



Using Dark Energy Explorers and Machine Learning to Enhance the Hobby–Eberly Telescope Dark Energy Experiment

Lindsay R. House^{1,10}, Karl Gebhardt¹, Keely Finkelstein¹, Erin Mentuch Cooper^{1,2}, Dustin Davis¹, Robin Ciardullo^{3,4}, Daniel J Farrow⁵, Steven L. Finkelstein¹, Caryl Gronwall^{3,4}, Donghui Jeong^{3,4}, L. Clifton Johnson^{6,7}, Chenxu Liu^{1,8}, Benjamin P. Thomas¹, and Gregory Zeimann⁹

¹ Department of Astronomy, The University of Texas at Austin, 2515 Speedway Boulevard, Austin, TX 78712, USA; lindsay.r.house@gmail.com

² McDonald Observatory, The University of Texas at Austin, 2515 Speedway Boulevard, Austin, TX 78712, USA

³ Department of Astronomy & Astrophysics, The Pennsylvania State University, University Park, PA 16802, USA

⁴ Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, PA 16802, USA

⁵ Centre of Excellence for Data Science, Artificial Intelligence and Modelling (DAIM), University of Hull, Cottingham Road, Kingston-upon-Hull HU6 7RX, UK

⁶ Adler Planetarium, 1300 S. DuSable Lake Shore Drive, Chicago, IL 60605, USA

⁷ Center for Interdisciplinary Exploration and Research in Astrophysics (CIERA) and Department of Physics and Astronomy, Northwestern University, 1800 Sherman Avenue, Evanston, IL 60201, USA

⁸ South-Western Institute for Astronomy Research, Yunnan University, Kunming, Yunnan, 650500, People's Republic of China

⁹ Hobby Eberly Telescope, The University of Texas at Austin, Austin, TX, 78712, USA

Received 2023 February 21; revised 2023 April 5; accepted 2023 April 13; published 2023 June 13

Abstract

We present analysis using a citizen science campaign to improve the cosmological measures from the Hobby–Eberly Telescope Dark Energy Experiment (HETDEX). The goal of HETDEX is to measure the Hubble expansion rate, $H(z)$, and angular diameter distance, $D_A(z)$, at $z = 2.4$, each to percent-level accuracy. This accuracy is determined primarily from the total number of detected Ly α emitters (LAEs), the false positive rate due to noise, and the contamination due to [O II] emitting galaxies. This paper presents the citizen science project, Dark Energy Explorers (<https://www.zooniverse.org/projects/erinmc/dark-energy-explorers>), with the goal of increasing the number of LAEs and decreasing the number of false positives due to noise and the [O II] galaxies. Initial analysis shows that citizen science is an efficient and effective tool for classification most accurately done by the human eye, especially in combination with unsupervised machine learning. Three aspects from the citizen science campaign that have the most impact are (1) identifying individual problems with detections, (2) providing a clean sample with 100% visual identification above a signal-to-noise cut, and (3) providing labels for machine-learning efforts. Since the end of 2022, Dark Energy Explorers has collected over three and a half million classifications by 11,000 volunteers in over 85 different countries around the world. By incorporating the results of the Dark Energy Explorers, we expect to improve the accuracy on the $D_A(z)$ and $H(z)$ parameters at $z = 2.4$ by 10%–30%. While the primary goal is to improve on HETDEX, Dark Energy Explorers has already proven to be a uniquely powerful tool for science advancement and increasing accessibility to science worldwide.

Unified Astronomy Thesaurus concepts: Cosmology (343); Cosmological parameters (339); Astronomy education (2165); Dark energy (351)

1. Introduction

Supernovae observations discovered that the universe is undergoing an accelerated expansion (Riess et al. 1998; Perlmutter 1999; Riess et al. 2021), which has been confirmed by a myriad of follow-up cosmological observations (Colless et al. 2003; Tegmark et al. 2004; Dawson et al. 2013; Planck Collaboration et al. 2020; DES Collaboration et al. 2021). The community is struggling for a theoretical understanding of this acceleration (Albrecht et al. 2009), with the cosmological constant as the primary culprit (Weinberg 1989). There are experiments (Laureijs et al. 2011; DESI Collaboration 2016; DES Collaboration et al. 2021) to measure this acceleration both with an improved accuracy and longer-time baseline. Both improvements will help limit the available physical models for explaining the accelerated expansion (i.e., dark energy).

The initial experiments measure the accelerated expansion during the late time of the universe at $z < 1$. While planned experiments like DESI and Euclid provide significant increase in the accuracy of the expansion rate, they will have only a modest increase in redshift range (Laureijs et al. 2011; DESI Collaboration 2016). Currently, the uncertainties on $H(z)$ and $D_A(z)$ at different redshifts range from 1.8% to 3% (see the summary in DESI Collaboration 2016). The Baryon Oscillation Spectroscopic Survey (BOSS) and its extension, eBOSS, measure D_A with accuracy of 1.8% at $z = 0.5$ (Bautista et al. 2020; Gil-Marín et al. 2020), 2.0% at $z = 0.7$ (de Mattia et al. 2021), and $\sim 3\%$ at $z = 2.3$ (du Mas des Bourboux et al. 2020), and the Dark Energy Survey (DES) gives an uncertainty of 2.7% at $z = 0.84$ (DESI Collaboration et al. 2021). Various missions, such as the Dark Energy Spectroscopic Instrument (DESI) and Euclid are expected to achieve a precision of $\sim 0.5\%$ at $z \sim 1$ (Laureijs et al. 2011; DESI Collaboration 2016).

In order to detect any evolution in the nature of dark energy over cosmic time, it is necessary to cover as large a redshift range as possible. The Hobby–Eberly Telescope Dark Energy Experiment (HETDEX) is designed to study the expansion rate at $1.9 < z < 3.5$ with an accuracy comparable to even the best

¹⁰ NSF Graduate Research Fellow.



low- z experiments (Gebhardt et al. 2021). HETDEX will determine redshifts of at least one million Ly α -emitting (LAE) galaxies from $1.9 < z < 3.5$. Our approach is to observe an area of 540 square degrees with an instrument composed of 74 integral field units feeding 156 spectrographs, over a spectral range of 350–550 nm.

HETDEX will use these million LAE galaxies to measure and analyze the full shape of the galaxy power spectrum and galaxy correlation function, including baryon acoustic oscillations (BAOs), of the $1.9 < z < 3.5$ universe to determine the epoch’s dark energy density. Measurements of clustering in the directions parallel and perpendicular to the line of sight provide constraints on the Hubble parameter, $H(z)$, and the angular diameter distance, $D_A(z)$ through the Alcock–Paczynski test (Alcock & Paczyński 1979; Sánchez et al. 2014). The angle-averaged clustering measurements determine the average distance, $D_V(z)$, as

$$D_V(z) \propto \frac{D_A(z)^2}{H(z)}. \quad (1)$$

With the expected volume and density of the measured LAEs, HETDEX will determine the Hubble expansion rate, $H(z)$, and the angular diameter distance, $D_A(z)$, each to 0.8% around $z = 2.4$, which would be the most accurate measure of the Hubble expansion at this epoch (Gebhardt et al. 2021). These accuracies translate to an overall accuracy on the volume-averaged distance, D_V , below 0.7%.

The accuracy on the cosmological constraints coming from HETDEX is primarily determined by the number of LAE sources, the false positives (FPs) due to noise, and the [O II] contamination. We have optimized our detection and classification algorithm (Leung et al. 2017; Farrow et al. 2021b; Gebhardt et al. 2021; Davis et al. 2023; Mentuch Cooper et al. 2023) to reach specifications on each of these parameters. This paper presents how we utilize the human eye’s ability for pattern recognition to further improve on the project’s algorithm and possibly push into new regimes, such as lower signal-to-noise ratio (S/N) and poor emission-line fits.

Citizen science is a collaboration between scientists and the public to reach a larger science goal. Aiming to utilize the human eye’s ability for pattern recognition, we created a citizen science project, Dark Energy Explorers,¹¹ to improve HETDEX data products.

Other large surveys have used contributions from citizen science as part of the analysis pipeline, e.g., Gravity Spy for LIGO to improve noise rejection (Bahaadini et al. 2018), TESS’s Planet Hunters (Eisner et al. 2021), and Catalina Outer Solar System Survey’s (Drake et al. 2014) evaluation of candidates to improve completeness. Galaxy Zoo is a well-established project to classify galaxies from SDSS, and it has set the stage for large-survey citizen science (Land et al. 2008; Lintott et al. 2008). Beginning as a simple website, Galaxy Zoo has now expanded and created a host platform, Zooniverse,¹² which is now the world’s largest citizen science platform. Zooniverse is home to dozens of citizen science projects in various disciplines that have led to hundreds of publications. Dark Energy Explorers is one of these programs. Zooniverse’s mission is “to enable research that would not be possible, or practical, otherwise.” Building off of this mission, Dark Energy

Explorers aims to improve the accuracy of HETDEX while simultaneously allowing participants the first look at astronomical sources by teaching them to classify millions of sources from the Hobby–Eberly Telescope.

HETDEX has no preselection of sources, which is quite different from other large-scale surveys. We tile the sky over 540 square degrees with a fill factor of 1/4.5 using 74 integral field units. These units feed 156 spectrographs covering a wavelength range of 350–550 nm, with a resolving power that ranges from 750 to 950. The software then searches through every spectral and spatial resolution element for emission lines, including those from LAEs. Over its lifetime, HETDEX will acquire about one billion spectra and one trillion resolution elements (spatial and spectral). From these trillion resolution elements we expect to find about 1.3 million LAEs, 0.92 million [O II] emitters, and many stars, meteors, asteroids, etc. The reduction from one trillion resolution elements to the one million LAEs is dependent on the S/N limit adopted by the experiment. At a lower S/N, more emission lines are detected but at the expense of more FPs due to artifacts (Mentuch Cooper et al. 2023). HETDEX is exploring ways to use all trillion resolution elements, which requires accurate control of pixel-level defects.

One fundamental challenge of HETDEX involves sifting through the data and distinguishing the LAEs from the other detections, including FPs, [O II] emitting galaxies, meteor trails, and other line-like features. HETDEX has optimized its algorithms to distinguish the various sources, but there are advantages that the human eye can provide that will improve the survey. While it is impractical for a small research team to visually vet many millions of sources, we realize that, if possible, such vetting would enable significant improvements toward the measurements of $H(z)$ and $D_A(z)$. Our goal is to use citizen science to classify millions of HETDEX sources and help keep the contamination rate low.

Visual vetting has two important aspects. First, we use citizen science to help identify sources caused by non-Gaussian noise, therefore reducing the FP rate and generating a cleaner training set for machine learning. Second, we are able to explore regimes that are more difficult to classify algorithmically, such as that for detections with low S/N and/or higher chi-squared from a single emission-line fit (see Section 3.2). For example, active galactic nuclei (AGNs) have a variety of emission profiles that are easy for the human eye to distinguish but can be difficult for a general algorithm that uses a single unresolved emission line. By including these additional sources, we improve our accuracy on the expansion rates. The current results suggest that we can increase the HETDEX LAE sample by up to 50%, which yields a $\sim 20\%$ improvement on the distance estimates (see Figure 1). Our goal is to combine the visual vetting with machine learning, and the first step toward doing this is to create labels for training the machine-learning model. Creating these labels to train the machine creates a rich opportunity for Dark Energy Explorers and HETDEX.

This paper discusses how we use Dark Energy Explorers to help HETDEX reach specifications. In Section 2, we discuss how to improve the accuracy of the distance measures using the collected data from Dark Energy Explorers. Section 3 focuses on Dark Energy Explorers interface and how we train the public to become HETDEX astronomers. Section 4 presents initial results of Dark Energy Explorers. Section 5 shows how

¹¹ <https://www.zooniverse.org/projects/erinmc/dark-energy-explorers>

¹² <https://www.zooniverse.org/>

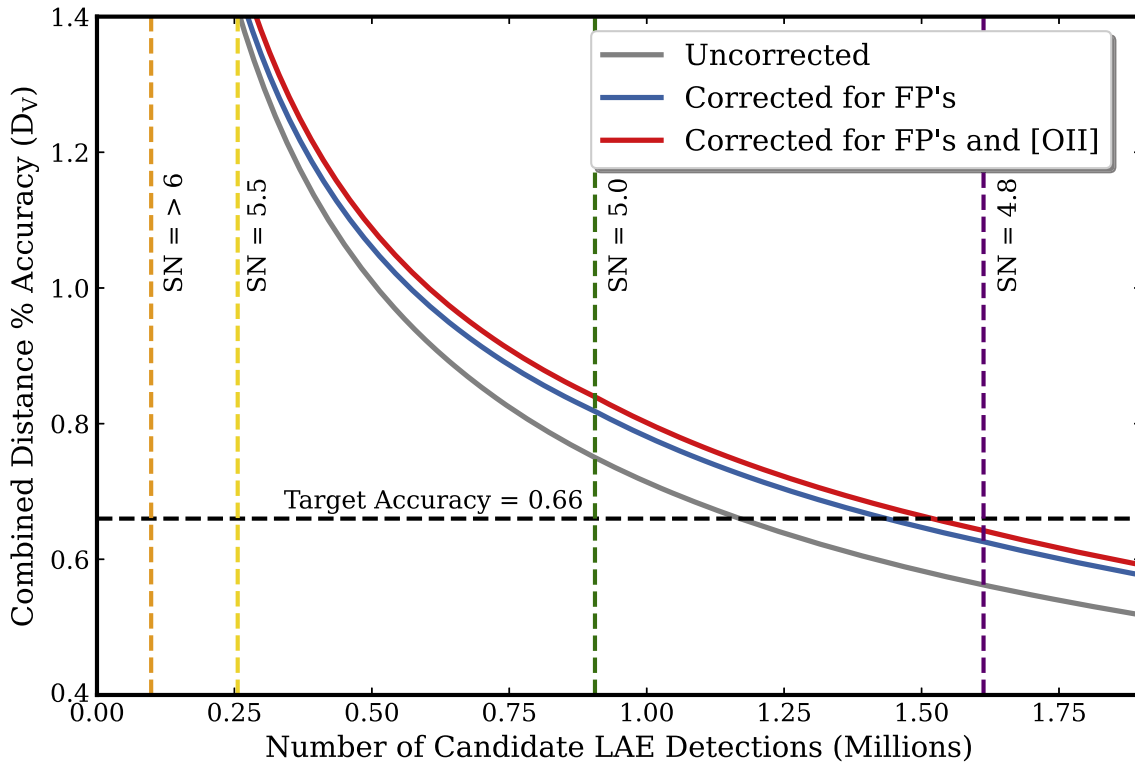


Figure 1. Based on the modeling of HDR2 results, this plot shows the accuracy of D_V as a function of the number of detected LAEs and signal-to-noise ratio (S/N). The primary relation is that D_V goes as the inverse square root of the number of LAEs, and this is given by the gray line assuming no corrections for false-positive (FP) rate or [O II] contamination. This demonstrates how the accuracy of the cosmological parameter, D_V is affected by FPs, as the blue line, and further including the [O II] contamination as the red line. The black dashed horizontal line represents the target goal for the whole HETDEX survey. The precision achieved for a given S/N limit is given by the intersection of the gray, blue, and red curves with the vertical dashed lines that represent the different S/N cuts. Note that the blue and red curves are less accurate due to contamination. Here, we assume a galaxy bias of $b_{\text{LAE}} = 2$. The S/N cut and/or the number of LAEs that is required to obtain the accuracy on D_V motivates the goal for the Dark Energy Explorers citizen science campaign.

we incorporate the Dark Energy Explorers in the HETDEX database using machine learning, and Section 6 will discuss conclusions.

2. Improving the Accuracy of the Distance Measures from HETDEX

HETDEX will obtain about one billion spectra and one trillion resolution elements over the lifetime of the survey (Gebhardt et al. 2021; Mentuch Cooper et al. 2023). Once HETDEX reaches completion, we expect to have over one million redshifts of distant LAEs between $1.88 < z < 3.52$ and over one million redshifts for nearby [O II] galaxies with $z < 0.5$. These galaxies are what we use for the cosmological analysis. The requirements are an FP rate $< 10\%$ due to noise, contamination due to [O II] emitters $< 2\%$, and the total number of LAEs of over one million. As shown in Mentuch Cooper et al. (2023) and Davis et al. (2023), we reach the specifications with little margin for error. Our goal here is to push to lower S/N and higher chi-squared, measured against a Gaussian fit. This will increase the number of LAEs and still maintain the low FP and contamination rates. In order to keep these rates this low, we employ visual vetting and machine learning.

The HETDEX HDR2 contains $\sim 50,000$ objects with well-fit emission-line profiles ($\text{S/N} > 5.5$ and $\chi^2 < 3$) that are classified as LAEs (Mentuch Cooper et al. 2023). This catalog is from contiguous fiber spectra coverage of 25 deg^2 of the sky. There are undoubtedly many sources outside these cuts; for example, AGNs with broad emission lines will deviate from a

single-line fit, causing the chi-square value to be high. We expect a robust visual vetting and machine-learning campaign to extract an additional $\sim 10\%$ of the catalog while keeping the FP and contamination rates low.

Figure 1 quantifies the trade between S/N cut, chi-squared cut, FP rate, and contamination rate for the primary cosmological parameter D_V . For this figure, we need to assume the LAE galaxy bias since the combined distance accuracy improves linearly with bias (higher bias values provide better accuracy). We are in the process of measuring this bias accurately (D. Farrow et al. 2023, in preparation), and for this figure we assume $b_{\text{LAE}} = 2.0$. For the properties we have control over (S/N, FP, chi-squared, [O II] contamination), we base these rates on analyses by Davis et al. (2023) and Mentuch Cooper et al. (2023). As demonstrated in Figure 1 the accuracy of D_V depends on all three factors. The black dashed horizontal line represents the target goal for the entire HETDEX survey. The primary relation is that D_V goes as the inverse square root of the number of LAEs, and this is given by the gray line assuming no corrections for FP rate or [O II] contamination. The vertical lines represent the different S/N cuts as given in the legend. The distance accuracy is degraded by the contribution from the FPs (the blue line), and then further including [O II] contamination (the red line). The uncertainty on D_V scales linearly with the FP rate as the FPs affect both the number of true LAEs and add in white noise. The uncertainty on D_V scales quadratically with the [O II] contamination as this affects the number of LAEs and imposes clustering power from nearby galaxies onto the LAE power spectrum. Farrow et al.

(2021a) show that an upper limit of 2% on [O II] contamination does not impact the scientific requirements significantly. Within Figure 1 we assume to have a [O II] contamination rate of 0.013, based on the work of Davis et al. (2023). The FP rate is a result of measuring a confirmation rate of LAEs, and therefore our FP rate will be less than the confirmation rate. Initial FP rate estimates are given in Gebhardt et al. (2021) and Mentuch Cooper et al. (2023) to be <0.24 at $S/N = 4.8$, <0.19 at $S/N = 5$, <0.05 at $S/N = 5.5$, and <0.01 at $S/N > 6$, providing upper limits.

The intersection of the gray, blue, and red curves with the black horizontal dashed line provides the uncertainty on D_V . The red line has the largest uncertainty on D_V since it accounts for contamination and FPs and is what we use for the final prediction. One can then read off the S/N cut that is required to obtain the desired accuracy on D_V , motivating our goal for the Dark Energy Explorers citizen science campaign.

Given that the primary corrections are due to FPs and [O II] contamination, we focus our citizen science campaign on those issues. As described below, the FP rate is a visual distinction between real and fake sources, and the [O II] contamination is a distinction between distant and nearby galaxies.

3. Developing Dark Energy Explorers

Through the end of 2022, the HETDEX data collection is 50% complete, and analysis from the first internal data release in 2019 showed the primary need is for careful visual vetting of the source catalog (Gebhardt et al. 2021). To solve this problem, we designed, created, and launched the worldwide citizen science project, Dark Energy Explorers, in late 2021 February.

The goal of Dark Energy Explorers is to solve the problem of having an infeasible amount of data to classify with a small in-house team. Initial visual vetting by HETDEX collaboration members provided source classifications but took too much time to make classification of the full data set feasible. Establishing a project on Zooniverse facilitated participation from thousands of volunteers across the world drawn from the platform's millions of registered volunteers where the project is accessible on any smartphone, tablet, or desktop computer with internet access. In addition to aiding in volunteer recruitment, the Zooniverse team has provided support from project creation, to launch, and beyond. In addition to reaching a larger science goal, the project tutorial (discussed more in Section 3.1) is intended to allow anyone to participate, even those without a science background.

We constructed two different workflows for Dark Energy Explorers: “Nearby versus Distant Galaxies” and “Fishing for Signal in a Sea of Noise.” The first workflow aims to differentiate between [O II] emitting objects at $z < 0.45$ (“Nearby Galaxies”) from $1.88 < z < 3.52$ LAEs (“Distant Galaxies”). The “Nearby versus Distant” galaxies workflow addresses [O II] contamination and helps us optimize our discrimination algorithm. Therefore, the target objects all have $S/N > 5$ and include candidate LAEs, AGNs, [O II] galaxies, and stars. The second, more recent workflow, “Fishing for Signal in a Sea of Noise,” addresses the FPs. As the FPs are caused by noise, they are more difficult to identify. This workflow therefore hones in on classifying the $S/N > 6$ LAE candidates. This is the subset that is addressed in detail with machine-learning efforts in this paper, and lower S/N regimes will be explored in future work.

3.1. Tutorial and Field Guide

Dark Energy Explorers is accessed via the Zooniverse app or website. Participants can create an account to save classifications and data or choose to participate anonymously. The Dark Energy Explorers project can be accessed directly via its URL or via selection from the Zooniverse projects page where it is listed under the “Space” or “Physics” categories. Volunteers can then choose a workflow from the project landing page, either “Nearby versus Distant Galaxies” or “Fishing for Signal in a Sea of Noise.” After choosing the workflow, the participants walk through a tutorial on how to classify the HETDEX data (discussed more in Section 3.2 and Figure 2).

The foundation of Dark Energy Explorers is the easy-to-understand tutorial. The tutorial trains members of the public to become amateur HETDEX astronomers without any prior astronomy or general science background by simplifying the classification process into digestible, jargon-free tasks. The participants read a tutorial and are provided a number of criteria to allow them to choose between two binary options in the workflows. We opted to limit the classifications to a binary choice to maximize classifying speed. For example, on a mobile device, which offers the swipe-left/swipe-right classifying option, a Dark Energy Explorer can classify 20 sources in less than a minute. This ensures that users have the ease of the swiping feature on the mobile device with simple choices, avoiding multiple or nested selections.

The primary criteria users have to consider for the “Nearby versus Distant” workflow are

1. the relative size of the object,
2. the strength of the emission line, and
3. the appearance of the emission line in at least one or more of the fiber spectra.

For example, a nearby galaxy often appears as a large, bright, resolved source in the images, while distant galaxies are generally small and faint. See Figure 3 for a comparison.

The primary criteria users have to consider for the “Fishing for Signal in a Sea of Noise” workflow are

1. the quality of the data collected,
2. the strength of the emission line, and
3. the appearance of the emission line in at least one or more of the fiber spectra.

See Figure 4 for a comparison. In addition to the tutorial, participants have a short, easy-access field guide that is available to them on the main classification page. This guide gives a quick reminder of the selection criteria for either workflow. Figure 2 shows the mobile version of Dark Energy Explorers where participants have access to both the field guide and an option to go back to the entire tutorial at any time during classification.

3.2. Interface and Data Inputs

We have combined multiple images and spectral data to create what we call a “mini” image. We refer to them as “minis” because they are a greatly reduced and compact representation of the more complete and complex Emission Line eXplorer (ELiXer; Davis et al. 2021). These images were designed to be compact and compatible with both a desktop and mobile device. Because we desired the workflow to be usable on mobile devices with a variety of phone resolutions

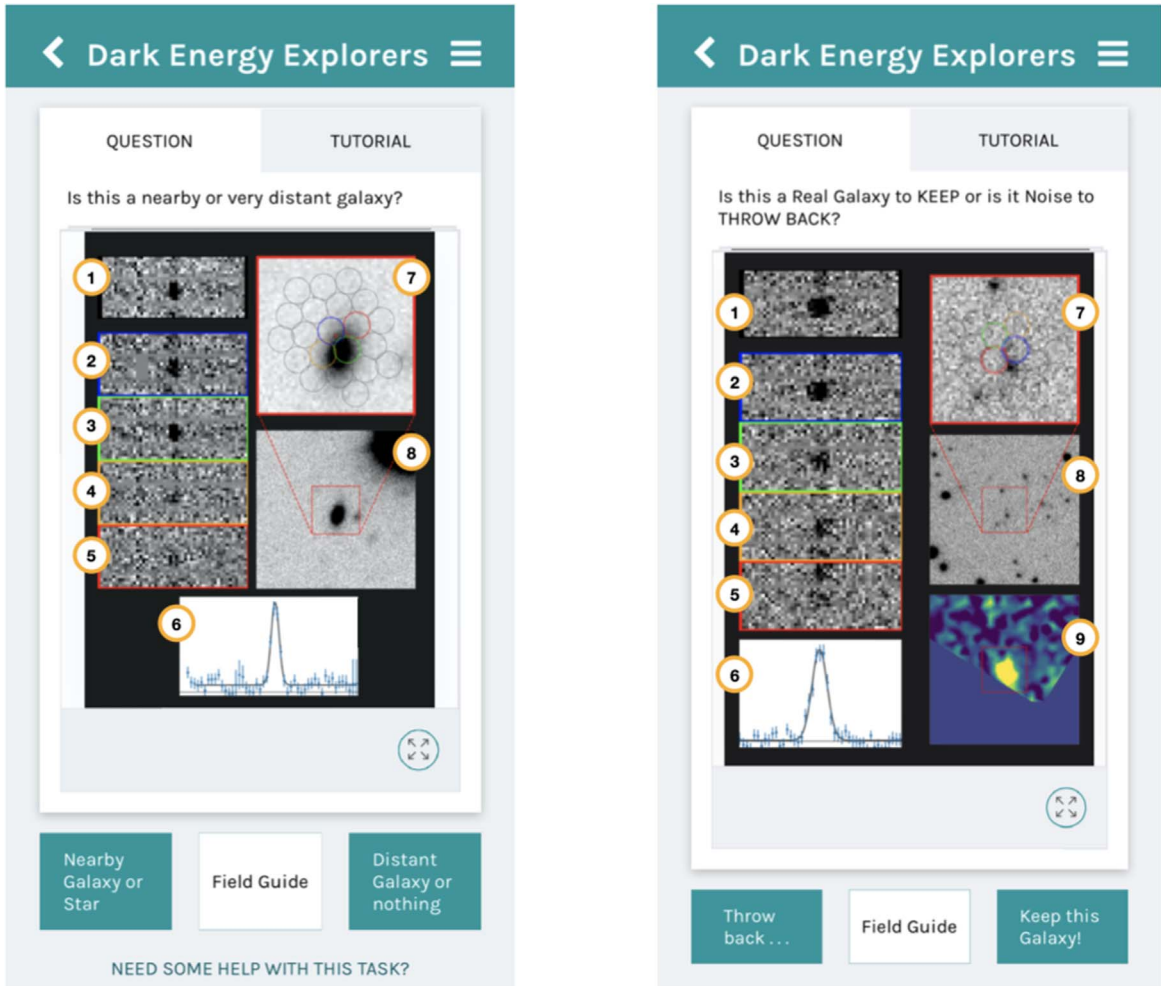


Figure 2. Above, we show the mobile version of Dark Energy Explorers. In either workflow (“Nearby versus Distant Galaxies” on the left; “Fishing for Signal in a Sea of Noise” on the right), the participants are shown various visualizations of an HETDEX detection from the Hobby–Eberly Telescope (HET) and given a binary choice to classify the astronomical source. Access to the tutorial is shown in the top right corner and the field guide in the bottom center. Image summary: (1)–(5) 2D fiber cutouts, (6) 1D emission-line fit, (7) imaging postage stamp with fiber positions (zoomed in, $9'' \times 9''$), (8) imaging postage stamp (zoomed out, $30'' \times 30''$), and (9) lineflux map. Image (9) is only included in the “Fishing for Signal in a Sea of Noise” workflow.

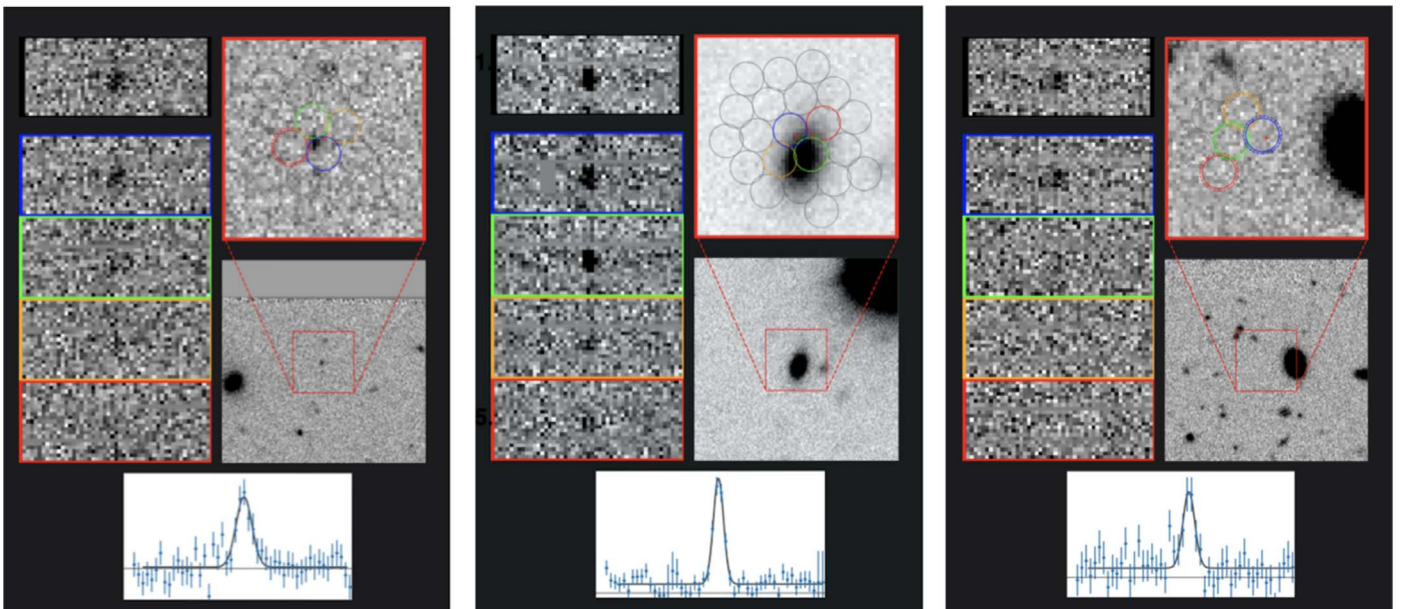


Figure 3. Above, we have examples of the “minis” from Dark Energy Explorers “Nearby vs. Distant” workflow. From left to right: a distant $\text{Ly}\alpha$ emitting galaxy, a nearby [O II] emitter, and a tricky case that might need more information.

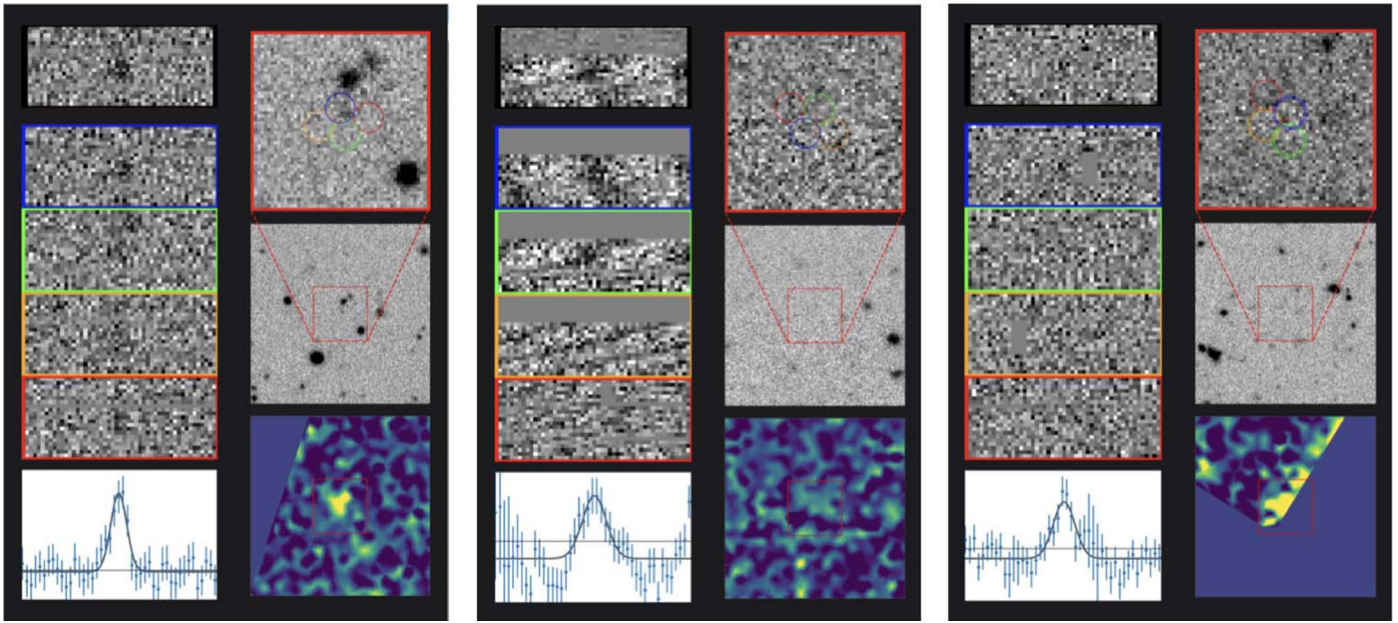


Figure 4. Above, we have examples of the “mini’s” from Dark Energy Explorers “Fishing for Signal in a Sea of Noise” workflow. From left to right: keep (real galaxy/emission line), throwback (bad detection), and a tricky case that might need more information.

and aspect ratios, we decided to keep the visualizations simple and compact. In addition, we wanted the site to be accessible to people without science backgrounds; therefore, the images are free of jargon and numbers, which could easily distract or turn away participants.

The “mini’s” consist of panels (1)–(8) for the “Nearby versus Distant Galaxies” workflow and images (1)–(9) for the “Fishing for Signal in a Sea of Noise” workflow. See Figure 2. These images are defined in detail here:

(1)–(5) *2D Fiber Cutouts*. Five cutouts within ± 40 Å of the detection line center in the spectral direction and ± 1 fiber across the detector. The spectral images are sky subtracted. Image (1) (highlighted in black) is the weighted sum of all contributing fibers. The rows below, images (2)–(5) (blue, green, orange, red), are the four fibers ordered by distance from the source position.

(6)—*1D Line Fit*. The resultant 1D spectrum and the emission-line fit.

(7)—*Postage Stamp with Fiber Positions*. The footprint of all fibers contributing to the detection plotted over deep ground-based imaging of a $9'' \times 9''$ region centered at the spatial position of the maximum S/N of the emission line. For internal classification purposes, a number of imaging catalogs are available as described in detail in Davis et al. (2023). However, for the current workflows, we only display ancillary *r*-band imaging obtained from Hyper Suprime-Cam (HSC) on the Subaru Telescope by the HSC Subaru Strategic Program (HSC-SSP; Aihara et al. 2018) and the HSC HETDEX Survey (HSC-DEX; Davis et al. 2023). This limits confusion created by heterogeneous image quality and flux sensitivity. The four colored fibers match the colors outlined to the left in images (1)–(5). Fibers with a dashed outer ring are at the edge of the detector. The point spread function (PSF)-weighted center of the detection is marked with a red cross.

(8)—*Imaging Postage Stamp*. A postage stamp cutout of the ancillary imaging data of a $30'' \times 30''$ region centered at the spatial position of the emission line. This is a zoomed-out

version of image (7) to show possible catalog counterparts or nearby sources.

(9)—*Lineflux map*. The wavelength-collapsed flux intensity map over a $\pm 3\sigma$ region from the emission-line center. The lower section of this particular map is blank as the region happens to fall off the edge of the detector. For more detail about how these minis are generated, see Gebhardt et al. (2021) and Davis et al. (2023).

These telescope images are of a detection on a random area of sky, centered on an object (i.e., galaxy, star, AGN) chosen at random from the sample database. Dark Energy Explorers avoids using numbers and astronomical jargon within the tutorial and field guide, making an approachable way to train anyone to become a volunteer for HETDEX.

3.3. Education and Engagement

While the main focus of Dark Energy Explorers is for science advancement of HETDEX, other primary goals have been education, outreach and public engagement. To do this, we have implemented a feature to educate participants while they conduct galaxy classifications. Other projects, such as Galaxy Zoo, have used this method to provide continued engagement and longer participation in the projects (Land et al. 2008; Lintott et al. 2008). Learning from the Galaxy Zoo project, Dark Energy Explorers aims to provide ways for participants to interact with each other and the HETDEX team to learn more of the science (Raddick et al. 2019). Following these methods, we have incorporated a similar objective for our consistent users of Dark Energy Explorers. For every 50 sources an individual classifies, they receive a pop-up message linking them to learn more about the science behind HETDEX, dark energy, and general astronomy. This “level-up” incentive provides continued engagement with the project while educating the public about astrophysics topics and empowering them as scientists.

In an effort to engage with our participants and address questions they may have, Dark Energy Explorers has a “Talk”

Table 1
Dark Energy Explorers Workflows

Workflow Name	Launch Date	Parent Sample	S/N cut	Down Selection	Sources Complete	Total Classifications
Nearby versus Distant	Feb 2021	500,000	5.5	209,000	114,000 LAE 95,000 [OII]	3.1 M
Fishing for Signal in a Sea of Noise	2021 October	700,000	6	60,000	14,000 ^a	140,000

Note.

^a Ongoing data collection.

board that allows participants to post comments and/or questions. These “Talk” boards are monitored by the HETDEX team and allows the participants to engage with the scientists or ask frequent questions. So far, we have had over 500 participants engage with the discussion boards and post over 5000 comments and images. The discussion boards provide a space for participants to see many examples of classified sources for each workflow. The HETDEX team member can comment back to confirm or deny the classifications to address these questions, and other participants can then see the discussion.

4. Dark Energy Explorers Results

4.1. Classification and User Statistics

Dark Energy Explorers launched in February 2021 and since then has collected roughly three and a half million classifications by 11,000 volunteers in over 85 different countries around the world. See Table 1 for workflow classifications and selection from the HETDEX data releases. This is a result of over three million classifications within the “Nearby versus Distant” workflow. These classifications amount to $\sim 209,000$ sources classified as LAEs or [O II] galaxies. The “Nearby versus Distant” workflow was the originally launched workflow and thus has more classifications in total. The “Fishing for Signal in a Sea of Noise” was launched in October of 2021 and has a total of 140,000 classifications resulting in $\sim 14,000$ completed subjects, which are the focus of the machine-learning efforts (see Section 5).

Since launch, roughly 60% of our users have classified more than 50 objects and have received the pop-up (see Section 3.3). Of those users, two-thirds have had continued engagement with the project by completing over 100 classifications. In other words, over 40% of total participants have made at least 100 classifications.

With visual vetting, we acknowledge human bias and also work to build in reliability in human classification. In order to have a robust entry for each source, we must average over some number of participants. The number of classifications selected per subject is a balance between classification efficiency and ensuring incorrect classifications are mitigated by statistics. We found that 10 classifications yield values that are just as consistent as the larger sample sizes and allow many more objects to be classified. The benefit of the binary classification is that it is very straightforward to obtain a classification statistic from a workflow by simply taking a median or average of the individual classifications. For this work, the 10 binary human classifications are averaged to build confidence in an accurate cataloging (Santos-Fernandez & Mengersen 2021). The final product yields a Dark Energy Explorers probability

(DEE probability) that the source is a distant galaxy or real signal by visual vetting, according to each workflow, respectively.

We note that the standard deviation of the N classifications is also of potential use as it indicates objects where there is strong disagreement on the subject among the citizen scientists. Internal HETDEX team members follow up these special cases by eye. The Dark Energy Explorers measurements thus reduce the number of sources the small internal HETDEX team must visually investigate and offers an effective method to identify rare, unique sources within the HETDEX survey. Out of 140,000 total classifications in “Fishing for Signal in a Sea of Noise,” 0% have a standard deviation greater than 0.5.

4.2. Nearby versus Distant Results

The citizen scientist data from the “Nearby versus Distant” workflow is working to validate current research by comparing the visual DEE probability to the model probability (Davis et al. 2023). We had over three million classifications by the Dark Energy Explorers that resulted in 209,000 sources classified since each is viewed by a minimum of 10 different participants. Once we calculate the DEE Probability, we can compare the [O II] and LAE classifications. Where they disagree allows us to delve into those sources and determine from where the disagreement stems. In some cases, the classification from the Dark Energy Explorers participants are correct, and we use these instances to improve our model (Davis et al. 2023). The Dark Energy Explorers classifications match the model, EliXer, with more than 92% agreement. The success of the Dark Energy Explorers significantly reduced the time spent by the HETDEX team manually inspecting sources (Davis et al. 2023). This is due to the fact that we used the “Nearby versus Distant” workflow to provide a way to incorporate visual vetting of the problem areas to tune our algorithm. With theses effort from Dark Energy Explorers, we now can rely on Davis et al. (2023) for the [O II] contamination rate (See Figure 1). Because of the work of the citizen scientists to confirm the [O II] emitting sources, we have retired this workflow and focus on the machine-learning efforts with “Fishing for Signal in a Sea of Noise.”

4.3. Fishing for Signal in a Sea of Noise Results

The raw data products of HETDEX contain many pixel-level events, such as charge traps, hot pixels, and time-variable changes in calibration. Since these imperfections are often hard for algorithms to identify (Gebhardt et al. 2021), most have to be found manually (Mentuch Cooper et al. 2023). The human eye is good at identifying these features, particularly at high S/N since they tend to have correlated residuals or create obvious

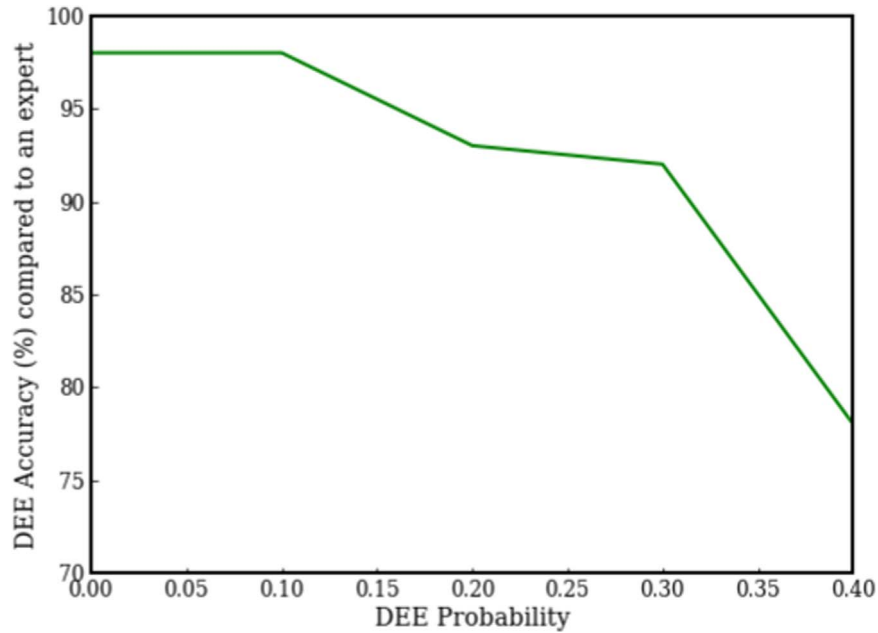


Figure 5. Above, we show the percent accuracy of the Dark Energy Explorers participants as compared to an HETDEX expert for the “Fishing for Signal in a Sea of Noise” workflow. This accuracy demonstrates how well the Dark Energy Explorers participants perform compared to an HETDEX team member. The DEE probability is the average of the 10 classifications for each source and represents the probability that a source is an FP–0–or a real detection–1. This figure and the analysis in this paper focus on identifying the FPs or DEE probabilities <0.4 .

features in the 2D charge-coupled detector imaging. Thus the main goal for Dark Energy Explorers is to generate an additional removal of false sources from the catalog. For this paper we focus on the FPs so that a DEE probability of 1.0 means an object is real, and a probability of 0.0 identifies a false detection. Thus, DEE probability is only in reference to the “Fishing for Signal in a sea of Noise” workflow.

Since the end of 2022, we have had over 140,000 classifications for the “Fishing for Signal in a Sea of Noise” workflow, resulting in 14,000 sources out of a total of 60,000 inputs. Figure 5 shows how well the participants do (i.e., the accuracy of the Dark Energy Explorers) compared to an HETDEX team member against the DEE probability. For the figure, the data have been binned to intervals of 0.1 in probability with 200 sources per bin. This analysis determines the cut in DEE probability where we accurately remove false detections without removing too many real sources. At this point, the main interest for the Dark Energy Explorers is to remove the obvious FPs. Again, when calculating the DEE probability, a false source and a real source correspond to 0 and 1, respectively. Thus, we focus on the accuracy of the Dark Energy Explorers classifications when the source is likely to be an artifact (i.e., having a probability less than 0.4). The plot indicates that Dark Energy Explorers most often agree with the HETDEX astronomers when the probability of the object being a real source is low. Below we describe the strategy for aggregating the results from the Dark Energy Explorers probabilities and combine their classifications of artifacts with machine-learning efforts.

5. Incorporating the Dark Energy Explorers Results with Machine Learning

The primary goal is to use the Dark Energy Explorers to improve the HETDEX catalog. Since HETDEX is a multiyear project that uses a large number of detectors, there are a

significant number of instrumental and calibration issues that can produce features that mimic real sources. A secondary goal is to use the classifications as a training set for unsupervised learning efforts. We analyze the citizen science outputs with both of these in mind.

To further ensure removal of the FPs from the data, we explore here including machine learning into the overall classifications. There are a few aspects worth considering. First, we can use the DEE classification as labels for machine learning. Second, we can use machine learning to help inform the Dark Energy Explorers training. Third, we can combine the Dark Energy Explorers classifications with machine learning to provide stronger leverage on the final classifications. For example, unsupervised learning combined with the Dark Energy Explorers probabilities would allow us to apply cuts to the full sample of sources as an additional technique to remove false detections from the catalog.

We have about 140,000 classifications by participants since each source is inspected by a minimum of 10 individuals. Of those 14,000 categorized detections, ~ 2000 have a DEE probability below 0.1 which are considered the mostly likely to be FPs. Furthermore, ~ 4000 sources have a DEE probability less than or equal to 0.3. From Figure 5 the agreement of identifying FPs (DEE probability ≤ 0.3) by the Dark Energy Explorers stays above 92%. When the DEE probability falls between 0.3 and 0.7, the accuracy also falls. These are the tricky cases as demonstrated in Figure 4 and we must turn to further information (such as EliXer) and the HETDEX team to gather more information to classify the source.

For the purpose of reducing the FPs within HETDEX we will focus our analysis on the FPs. We use unsupervised learning to map DEE probability values to spectral characteristics. We do this via t-distributed stochastic neighbor embedding, t-SNE (van der Maaten & Hinton 2008), from the `scikit-learn` Python package (Pedregosa et al. 2012).

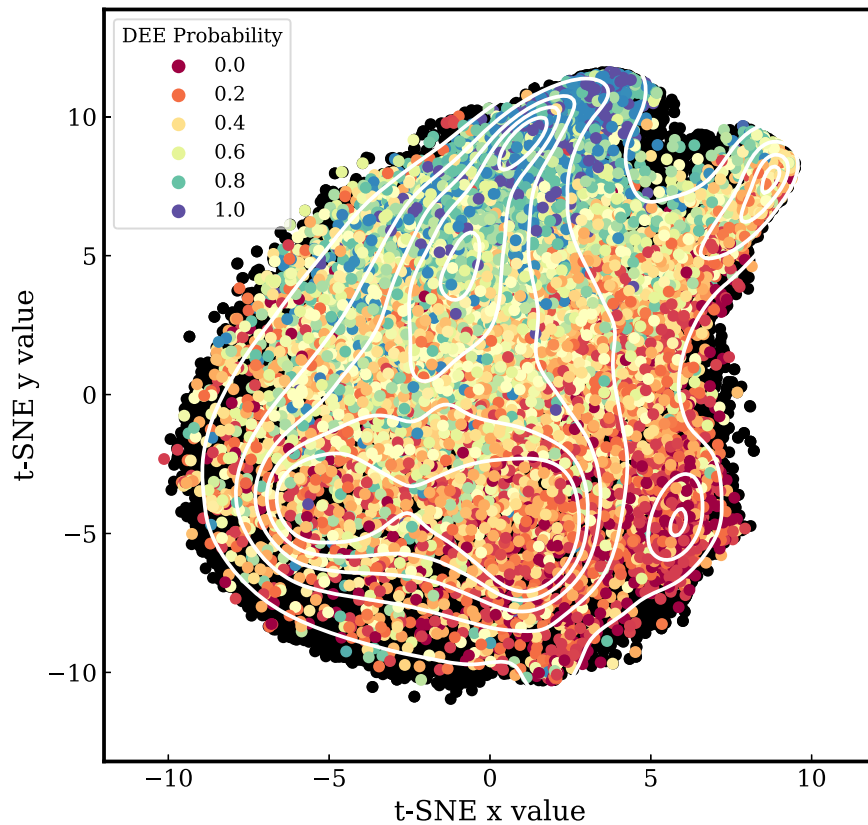


Figure 6. Above, we show a visual representation of the t-SNE machine-learning algorithm. The colored points are labels created from the Dark Energy Explorers probabilities ($\sim 14,000$), which are overplotted on the black points, which represent the total number of sources with $S/N > 5.5$ within the source catalog ($\sim 120,000$ sources). The contours of source density within the t-SNE space are shown in white.

We analyzed t-SNE inputs with both the full spectra and with a $\pm 50 \text{ \AA}$ region centered around the detected emission feature. t-SNE then takes the high-dimensional data set and produces a representation of those points in a lower-dimensional space. Here we will reduce the data to two dimensions, and therefore the n -components are set to 2. The next parameter we set is the perplexity. Following the most recent research, the perplexity is shown to have optimal results when set as the square root of the number of data points or fixed at 50 for large data sets (van der Maaten 2015).

Data sets of high dimensionality, like our spectral data (1036 dimensions for the full spectra) often present challenges when analyzing (van der Maaten & Hinton 2008; V. Poleo 2023, in preparation). For this analysis we use only a small wavelength region ($\pm 50 \text{ \AA}$) around the emission line. We do this cut in order to focus on the line itself and keep it centered for the t-SNE visualization. We have tried using the full spectrum, and in this case t-SNE tends to separate based on redshift. Since we are primarily focused on real emission or FPs, we currently remove the redshift information. Future work will continue to explore using the full spectrum. For example, applying a principal component analysis or an autoencoder might help in removing the redshift from the visualization and help to further discriminate sources. Therefore, reducing the dimensions of the data set, from 1036 for the full spectra down to 100 for the cutout around the emission line allows us to avoid the complications and simplifies the data visualization. Making this dimension reduction gave results with better discrimination, and those results are discussed here.

In Figure 6 we show the combination of the Dark Energy Explorers participants with the t-SNE machine-learning technique, which results in the highest removal of fake sources and the lowest percentage of removal of real detections. Figure 6 shows a visual representation of the t-SNE machine-learning algorithm. The colored points are labels created from the Dark Energy Explorers probabilities ($\sim 14,000$), and the black points are the total number of sources with $S/N > 5.5$ within the source catalog ($\sim 120,000$ sources). The contour of source density within the t-SNE space is overplotted in white. These results demonstrate how the combination of citizen science visual vetting and t-SNE can yield the more consistent results.

Going through all the sources that have a DEE probability of below 0.1, we find a result of 98% agreement, as seen in Figure 5. All of these sources are reviewed by members of the HETDEX team and then compared to the DEE probability to determine this accuracy.

The variety of problems we see in the significant number of detectors and resolution elements makes it hard for us to quantify all categories of FPs, which is why we first employ members of the HETDEX team. The t-SNE visualization will eventually be combined with a supervised learning algorithm where we can group regions, providing a well-trained classification scheme. Our goal for this paper is to determine the FPs and remove them from the sample with the lowest amount of real detections being removed. Our next goal is to use these vetted FPs to enhance the classifications and the selection configuration.

We find that the highest accuracy of both Dark Energy Explorers and t-SNE is within the region of t-SNE x values >3 and t-SNE y values <-1 using the visualization from Figure 6. The selected t-SNE region also yields the lowest removal of real detections with none of the sources in this region having a DEE probability of >0.9 . In other words, using this combination of methods no real sources are removed in the process. In addition, only 0.08% of sources in this region have a DEE probability of >0.6 , which are real detections according to the Dark Energy Explorers. Since we have visual confirmation that professional astronomers believe the sources to be real, we will use the DEE probability to assure that these sources stay in the catalog, meaning that no real detections will be removed. This technique demonstrates how using the combination of machine learning and citizen science allows for isolated removal of false detections while the real detections remain in the catalog.

These results from Dark Energy Explorers allow us to use t-SNE and expand on the full catalog of $\sim 120,000$ sources, which are the black points in Figure 6. By applying the selected region of t-SNE space, we can remove sources while assuring real detections will remain in the sample because we have visual confirmation from the Dark Energy Explorers. This work results in the removal of 7871 false detections from the catalog.

We can expand on this cleaning technique to see where other sources in the current catalog lie. This provides an efficient method to remove sources from the catalog without having to be completely vetted by humans. In conclusion, citizen science and machine learning when used synergistically can enhance science goals while creating a unique educational opportunity for the public.

6. Conclusions

Since the end of 2022, Dark Energy Explorers has collected roughly 4 million classifications by 11,000 volunteers in over 85 different countries around the world. An additional goal of this campaign is to have visual vetting by the Dark Energy Explorers of 100% of all sources down to $S/N > 4.8$ or the full HETDEX catalog, which we expect to contain about three million sources. Repeating this process with all sources at a lower S/N will provide a robust classification using a combination of algorithms, data visualization, and human vetting. While this number is significantly larger than what we have covered already, we will explore ways to engage a larger audience, to use a smaller number of individual classifications, and to incorporate targeted workflows.

So far, Dark Energy Explorers has reduced our team's visual vetting work, making this detailed level of research viable; it would have otherwise been impossible with our small team. This project is an innovative approach to studying dark energy and to classifying large data sets that require visual vetting.

By using Dark Energy Explorers we expect to improve constraints on the cosmological parameters by ~ 10 to 30%. For example, allowing us to go from $S/N > 5.2$ to $S/N > 4.8$ would result in a 30% better accuracy on the cosmological parameter D_V , following the motivation from Figure 1. In addition to Dark Energy Explorers being a uniquely powerful tool for science advancement, the project is increasing accessibility to science worldwide.

Acknowledgments

The results of the Dark Energy Explorers would not be as robust and useful if not for the care and dedication of the

volunteers. We are extremely grateful to their work. It is having a large impact and is motivational.

L.H. acknowledges support from NSF GRFP DGE 2137420 and NASA 21-CSSFP21-0009. K.G. acknowledges support from the NSF-2008793 and from NASA 21-CSSFP21-0009.





Dark Energy Explorers is recognized as an official NASA Citizen Science partner. This publication uses data generated via the www.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.

HETDEX is led by the University of Texas at Austin McDonald Observatory and Department of Astronomy with participation from the Ludwig-Maximilians-Universität München, Max-Planck-Institut für Extraterrestrische Physik (MPE), Leibniz-Institut für Astrophysik Potsdam (AIP), Texas A&M University, The Pennsylvania State University, Institut für Astrophysik Göttingen, The University of Oxford, Max-Planck-Institut für Astrophysik (MPA), The University of Tokyo, and Missouri University of Science and Technology. In addition to Institutional support, HETDEX is funded by the National Science Foundation (grant AST-0926815), the State of Texas, the US Air Force (AFRL FA9451-04-2-0355), and generous support from private individuals and foundations.

The Hobby–Eberly Telescope (HET) is a joint project of the University of Texas at Austin, the Pennsylvania State University, Ludwig-Maximilians-Universität München, and Georg-August-Universität Göttingen. The HET is named in honor of its principal benefactors, William P. Hobby and Robert E. Eberly. The Institute for Gravitation and the Cosmos is supported by the Eberly College of Science and the Office of the Senior Vice President for Research at the Pennsylvania State University.

The authors acknowledge the Texas Advanced Computing Center (TACC)¹³ at The University of Texas at Austin for providing high performance computing, visualization, and storage resources that have contributed to the research results reported within this paper.

ORCID iDs

Lindsay R. House  <https://orcid.org/0000-0002-1496-6514>
 Karl Gebhardt  <https://orcid.org/0000-0002-8433-8185>
 Keely Finkelstein  <https://orcid.org/0000-0003-0792-5877>
 Erin Mentuch Cooper  <https://orcid.org/0000-0002-2307-0146>
 Dustin Davis  <https://orcid.org/0000-0002-8925-9769>
 Robin Ciardullo  <https://orcid.org/0000-0002-1328-0211>
 Daniel J Farrow  <https://orcid.org/0000-0003-2575-0652>
 Steven L. Finkelstein  <https://orcid.org/0000-0001-8519-1130>
 Caryl Gronwall  <https://orcid.org/0000-0001-6842-2371>
 Donghui Jeong  <https://orcid.org/0000-0002-8434-979X>
 L. Clifton Johnson  <https://orcid.org/0000-0001-6421-0953>
 Chenxu Liu  <https://orcid.org/0000-0001-5561-2010>
 Benjamin P. Thomas  <https://orcid.org/0000-0002-0977-1974>
 Gregory Zeimann  <https://orcid.org/0000-0003-2307-0629>

¹³ <http://www.tacc.utexas.edu>

References

- Aihara, H., Arimoto, N., Armstrong, R., et al. 2018, *PASJ*, **70**, S4
- Albrecht, A., Amendola, L., Bernstein, G., et al. 2009, arXiv:0901.0721
- Alcock, C., & Paczyński, B. 1979, *Natur*, **281**, 358
- Bahaadini, S., Noroozi, V., Rohani, N., et al. 2018, *Inf. Sci.*, **444**, 172
- Bautista, J. E., Paviot, R., Vargas Magaña, M., et al. 2020, *MNRAS*, **500**, 736
- Colless, M., Peterson, B. A., Jackson, C., et al. 2003, arXiv:astro-ph/0306581
- Davis, D., Gebhardt, K., Mentuch Cooper, E., et al. 2021, *ApJ*, **920**, 122
- Davis, D., Gebhardt, K., Mentuch Cooper, E., et al. 2023, *ApJ*, **946**, 86
- Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, *AJ*, **145**, 10
- de Mattia, A., Ruhlmann-Kleider, V., Raichoor, A., et al. 2021, *MNRAS*, **501**, 5616
- DESICollaboration 2016, arXiv:1611.00036
- DESCollaboration, Abbott, T. M. C., & Aguena, M. 2021, *ApJS*, **255**, 20
- Drake, A. J., Graham, M. J., Djorgovski, S. G., et al. 2014, *ApJS*, **213**, 9
- du Mas des Bourboux, H., Rich, J., Font-Ribera, A., et al. 2020, *ApJ*, **901**, 153
- Eisner, N. L., Barragán, O., Lintott, C., et al. 2021, *MNRAS*, **501**, 4669
- Farrow, D., Sánchez, A., Ciardullo, R., et al. 2021a, *MNRAS*, **507**, 3187
- Farrow, D. J., Sánchez, A. G., Ciardullo, R., et al. 2021b, *MNRAS*, **507**, 3187
- Gebhardt, K., Mentuch Cooper, E., Ciardullo, R., et al. 2021, *ApJ*, **923**, 217
- Gil-Marín, H., Bautista, J. E., Paviot, R., et al. 2020, *MNRAS*, **498**, 2492
- Land, K., Slosar, A., Lintott, C., et al. 2008, *MNRAS*, **388**, 1686
- Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv:1110.3193
- Leung, A. S., Acquaviva, V., Gawiser, E., et al. 2017, *ApJ*, **843**, 130
- Lintott, C. J., Schawinski, K., Slosar, A., et al. 2008, *MNRAS*, **389**, 1179
- Mentuch Cooper, E., Gebhardt, K., Davis, D., et al. 2023, *ApJ*, **943**, 177
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *JMLR*, **12**, 2825
- Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, *ApJ*, **517**, 565
- Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, *A&A*, **641**, A6
- Raddick, M. J., Prather, E. E., & Wallace, C. S. 2019, *Public Understand. Sci.*, **28**, 636
- Riess, A. G., Casertan, S., Yuan, W., et al. 2021, *ApJL*, **908**, L6
- Riess, A. G., Filippenko, V., Challis, A., & Clocchiatti, A. P. 1998, *AJ*, **116**, 1009
- Sánchez, A. G., Montesano, F., Kazin, E. A., et al. 2014, *MNRAS*, **440**, 2692
- Santos-Fernandez, E., & Mengersen, K. 2021, *Methods Ecol. Evol.*, **12**, 1533
- Tegmark, M., Strauss, M. A., Blanton, M. R., et al. 2004, *PhRvD*, **69**, 103501
- van der Maaten, L. 2015, *JMLR*, **15**, 3221, <https://jmlr.org/papers/volume15/vandermaaten14a/vandermaaten14a.pdf>
- van der Maaten, L., & Hinton, G. 2008, *JMLR*, **9**, 2579, <https://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf>
- Weinberg, S. 1989, *RvMP*, **61**, 1