LIGO detectors at the beginning of gravitational wave astronomy. *Phys. Rev. D* **93**(11), 112004 (2016). <https://doi.org/10.1103/PhysRevD.93.112004>. arXiv:1604.00439 [astro-ph.IM]. [Addendum: *Phys. Rev. D* **97** 059901 (2018)]
96. B.P. Abbott et al., GW170104: Observation of a 50-Solar-Mass Binary Black Hole Coalescence at Redshift 0.2. *Phys. Rev. Lett.* **118**(22), 221101 (2017). <https://doi.org/10.1103/PhysRevLett.118.221101>. arXiv:1706.01812 [gr-qc]. [Erratum: *Phys. Rev. Lett.* **121** 129901 (2018)]
97. A. Buikema et al., Sensitivity and performance of the Advanced LIGO detectors in the third observing run. *Phys. Rev. D* **102**(6), 062003 (2020). <https://doi.org/10.1103/PhysRevD.102.062003>. arXiv:2008.01301 [astro-ph.IM]
98. K. Crowston, C. Østerlund, T.K. Lee, C. Jackson, M. Harandi, S. Allen, S. Bahaadini, S. Coughlin, A.K. Katsaggelos, S.L. Larson et al., Knowledge tracing to model learning in online citizen science projects. *IEEE Trans. Learn. Technol.* **13**(1), 123–134 (2019)
99. P. Nguyen et al., Environmental noise in Advanced LIGO detectors. *Class. Quant. Grav.* **38**(14), 145001 (2021). <https://doi.org/10.1088/1361-6382/ac011a>. arXiv:2101.09935 [astro-ph.IM]
100. M. Razzano, F. Di Renzo, F. Fidecaro, G. Hemming, S. Katsanevas, GWitchHunters: machine learning and citizen science to improve the performance of gravitational wave detector. *Nucl. Instrum. Methods A* **1048**, 167959 (2023). <https://doi.org/10.1016/j.nima.2022.167959>. arXiv:2301.05112 [gr-qc]
101. M. Walmsley, L. Smith, C. Lintott, Y. Gal, S. Bamford, H. Dickinson, L. Fortson, S. Kruk, K. Masters, C. Scarlata, B. Simmons, R. Smethurst, D. Wright, Galaxy Zoo: probabilistic morphology through Bayesian CNNs and active learning. *Mon. Not. R. Astron. Soc.* **491**(2), 1554–1574 (2020). <https://doi.org/10.1093/mnras/stz2816>. arXiv:1905.07424 [astro-ph.GA]
102. C.N. Cardamone et al., Galaxy Zoo green peas: discovery of a class of compact extremely star-forming galaxies. *Mon. Not. R. Astron. Soc.* **399**, 1191–1205 (2009). <https://doi.org/10.1111/j.1365-2966.2009.15383.x>. arXiv:0907.4155 [astro-ph.CO]
103. N.L. Eisner, O. Barragán, C. Lintott, S. Aigrain, B. Nicholson, T.S. Boyajian, S. Howell, C. Johnston, B. Lakeland, G. Miller, A. McMaster, H. Parviainen, E.J. Safron, M.E. Schwamb, L. Trouille, S. Vaughan, N. Zicher, C. Allen, S. Allen, M. Bouslog, C. Johnson, M.N. Simon, Z. Wolfenbarger, E.M.L. Baeten, D.M. Bundy, T. Hoffman, Planet hunters TESS II: findings from the first two years of TESS. *Mon. Not. R. Astron. Soc.* **501**(4), 4669–4690 (2021). <https://doi.org/10.1093/mnras/staa3739>. arXiv:2011.13944 [astro-ph.EP]