Classifying the unknown: Discovering novel gravitational-wave detector glitches using similarity learning

S. Coughlin,[1,2,*] S. Bahaadini,[3] N. Rohani,[3] M. Zevin,[2] O. Patane,[4] M. Harandi,[5] C. Jackson,[5] V. Noroozi,[6] S. Allen,[7] J. Areeda,[4] M. Coughlin,[8] P. Ruiz,[3] C. P. L. Berry,[2] K. Crowston,[5] A. K. Katsaggelos,[3] A. Lundgren,[9] C. Østerlund,[5] J. R. Smith,[4] L. Trouille,[7] and V. Kalogera[2]

[1]*Physics and Astronomy, Cardiff University, Cardiff CF10 2FH, United Kingdom*
[2]*Center for Interdisciplinary Exploration & Research in (CIERA), Northwestern University, Evanston, Illinois 60208, USA*
[3]*Electrical Engineering and Computer Science, Northwestern University, Evanston, Illinois 60201, USA*
[4]*Department of Physics, California State University Fullerton, Fullerton, California 92831, USA*
[5]*School of Information Studies, Syracuse University, Syracuse, New York 13210, USA*
[6]*Department of Computer Science, University of Illinois at Chicago, Illinois 60607, USA*
[7]*Adler Planetarium, Chicago, Illinois 60605, USA*
[8]*Division of Physics, Math, and Astronomy, California Institute of Technology, Pasadena, California 91125, USA*
[9]*Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth PO1 2UP, United Kingdom*

The observation of gravitational waves from compact binary coalescences by LIGO and Virgo has begun a new era in astronomy. A critical challenge in making detections is determining whether loud transient features in the data are caused by gravitational waves or by instrumental or environmental sources. The citizen-science project *Gravity Spy* has been demonstrated as an efficient infrastructure for classifying known types of noise transients (glitches) through a combination of data analysis performed by both citizen volunteers and machine learning. We present the next iteration of this project, using similarity indices to empower citizen scientists to create large data sets of unknown transients, which can then be used to facilitate supervised machine-learning characterization. This new evolution aims to alleviate a persistent challenge that plagues both citizen-science and instrumental detector work: the ability to build large samples of relatively rare events. Using two families of transient noise that appeared unexpectedly during LIGO's second observing run, we demonstrate the impact that the similarity indices could have had on finding these new glitch types in the Gravity Spy program.

I. INTRODUCTION

The recent detection of gravitational waves (GWs) from the inspiral and merger of binary black holes [1–6] and binary neutron stars [6,7] by the ground-based interferometric detectors Advanced LIGO [8] and Advanced Virgo [9] has begun the era of GW astronomy. The analysis of GW data is complicated by the presence of noise transients of both instrumental and environmental origin known as *glitches*. These noise transients are caused by a wide variety of phenomena, including up-conversion of ground motion into the optical system, laser power fluctuations, and magnetic fields at the local site [10]; however, many persistent noise transients have no known cause, and are not coincidentally witnessed by any auxiliary monitoring channel. Glitches can impact the detection of signals, as

they can be confused with astrophysical signals with short durations or significant theoretical uncertainties, and, if occurring at the same time as a GW (as was the case for the first binary neutron star observation [7]), they make it more challenging to accurately infer the properties of the astrophysical source [11,12]. Over the next decade [13], sensitivity improvements to the GW detectors are expected due to increased laser power, improved optical coatings, and improved seismic isolation. These improvements will also lead to an altered sensitivity to instrumental and environmental noise transients, as well as to changes in the character of some types of glitches [14]. As the Advanced LIGO and Advanced Virgo detectors evolve, glitches will continue to remain a challenge for GW analysis.

Significant efforts are made to characterize and identify these noise transients so that the times that they are present can be removed from analyzed data, and, if possible, their causes eliminated altogether. This can be accomplished

---

[*]scottcoughlin2014@u.northwestern.edu

through the use of algorithms that use auxiliary sensors placed in and around the GW detectors to identify those sensors that are highly correlated witnesses with a particular type of glitch [15–17]. The idea is to identify activity in witness sensors that can be attributed to the glitch, thereby requiring the removal of as little GW data as possible while also successfully removing the classes of data transients. GW signals will not appear in these auxiliary sensors, so vetoing transients based on auxiliary sensor data does not erroneously remove true GW signals. An essential first step in diagnosing the causes of glitches is identifying them as a feature in the data.

There has been significant recent interest in the application of machine learning (ML) to glitch identification [18–23], to the identification of correlated witness channels [16,17], and to GW searches [24–26]. Though the classification of glitches through ML approaches has shown promise, these approaches suffer from some shortcomings. First, the supervised ML methods, where previously known classes of transients are given as a training set to the algorithm, have no immediate way to identify other classes also present in the data. Alternatively, unsupervised ML methods, where the algorithm seeks to learn the discriminative features of the data set in order to create its own classes or clusters of similar data, have the downside of decoupling the analysis from the understanding of how known classes relate to the detector. Moreover, unsupervised methods inevitably suffer from the need to validate the self-identified classes, as the clusters are far from exclusive because the features the algorithm learns from the unlabeled data set are not discriminative enough. Both supervised and unsupervised ML techniques have merits, but neither is a perfect solution.

In an effort to address the glitch classification problem, we previously introduced *Gravity Spy* [14].[1] This combines the crowd-sourcing power of citizen science with the rapid classification ability of ML [27] to support the characterization of glitches in GW data. Gravity Spy is hosted on Zooniverse, a leading online platform that has enabled over 1.5 million citizen scientists to analyze scientific data. Gravity Spy users are asked to classify time-frequency plots depicting glitches into one of a number of classes. The large number of people supporting this work provides data sets of known glitch classes, which are then used as training sets for the ML algorithms. The ML algorithms can then rapidly classify the entire data set of known glitches. These statistically pure data sets are then used for the purpose of long-term trend studies as well as targeted auxiliary channel follow-up, e.g., comparing humidity at the detector with the rate of the blip glitch [28]. Although the classification and verification of known classes has proven effective in Gravity Spy, it remains challenging to collect sufficient numbers of novel glitches to identify new classes. We note that unsupervised clustering has been shown to classify new types of glitches, e.g., "Reverse Chirp," shown in Fig. C2 of Ref. [22].

To solve this problem we employ techniques from transfer learning. Transfer learning applies the knowledge from a labeled data set to an unlabeled data set with different features. In this case, we are interested in transferring the knowledge of what makes the known glitch classes in Gravity Spy similar and different from each other to the domain of images that do not belong to any known class. Although this method proves useful in helping the algorithm extract more discriminating features that make for cleaner clustering of the unlabeled data, they are still too inclusive to confidently contain a single new class of glitch. Therefore, combining the feature space obtained through transfer learning techniques with human-controlled clustering of this space may prove the most effective way to rapidly identify new glitch classes.

We introduce a new method for the rapid identification of novel transients that combines techniques within the field of transfer learning with the crowd-sourcing power of Gravity Spy. In Sec. II, we discuss the specifics of our transfer learning algorithm. We then discuss the new proposed infrastructure for Gravity Spy in Sec. III. In Sec. IV, we highlight the impact the proposed methodology could have had on discovering two sets of new LIGO glitches from the second observing (O2) run. In Sec. IV A, we summarize the impact of different settings of the transfer learning algorithm on the discriminative ability of the feature space. In Sec. V, we discuss future iterations of the Gravity Spy project and its role in GW detector characterization.

II. TRANSFER LEARNING

Transfer learning applies knowledge obtained from a model that was trained on one data set to another data set. Specifically, for our method, we hope to transfer knowledge about what makes the spectrograms of the known Gravity Spy classes similar and different from each other to the unlabeled Gravity Spy glitches. We anticipate that this knowledge will enable a better clustering of these glitches that will lead to the discovery of new classes.

To accomplish this goal, we must first train an algorithm designed to model the similarity and differences between images on the known set of Gravity Spy glitches. For this analysis, we use the transfer-learning algorithm DIRECT [29] to quantify the similarity between Gravity Spy images. In short, this algorithm solves for a nonlinear embedding function f_θ, i.e., the discriminative feature space, by using a deep neural network. Using pairs of labeled images as input, the neural network is trained by solving for the f_θ that minimizes the function

$$
\begin{aligned}
\mathcal{L} &= \sum_{i=1}^{N} l(y^i, x_1^i, x_2^i) \\
&= \sum_{i=1}^{N} y^i \mathrm{dist}(f_\theta(x_1^i), f_\theta(x_2^i)) \\
&\quad + (1 - y^i) \max\{0, m - \mathrm{dist}(f_\theta(x_1^i), f_\theta(x_2^i))\}. \quad (1)
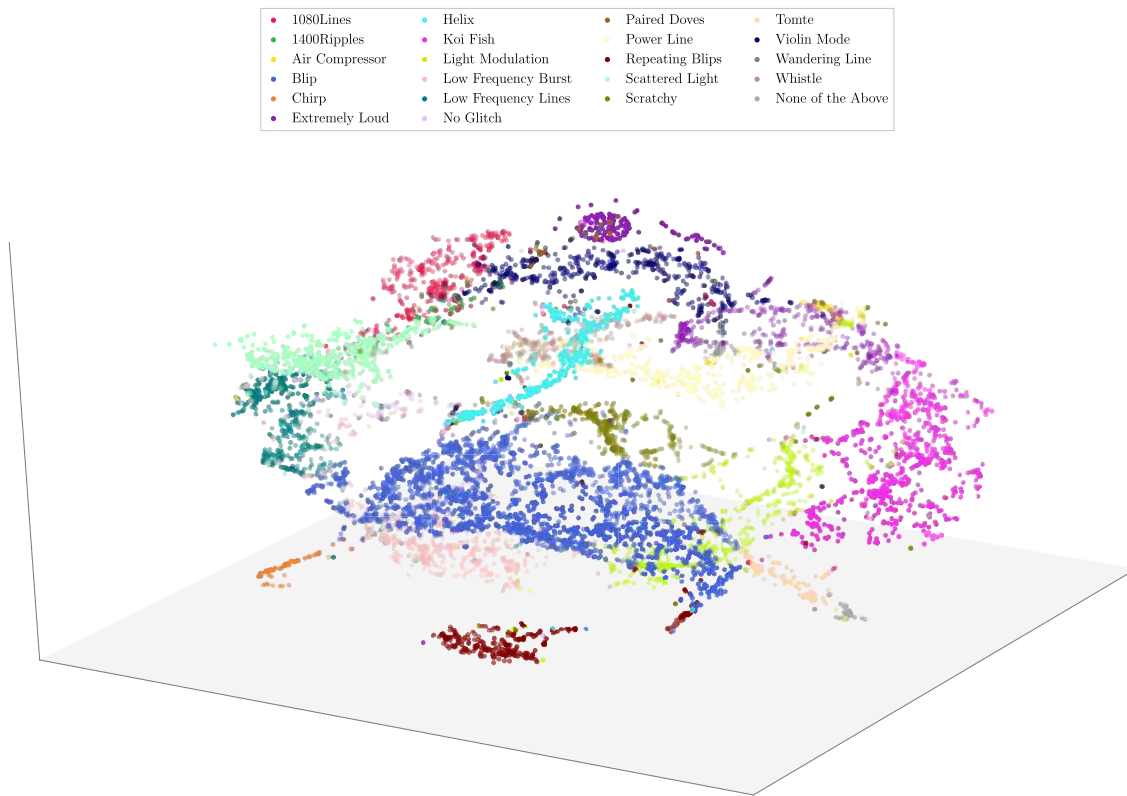\end{aligned}
$$

FIG. 1. Visual representation of the training set in the DIRECT feature space using the t-distributed stochastic neighbor embedding (t-SNE) statistic. This metric is purely designed to project groups of samples in the N-dimensional feature space into three dimensions and has no physical meaning.

Here N is the number of training pairs; x_1^i and x_2^i are the first and second items of the ith pair; y^i is the binary label of the ith pair, which is one when the two items of the pair belong to the same class and zero when they belong to different classes; dist is a distance function (such as Euclidean or cosine); and m is the margin that is used to bound the distance between the items of pairs from different classes. A convolutional neural network models the nonlinear function f_θ by adding a fully connected dense layer onto the pretrained VGG16 network [30]. The VGG16 network consists of 13 convolutional layers and two fully connected layers and was pretrained on the ImageNet [31] database of images. We use the cosine distance metric as our distance function,[2] and to train the model we use the Gravity Spy training set described in Ref. [32]. Each glitch is portrayed as four spectrograms with different temporal durations. These are generated using GWPY [33]. We take these four images to create a single merged image for each glitch, identical to the input currently used for the convolutional neural network classifications in Gravity Spy [32]. By propagating through the DIRECT network described above, the pixel data of the

input image is mapped to a smaller, 200-dimensional feature space. The dimension of the feature space is fixed at 200 based on Fig. 2 of Ref. [29]. In Fig. 1 (cf. Fig. 3 of Ref. [32]) we show a visual representation of the training set in the DIRECT feature space using the t-distributed stochastic neighbor embedding statistic (t-SNE) [34]. As can be seen samples from the same glitch class are put closer to each other while samples of different classes are far from each other. Such a property is called *discriminative* feature representation. Having trained this model on the known set of glitches, we can now apply it to the unlabeled glitches so that in this new, more discriminative feature space we can cluster similar images together and find new classes.

Having established a means of clustering glitches, in the next section, we describe how we use the discriminative feature space obtained using DIRECT on the Gravity Spy data set to empower volunteers to build large data sets of unknown glitches.

III. IDENTIFYING NOVEL GLITCHES

In our previous work, we relied on the None-of-the-Above classification to identify glitches from previously unknown classes, and the volunteer *Talk* forum (a thread of comments on the image from other volunteers) to

---

[2]The cosine distance between two vectors $\vec{a}$ and $\vec{b}$ is $1 - \vec{a} \cdot \vec{b}/(|\vec{a}||\vec{b}|)$.
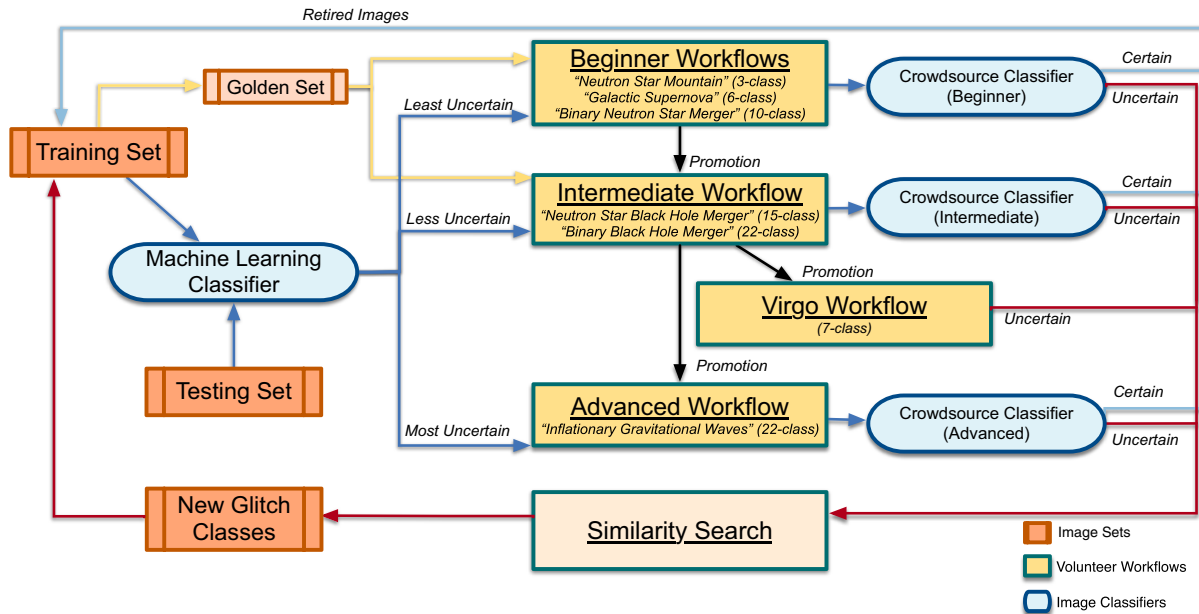
FIG. 2. New infrastructure proposal for Gravity Spy. This design differs from that described in Ref. [14] by facilitating the direct follow-up of single examples of unknown transients through the similarity search algorithm. This is in contrast to the reliance on the None-of-the-Above classifications for filtering out novel glitches from the data set.

consolidate examples in order to develop training sets and add new classes to the supervised model. As the volume of data increases, this design will prove ineffective as a user would have to go through too many classifications before seeing multiple examples of a novel glitch.

To this end, we are introducing a similarity search tool to the Gravity Spy infrastructure. This tool uses the feature space output for every Gravity Spy glitch, which is obtained using the DIRECT algorithm described in Sec. II, to enable users to take a single example of a glitch and perform a search querying for glitches in the data which are morphologically similar to the example. In order to determine the similarity, we subtract one from the same cosine distance metric described in Sec. II. In this case, identical images have a distance metric of 1, while orthogonal images, that is two images that share no common components in DIRECT feature space, have a distance metric close to 0. This metric has the advantage of being efficient to evaluate and not being affected by the overall scaling of the input vectors. Specifically, this tool is introduced in the form of a supplementary web service called gravityspytools.[3] The new infrastructure design is highlighted in Fig. 2.

We anticipate that this tool, which outputs results in the form of *Collections* (galleries of user-selected Gravity Spy glitches), will reduce the size of the data set to be searched for examples of a new class, such that building substantial training sets for unknown glitch types is manageable by a single user. In the next section, we demonstrate how this

tool could have proved useful in the rapid identification of two glitch classes that appeared in O2 which were *not* previously included in the Gravity Spy classification.

IV. RESULTS

We now highlight the application of the similarity search tool on data from O2. Specifically, we assess the impact the similarity search tool could have had on the identification of two glitch classes that appeared during O2: the Water Jet [35] and the Raven Peck [36] glitches. The Water Jet glitch was caused by local seismic noise which resulted in loud bangs near the input optics, and the Raven Peck glitch was caused by ravens pecking on ice built up along vent lines transporting nitrogen outside of the detector [10]. The resulting time-frequency morphologies of these glitches as they appear in the GW data channel can be seen in the top panel of Fig. 3. These glitches occurred in the LIGO-Hanford detector. The Raven Peck glitch is found in Gravity Spy data between 14 April and 9 August, 2017, and the Water Jet glitch in data between 4 January and 28 May, 2017. Over these durations, there are a total of 13 513 and 26 871 Gravity Spy glitches, respectively.

We use these two glitches because they highlight the strengths and weaknesses of the DIRECT algorithm, and they allow us to emphasize the importance of incorporating crowd-sourcing methods into the identification of new glitch classes. As DIRECT is a transfer learning algorithm, it is only able to employ concepts of similarity and difference learned from the training set to the unlabeled data. If the distinguishing characteristic of a particular glitch is not also something present or extractable from the

---

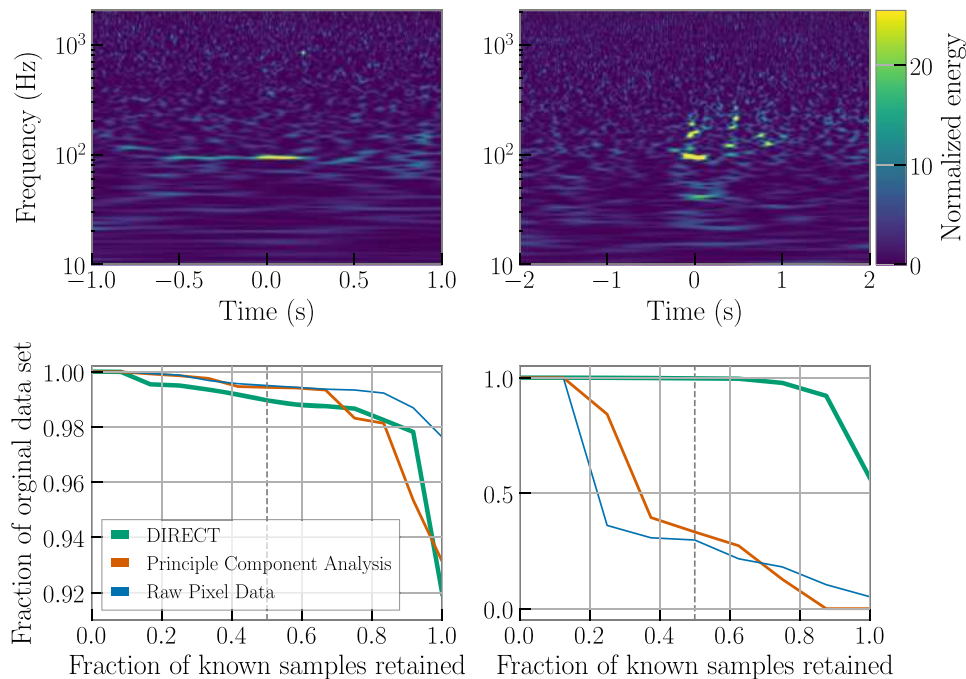[3]gravityspytools.ciera.northwestern.edu.

FIG. 3.   Top: Nominal examples of the Raven Peck (left) and Water Jet (right) glitches. Bottom left: The fraction of known Raven Peck samples that have a higher similarity score than a given percentage of other data set samples when calculating the similarity to a single known Raven Peck glitch. For example, while retaining 50.0% of known Raven Peck glitches, we can disregard about 99.0% percent of the other data set samples, increasing the purity of the set to be examined by the user. Bottom right: Same for Water Jet glitches. Similarly, while retaining 50.0% of known Water Jet glitches, we can disregard about 99.0% percent of the other data set samples.

training set, it may be more difficult for the algorithm to find other examples easily. This is demonstrated well here because the Raven Peck glitch is most prominently defined as a line feature which *is* present in many of the Gravity Spy classes used in training (such as Power Line, Low Frequency Line and 1080 Line); on the other hand, the unique aspect of the Water Jet glitch is the subtle frequency decay that occurs after the initial pulse, which *is not* obviously extractable from glitches in the training set.

We demonstrate the ability of the similarity search tool to organize testing images by their similarity to a queried glitch. As was done in the DIRECT paper [29], we also compare finding similar images with DIRECT to more straightforward approaches such as using the raw pixel data or doing a principle components analysis (PCA). The bottom panel of Fig. 3 shows the fraction of known samples that have a higher similarity score than a given percentage of the other data set samples. For example, while retaining 50.0% of known Raven Peck glitches, we can remove about 99.9% percent of the other data set samples, increasing the purity of the set to be examined by the user. For the same glitch, the raw pixel data approach and the PCA approach perform similarly to the Raw Pixels approach doing best at near 100.0%. For Water Jet glitches, DIRECT also gives a similar performance retaining 50.0% of known samples as it did for the Raven Peck. However, if a retention rate of 100.0% of the known samples is desired,

the data set reduction rate for Raven Peck is 92.0% compared to 55.0% for the Water Jet glitch. For this glitch, the methods of raw pixel data and PCA prove ineffective. For a retention rate of 50.0% only about 30.0% of samples in the data set have lower similarity scores. We believe these examples represent a challenging and less challenging task, respectively, for the model, and in both cases DIRECT performs well, and the other approaches fail to be effective in the case of the Water Jet glitch. We anticipate that the reduction in the size of the original data set combined with the retention rate of similar samples will be significant enough that a single user can produce large data sets of novel glitches.

A. Different configurations

To test the best training and setting configuration for DIRECT, we tried two different activation layers, tanh and leakyReLU, for the custom fully connected layer that DIRECT adds to the VGG16 model. In addition, we varied the number of training rounds and the number of pairs of similar and dissimilar images that are drawn from the training set each time. As training this model can be expensive because of the possible pairs of images one can produce from the training set, it is critical to understand what the minimal expense is that still produces an effective model. To judge effectiveness, we quote the percentage reduction in the data set samples at which we still retain

TABLE I. The fraction of the original data set with similar scores lower than the similarity score of 50.0% of other known Raven Peck (left) and Water Jet glitches (right). Columns refer to different choices in the activation layer used in the dense layer of the model and the number of training rounds where each round draws a new set of X number of similar and dissimilar pairs. In bold is the configuration(s) that yielded the best reduction vs retention rate for both glitches.

Act. Layer	Train Rnds	10 000 pr.	50 000 pr.	100 000 pr.
tanh	30	(0.93, 0.78)	(0.96, 0.92)	(0.94, 0.87)
	100	(0.90, 0.85)	(0.93, 0.82)	(0.94, 0.72)
	200	(0.93, 0.82)	(0.87, 0.80)	(0.93, 0.93)
leakyReLU	30	(0.96, 0.94)	(0.95, 0.92)	(0.95, 0.91)
	100	(0.95, 0.91)	(0.96, 0.91)	(0.96, 0.94)
	200	(0.95, 0.91)	(0.97, 0.92)	(0.96, 0.92)

50.0% of other known Raven Peck and Water Jet glitches, respectively. This value for each model is shown in Table I. The models using tanh as the activation layer perform worse than those with leakyReLU. We anticipate that this is due to the restricted range allowed by the tanh activation layer, $[-1, 1]$, compared to that of leakyReLU, $(-\infty, \infty)$. Specifically, the distances away from each other similar and dissimilar images can be is restricted in the one case causing the discriminative feature space that is created to be less discerning than the other. In terms of number of training pairs and rounds of training, it appears that increasing each does not lead to significantly improved results. We anticipate that this is due to the fact that most of the Gravity Spy classes are quite distinct from each other, and therefore using or drawing more pairs of images is unnecessary to produce a useful discriminative feature space representation of the data. Therefore, this method can still be effective without an extremely costly training stage.

V. CONCLUSIONS

We have described a novel extension of current GW data transient class identification combining the power of citizen scientists with the latest techniques in ML. In the original paper, we allowed volunteers to classify glitches as None of the Above in order to identify individual subjects which belong to an unknown class. Utilizing DIRECT, we have shown the ability to expedite the identification of new glitch classes compared with this original method, which will be important as we get new data from upcoming observing runs.

Using two noise transients from LIGO's O2 data, the Raven Peck and water Jet glitches, we demonstrated that DIRECT creates a discriminative feature space representation of the Gravity Spy data set such that single examples of each glitch can efficiently lead to the discovery in the data set of other Raven Peck and Water Jet glitches.

We compared DIRECT to simpler approaches such as using the raw pixel data or PCAs to find similar images and found that DIRECT produces either comparable or better results depending on the glitch.

There are a variety of plans for future related research. For example, we can explore the use of other metrics to further the inter- and intraclass separation, thereby identifying separate classes that are otherwise improperly associated. In addition, there are lessons learned that are applicable in other areas of astronomy, in line with the ongoing applications of unsupervised learning to large-scale astronomical surveys (e.g., Refs. [37–40]). For example, the Large Synoptic Survey Telescope (LSST) [41], an 8 meter class telescope being constructed on Cerro Pachon near La Serena, Chile will take millions of images over its lifetime, identifying approximately 100 000 objects each night. Although it is impossible for most of them to be analyzed by astronomers directly, citizen scientists can contribute to the monitoring, classifying, and annotating of spurious and surprising data. The expectation is that LSST, with its unprecedented field of view and rapid cadence, will discover a multitude of astrophysical phenomena, and can benefit from the ability to rapidly identify unique signals. Fast transients which fade on time scales of a few days, such as kilonova, which have not been identified in previous surveys are likely to be found and will constitute a new class of transient for this survey.

In general, the possibilities for further science education with citizen science initiatives, which place students on the edge of the scientific frontier, lie strongly in the identification of previously unknown phenomena. Projects such as these create an environment where not all phenomena are known and understood, in contrast to textbook science lessons, and achieve a more realistic view of the wonder and challenges of science [42–44]. For this reason, these initiatives help provide the foundation for further education in any scientific field, where the goal is to be able to follow a logical account of a problem to a solution, through the creation of a hypothesis, the taking of data, and the eventual explanation to understand the phenomena. Projects like this will have significant educational benefits and will impact the research projects both inside of LIGO and LSST and outside in numerous research groups conducting other astrophysical studies. We anticipate that the combination of more novel ML techniques with web applications will continue to help with the efficacy of this work.

---

[1] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. Lett. **116,** 061102 (2016).

[2] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. Lett. **116,** 241103 (2016).

[3] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. Lett. **118,** 221101 (2017); **121,** 129901(E) (2018).

[4] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. Lett. **119,** 141101 (2017).

[5] B. P. Abbott *et al.* (Virgo and LIGO Scientific Collaborations), Astrophys. J. **851,** L35 (2017).

[6] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), arXiv:1811.12907.

[7] B. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. Lett. **119,** 161101 (2017).

[8] J. Aasi *et al.* (LIGO Scientific Collaboration), Classical Quantum Gravity **32,** 074001 (2015).

[9] F. Acernese *et al.* (Virgo Collaboration), Classical Quantum Gravity **32,** 024001 (2015).

[10] L. K. Nuttall, Phil. Trans. R. Soc. A **376,** 20170286 (2018).

[11] C. Pankow *et al.*, Phys. Rev. D **98,** 084016 (2018).

[12] J. Powell, Classical Quantum Gravity **35,** 155017 (2018).

[13] B. P. Abbott *et al.* (KAGRA, LIGO Scientific, and Virgo Collaborations), Living Rev. Relativity **21,** 3 (2018).

[14] M. Zevin *et al.*, Classical Quantum Gravity **34,** 064003 (2017).

[15] J. R. Smith, T. Abbott, E. Hirose, N. Leroy, D. Macleod, J. McIver, P. Saulson, and P. Shawhan, Classical Quantum Gravity **28,** 235005 (2011).

[16] R. Biswas *et al.*, Phys. Rev. D **88,** 062003 (2013).

[17] M. Cavaglia, K. Staats, and T. Gill, Commun. Comput. Phys. **25,** 963 (2019).

[18] R. Biswas *et al.*, Phys. Rev. D **88,** 062003 (2013).

[19] J. Powell, D. Trifirò, E. Cuoco, I. S. Heng, and M. Cavaglià, Classical Quantum Gravity **32,** 215012 (2015).

[20] J. Powell, A. Torres-Forné, R. Lynch, D. Trifirò, E. Cuoco, M. Cavaglià, I. S. Heng, and J. A. Font, Classical Quantum Gravity **34,** 034002 (2017).

[21] N. Mukund, S. Abraham, S. Kandhasamy, S. Mitra, and N. S. Philip, Phys. Rev. D **95,** 104059 (2017).

[22] D. George, H. Shen, and E. Huerta, Phys. Rev. D **97,** 101501 (2018).

[23] M. Razzano and E. Cuoco, Classical Quantum Gravity **35,** 095016 (2018).

[24] X. Li, W. Yu, and X. Fan, arXiv:1712.00356.

[25] D. George and E. A. Huerta, Phys. Lett. B **778,** 64 (2018).

[26] H. Gabbard, M. Williams, F. Hayes, and C. Messenger, Phys. Rev. Lett. **120,** 141103 (2018).

[27] S. Bahaadini, N. Rohani, S. Coughlin, M. Zevin, V. Kalogera, and A. K. Katsaggelos, in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2017), pp. 2931–2935, DOI: 10.1109/ICASSP.2017.7952693.

[28] D. Davis, L. White, and M. Cabero, Correlations between O2 Blip Glitches and Low Relative Humidity (2018), Advanced LIGO electonic log 41263, https://alog.ligo-wa.caltech.edu/aLOG/index.php?callRep=41263.

[29] S. Bahaadini, V. Noroozi, N. Rohani, S. Coughlin, M. Zevin, and A. K. Katsaggelos, in *2018 25th IEEE International Conference on Image Processing (ICIP)* (2018), pp. 748–752, DOI: 10.1109/ICIP.2018.8451708.

[30] K. Simonyan and A. Zisserman, arXiv:1409.1556.

[31] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, in *2009 IEEE Conference on Computer Vision and Pattern Recognition* (2009), pp. 248–255, DOI: 10.1109/CVPR.2009.5206848.

[32] S. Bahaadini, V. Noroozi, N. Rohani, S. Coughlin, M. Zevin, J. Smith, V. Kalogera, and A. Katsaggelos, Information Sciences (NY) **444,** 172 (2018).

[33] D. Macleod, A. L. Urban, S. Coughlin, T. Massinger, paulaltin, J. Areeda, E. Quintero, and T. G. Badger, gwpy/gwpy: 0.12.2, DOI: 10.5281/zenodo.1470355 (2018).

[34] L. van der Maaten and G. Hinton, J. Mach. Learn. Res. **9,** 2579 (2008).

[35] T. Dent, Severe transient scattering events in darm caused by loud 20–30 Hz disturbances ('thuds') in cs/lvea (2017), Advanced LIGO electronic log 33761, https://alog.ligo-wa.caltech.edu/aLOG/index.php?callRep=33761.

[36] R. Schofield, Why the GW channel detects thirsty black ravens along with colliding black holes (2017), Advanced LIGO electronic log 37630, https://alog.ligo-wa.caltech.edu/aLOG/index.php?callRep=37630.

[37] H. Domínguez Sánchez *et al.* (DES Collaboration), Mon. Not. R. Astron. Soc. **484,** 93 (2019).

[38] A. Khan, E. A. Huerta, S. Wang, and R. Gruendl, arXiv:1812.02183.

[39] H. Domínguez Sánchez, M. Huertas-Company, M. Bernardi, D. Tuccillo, and J. L. Fischer, Mon. Not. R. Astron. Soc. **476,** 3661 (2018).

[40] A. Mahabal, U. Rebbapragada, R. Walters, F. J. Masci, N. Blagorodnova, J. van Roestel, Q.-Z. Ye, R. Biswas, K. Burdge, C.-K. Chang *et al.*, Publ. Astron. Soc. Pac. **131,** 038002 (2019).

[41] Z. Ivezic, J. A. Tyson, R. Allsman, J. Andrew, and R. Angel (LSST Collaboration), Astrophys. J. **873,** 111 (2019).

[42] C. Doyle, Y. Li, M. Luczak-Roesch, D. Anderson, B. Glasson, M. Boucher, C. Brieseman, D. Christenson, and M. Coton, arXiv:1805.00441.

[43] G. Tsueng, A. Kumar, S. M. Nanis, and A. I. Su, Aligning Needs: Integrating Citizen Science Efforts into Schools Through Service Requirements (Cold Spring Harbor Laboratory, 2018), https://www.biorxiv.org/content/early/2019/03/28/304766.

[44] H. R. Shah and L. R. Martinez, J. Microbiol. Biol. Educ. **17,** 17 (2016).