

Building Machine Learning Challenges for Anomaly Detection in Science

Editors: Elizabeth G. Campolongo³, Yuan-Tang Chou², Ekaterina Govorkova¹, Wahid Bhimji¹⁰, Wei-Lun Chao³, Chris Harris¹⁰, Shih-Chieh Hsu², Hilmar Lapp⁴, Mark S. Neubauer⁶, Josephine Namayanja⁹, Aneesh Subramanian⁵, Philip Harris¹,

Students: Advaith Anand², David E. Carlyn³, Subhankar Ghosh⁷, Christopher Lawrence⁸, Eric Moreno¹, Ryan Raikman¹, Jiaman Wu³, Ziheng Zhang³,

Endorsers: Bayu Adhi⁹, Shazeena Ashraf⁵¹, Marta Babicz²⁸, Furqan Baig⁶, Namrata Banerji³, William Bardon⁹, Tyler Barna¹⁴, Tanya Berger-Wolf³, Micah Brachman⁴⁰, Quentin Buat², Franco Cerino³⁸, Yi-Chun Chang²³, Shivaji Chaulagain⁴⁶, An-Kai Chen²³, Eric Chen³⁷, Deming Chen⁶, Artur Cordeiro Oudot Choi², Chia-Jui Chou³², Zih-Chen Ciou²³, Miles Cochran-Branson², Michael Coughlin¹⁴, Matteo Cremonesi¹⁹, Maria Dadarlat¹⁷, Peter Darch⁶, Malina Desai¹, Daniel Diaz³⁷, Adji Bousso Dieng⁸, Steven Dillmann^{15,16}, Javier Duarte³⁷, Isla Duporge⁸, Urbas Ekka⁷, Hao Fang², Rian Flynn¹⁷, Geoffrey Fox⁴³, Emily Freed⁴¹, Hang Gao¹⁴, Jing Gao³⁶, Mohammad Ahmadi Gharehtoragh¹⁷, Julia Gonski¹⁶, Matthew Graham¹³, Abolfazl Hashemi¹⁷, Scott Hauck², James Hazelden², Saba Entezari Heravi², Duc Hoang⁶, Mirco Huenefeld²⁵, Wei Hu⁶, David C.Y. Hui²², David Hyde³⁹, Vandana Janeja⁹, Nattapon Jaroenchai⁶, Haoyi Jia¹⁶, Yunfan Kang⁶, Elham E. Khoda³⁷, Maksim Kholiavchenko³¹, Sangin Kim²², Aditya Kumar⁶, Bo-Cheng Lai³², Chi-Wei Lee²³, JangHyeon Lee⁹, Shaocheng Lee²¹, Suzan van der Lee¹¹, Charles Lewis², Trung Le², Henry Liao²¹, Mia Liu¹⁷, Xiulong Liu², Xiaolin Liu⁶, Haitong Li¹⁷, Haoyang Li², Vladimir Loncar⁴⁴, Fangzheng Lyu¹⁸, Phuong M. Cao⁶, Ilya Makarov⁴⁵, Abhishek Mallampalli²⁵, Chen-Yu, Mao³², Ann Mariam Thomas¹¹, Alexander Michels⁶, Alexander Migala³⁷, Farouk Mokhtar³⁷, Saúl Alonso Monsalve²⁷, Mathieu Morlighem¹², Min Namgung¹³, Andrzej Novak¹, Andrew Novick⁴¹, Amy Orsborn², Anand Padmanabhan⁶, Sneh Pandya⁴⁷, Jia-Cheng Pan²³, Ana Peixoto², Zhiyuan Pei³⁵, George Percivall⁴⁸, Joshua Henry Peterson²⁵, Alex Po Leung³⁴, Sanjay Purushotham⁹, Zhiqiang Que²⁹, Melissa Quinnan³⁷, Eric R. Sokol⁴², Arghya Ranjan¹⁷, Dylan Rankin²⁴, Christina Reissel¹, Benedikt Riedel²⁵, Dan Rubenstein⁸, Argyro Sasli¹⁴, Eli Shlizerman², Kim Singh², Arushi Singh³⁷, Arturo Sorensen³⁷, Yu Su³, Mitra Taheri⁴⁹, Vaibhav Thakkar²⁴, Eric Toberer⁴¹, Chenghan Tsai²³, Rebecca Vandewalle⁶, Ricco C. Venterea²⁶, Arjun Verma²⁴, He Wang^{20,50}, Sam Wang³⁷, Shaowen Wang⁶, Jianwu Wang⁹, Gordon Watts², Jason Weitz³⁷, Andrew Wildridge¹⁷, Rebecca Williams⁹, Scott Wolf⁸, Yue Xu², Jianqi Yan³⁴, Jai Yu³⁰, Yulei Zhang², Haoran Zhao², Ying Zhao³³, and Yibo Zhong¹⁷

¹MIT, Cambridge, MA 02139, USA

²University of Washington, Seattle, WA 98109, USA

³The Ohio State University, Columbus, OH 43210, USA

⁴Duke University, Durham, NC 27708, USA

⁵University of Colorado, Boulder, Colorado 80309, USA

⁶University of Illinois Urbana-Champaign, Urbana, IL 61801, USA

⁷University of Minnesota, Minneapolis, MN 55455, USA

⁸Princeton University, Princeton, NJ 08544, USA

⁹University of Maryland Baltimore County, Baltimore, MD 21250 USA

¹⁰Lawrence Berkeley National Laboratory, Berkeley, CA 94720 USA

¹¹Northwestern University, Evanston, IL 60208, USA

- ¹²Dartmouth College, Hanover, NH 03755, USA
¹³Caltech, Pasadena, CA 91125 USA
¹⁴University of Minnesota, Minneapolis MN 55455, USA
¹⁵Stanford University, Stanford, CA 94305, USA
¹⁶SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA
¹⁷Purdue University, West Lafayette, IN 47907, USA
¹⁸Virginia Tech, Blacksburg, VA 24061, USA
¹⁹Carnegie Mellon University, Pittsburgh, PA 15213, USA
²⁰University of Chinese Academy of Sciences (UCAS), Beijing, 100049, China
²¹National Taiwan University, Taipei, 10617, Taiwan
²²Chungnam National University, Daejeon, 34134, South Korea
²³National Tsinghua University, Hsinchu, 30013, Taiwan
²⁴University of Pennsylvania, Philadelphia, PA 19104, USA
²⁵University of Wisconsin-Madison, Madison, WI 53707, USA
²⁶University of Perugia, Perugia, 06123, Italy
²⁷ETH Zürich, Zürich, 8092, Switzerland
²⁸University of Zürich, Zürich, 8057, Switzerland
²⁹Imperial College London, London, SW7 2AZ, UK
³⁰University of Chicago, Chicago, IL 60637, USA
³¹Rensselaer Polytechnic Institute, Troy, NY 12180, USA
³²National Yang Ming Chiao Tung University, Taipei City 112304, Taiwan
³³Ricoh Software Research Center (Beijing) Co., Beijing, China
³⁴University of Hong Kong, Hong Kong, China
³⁵Macau University of Science and Technology, Macau, China
³⁶University of Delaware, Newark, DE 19716, USA
³⁷University of California San Diego, La Jolla, CA 92093, USA
³⁸Universidad Nacional de Córdoba, Córdoba X5000GYA, Argentina
³⁹Vanderbilt University, Nashville, TN 37235, USA
⁴⁰Open Geospatial Consortium, Arlington, VA 22201, USA
⁴¹Colorado School of Mines, Golden, CO 80401, USA
⁴²NEON, Battelle, Boulder, CO 80301, USA
⁴³MLCommons and University of Virginia, Charlottesville, VA 22903, USA
⁴⁴CERN, Geneva, 1211, Switzerland
⁴⁵Ilya Makarov, AIRI & ISP RAS, Moscow 109004, Russia
⁴⁶St. Xavier's College, Kathmandu 44600, Nepal
⁴⁷Northeastern University, Boston 02115, USA
⁴⁸GeoRoundtable IEEE SA, Annapolis, USA
⁴⁹Johns Hopkins University, Baltimore 21218, USA
⁵⁰University of Chinese Academy of Sciences (UCAS), Beijing 100049, China
⁵¹University of Arkansas for Medical Sciences (UAMS), Little Rock, AR 72205, USA

ABSTRACT

Scientific discoveries are often made by finding a pattern or object that was not predicted by the known rules of science. Oftentimes, these anomalous events or objects that do not conform to the norms are an indication that the rules of science governing the data are incomplete, and something new needs to be present to explain these unexpected outliers. Finding anomalies can be confounding since it requires codifying a complete knowledge of the known scientific behaviors and then projecting these known behaviors on the data to look for deviations. When utilizing machine learning, this presents a particular challenge since we require that the model not only understands scientific data perfectly but also recognizes when the data is inconsistent and outside the scope of its trained behavior. In this paper, we present three datasets aimed at developing machine learning-based anomaly detection for disparate scientific domains covering astrophysics, genomics, and polar science. We provide a scheme to make machine learning challenges around the three datasets **Findable**, **Accessible**, **Interoperable**, and **Reusable** (FAIR). Furthermore, we present an approach that generalizes to future machine learning challenges, enabling the possibility of large, more compute-intensive challenges that can ultimately lead to scientific discovery.

1 Introduction

Anomaly detection, the process of finding an outlier in a distribution, has long been a critical element of scientific discovery. New anomalies often trigger the question of how to construct a broadly inclusive model allowing for a deeper understanding of the data and, ultimately, a better scientific model. In some cases, this has happened with the observation of a new event or process. Other times, discoveries have been made through aggregating data to observe unexplainable trends¹⁻⁴.

In this paper, we present a series of machine learning (ML) challenges targeting anomaly detection in different scientific domains. We define this as the ability to identify events, objects, or phenomena that are rare and do not match ordinary detector noise or otherwise known signals or phenomena. Anomaly detection is thus searching for the “unknown,” including “known-unknowns” (hypothesized and anticipated but otherwise poorly modeled or quantified) and “unknown-unknowns” (not accommodated by prior knowledge in any given domain that—if detected—push the frontier in understanding the theory and phenomenology of the underlying observations and measurements). This could be the identification of poorly modeled phenomena, such as a core-collapse supernova signal in a gravitational-wave detector⁵⁻⁹, unexplained sea ice retreat¹⁰, extreme snow melts in the Arctic¹¹, or a new species of animal¹²⁻¹⁴. The challenges presented here aim to bring awareness to the difficulty of using ML for these types of problems: the options of well-labeled datasets or even the possibility of training through a systematic procedure, such as token masking, are often insufficient to answer this type of question. Despite that, there is tremendous potential for algorithms focused on anomaly detection to enable scientific discovery.

The process of using ML for anomaly detection, sometimes referred to as outlier detection, has a long history¹⁵⁻¹⁸. However, success has been limited due to the ambiguity of quantifying scientifically meaningful deviation^{19,20}. Nevertheless, several approaches have been tried^{18,20,21}, including the use of unsupervised algorithms²², which aim to build a mapping for what is regular through an embedding in a latent space, semi-supervised algorithms, which aim to use partially labeled data to construct representation space characteristic to the problem at hand²³, and weakly-supervised algorithms, which aim to play regions of data off of each other to find the largest incongruous regions of the data²⁴.

The ML challenge project brings together communities involved as part of the NSF Harnessing the Data Revolution (HDR) program, dedicated to connecting domain science with AI to enable groundbreaking scientific discoveries. All five HDR institutes were engaged in developing datasets with a tractable problem defined across their domains, each targeting a single anomaly detection problem and dataset. Ultimately, we narrowed the challenge to three datasets to ensure that they were sufficiently aligned that the same algorithmic toolkit was applicable to each of the challenges. Lastly, the most critical component of this challenge was the ability to hide pre-identified anomalies in such a way that there was no ambiguity in the anomaly definition in the final dataset. The challenge is to see if the proponents can find these pre-hidden anomalies with the smallest possible false positive rate.

The three challenges cover a wide variety of data types and problems, ranging from time series data, to photographic images, and satellite imagery over time, as described below.

- **Anomalous Gravitational-Wave Transients:** Short-duration ($O(1s)$) astrophysical signals present in public data from the two LIGO gravitational-wave detectors. The data that is received from each of the LIGO interferometers is a gravitational-wave strain from a signal, indicating a fractional stretching and squeezing of space-time on the order of 10^{-22} the measured length scale. The challenge is to find a signal that appears in both detectors and has broad features representative of an astrophysical signature. Such a signature may correspond, for example, to a core-collapse supernova, a white noise burst of unknown origin, or a black hole merger beyond our nominal modeling and expectation from general relativity.
- **Butterfly Hybrids:** Images of butterflies from two mimetic species, each with highly varied subspecies²⁵. It includes images of the various subspecies, as well as different “hybrids” among each species’ subspecies. The challenge is to find the anomalous butterfly hybrids without confusing them for subspecies of the other species.
- **Anomalous Sea Level Rise:** Satellite imagery²⁶ depicting observed sea levels over time. The challenge is to use this imagery to identify anomalous sea level rise at ports on the east coast of the US.

A diagram of the challenge flow is provided in Figure 1. For the challenge, teams must develop an algorithm that processes the data and outputs an anomaly score. The scoring is then determined by seeing if the algorithm is capable of singling out the anomalous features of the dataset. Each challenge has three core data subsets:

- **Training Dataset:** A publicly released subset consisting of the background (non-anomalous) cases and sample anomalous signals to enable the model construction. These signals do not encompass all variations of the challenge dataset, instead providing examples of what *might* be present. Moreover, the background dataset need not have the same proportional composition as the challenge dataset.

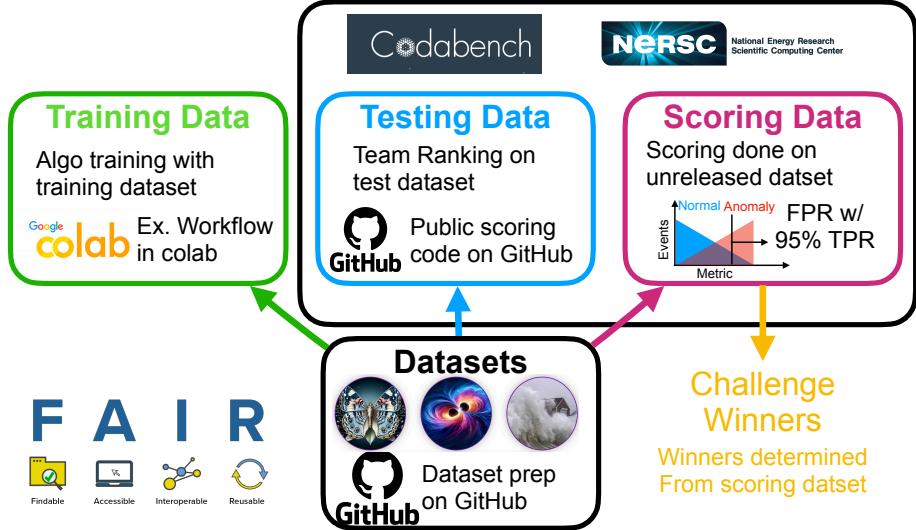


Figure 1. Overview of the flow of the ML challenge, and the datasets used for training. All preparation was explicitly made following FAIR principles, with the curation and data cleaning workflows stored on GitHub. The Codabench platform and NERSC computing cluster are used for scoring and ranking of submissions.

- **Testing Dataset:** During the competition, this subset is used for benchmarking algorithms and presenting a leaderboard on the competition platform. It largely consists of background non-anomalous events with some hidden anomalies (known-unknowns identifiable through additional information not available to participants). This dataset is not released publicly; however, participants may run their algorithm on it through the collaboration platform.
- **Challenge Dataset:** This subset is used to determine the winner. It is constructed in the same way as the testing dataset, but is not used before the final scoring. Thus, no participating team can overly tune the algorithm on the dataset.

A driving goal of this work is to codify a list of best practices that ensure workflows, no matter the scientific domain, are easy to follow and reproduce. Through this work, we endeavor to establish an exemplar for future ML challenges, preserving the level of Findability, Accessibilty, Interoperability, and Reusability (FAIR) present in this challenge in future challenges. Moreover, by ensuring a high level of FAIRness we aim to use these datasets for future benchmarking^{19,27,28}.

The ensuing paper presents the details of the challenge. In Section 2, we outline the motivation for the challenge and the scientific problem of anomaly detection. Subsection 2.1 highlights how the challenge was made FAIR. We then present the various datasets comprising the greater challenge in Section 3, followed by a detailed description of the evaluation metric in Section 4, and we end with the conclusions.

2 Scientific and FAIR Motivation

The anomaly detection challenges presented in this paper build on the traditional anomaly detection by providing unique, scientific elements that distinguish them from previous challenges. With the genomics problem, we utilize subtle features of images to replace the need for tedious DNA analysis to differentiate various species and subspecies. With the gravitational-wave challenge, we aim to codify all allowed laws of astrophysics within a time series to enhance the identification of unknown anomalies. With the climate science problem, we replace dedicated sensor data with large-scale, coarse satellite imagery and utilize this to find deviations from the norm. More conventional anomaly detection simply targets outliers; with each challenge here, we add a scientific element that pushes the boundaries of conventional approaches. The challenge of codifying scientific elements within ML adds an additional layer of complexity and highlights the uniqueness of this work.

For each component of this challenge, we constructed anomalies in a controlled manner, utilizing scientific knowledge or data not provided in the challenge. With the genomics problem, this involved using DNA. For the climate science problem, we utilized local water level sensors. For the astrophysics problem, knowledge of astrophysical phenomena is used to simulate hypothetical anomalous signals not observed within the data. With each of these challenges, our use of hidden labels, sensors and simulations is critical to the dataset construction. However, in real-world scenarios this additional information is not present, nor easily obtained. As a result, devising a strategy that can rely on the limited information—even if imperfect—would lead to impactful scientific results applicable to a large amount of unexplored data.

The choice of these datasets or tasks can be viewed as a seed to more complex problems. By constructing a FAIR challenge, with well-defined scoring metric and clear datasets, we aim to understand effective scientific approaches to anomaly detection. Future challenges can then be adapted to explore more difficult problems, and ultimately to true problems in the scientific domain where solutions might not yet exist at all.

2.1 FAIR

We endeavor to create an entirely **F**indable **A**ccessible **I**nteroperable **R**eusable, and ultimately reproducible, ML challenge. This choice of FAIRness extends beyond conventional FAIR datasets to include all aspects of the ML challenge. To this end, we require that all components—the challenge code, software environment, datasets, and metadata—adhere to the FAIR principles²⁹ and utilize methods described in ^{27,30,31}. Thus, we ensure the following for each dataset in the challenge:

- **Public Code:** All elements of our challenges are on GitHub, providing the full code used for running the challenges through Codabench³².
- **Common Base Container:** All challenges are built off the same base Docker container with common, standard software tools including tensorflow, Pytorch, and NumPy; it is specified publicly so that the container construction, itself, is reproducible.
- **Requirements list:** Any changes to the container in terms of packages—either through version upgrades or additional packages—must be specified in a requirements file submitted by participants.
- **Software Package Whitelist:** Only open-source software packages that the team has reviewed are eligible for installation; in addition to listing the software upgrade in the requirements, these packages are checked against a whitelist of allowed packages to ensure that all software is open-source and thus reproducible. Public GitHub repositories enabled participants to request whitelist additions, while ensuring that these would be documented along with the remaining challenge code.

We place similar conditions on submissions. Thus, requiring they consist of the trained model, with an inference script (containing a model class with at least two specified methods), and a list of software requirements. To illustrate how this is constructed, sample submissions, along with training code and notebooks are provided in sample repositories on GitHub, demonstrating the expected format of the final submissions. The challenge repositories are [Gravitational Waves](#) (includes the Docker container), [Hybrid Butterflies](#), and [Sea Level Rise](#).

For the ML Challenge framework, we choose Codabench³², a flexible open-source platform designed for benchmarking ML algorithms—particularly through ML challenges—that enables custom code for scoring and presentation of the results. Despite its robust deployment, the above listed points are not built into the Codabench framework, requiring us to develop these elements to ensure the challenge was fully FAIR. We thus publish all components of our challenges on GitHub, providing all source code used to run the challenges through Codabench³². The front end has a standard interface to view scoring, the leaderboard, and instructions for how to participate in the challenge. The inference backend is flexible, so submissions on the Codabench platform are run at National Energy Research Scientific Computing (NERSC) Center at Lawrence Berkeley National Laboratory³³. Though this adds additional complications in ensuring the submissions are FAIR, it furthers the reproducibility by standardizing the scoring and requiring the whitelist.

Through the use of both NERSC and the insistence on a FAIR framework, we endeavor to democratize the ML Challenge. Further extensions can be made to ensure FAIR principles can be adapted to other aspects of the ML challenge, such as model construction, in future editions. The final FAIRness component of the challenge is the requirement that all participants publish their challenge submissions (training code, requirements, model weights, etc.) in fully documented GitHub repositories (following the [Butterfly sample repository](#) setup). Additionally, when working from pre-trained models, we require that the participants select only those that are also open-source. This ensures that the challenge adheres to a more general notion of fairness in addition to the FAIR principles.

3 The Challenges

3.1 Detecting Anomalous Gravitational-Wave Signals

Gravitational waves as detected by the large ground-based interferometers LIGO³⁴, Virgo³⁵ and KAGRA³⁶ result from astrophysical phenomena involving the bulk motion of mass at high velocity. They appear as stretching and squeezing of the interferometers' arms due to perturbations of the spacetime metric. So far, all gravitational-wave detections that have been announced correspond to short-duration signals (referred to as transients, or bursts) from binary compact mergers involving black holes and neutron stars^{37–40}. These astrophysical systems and their corresponding gravitational-wave emission are well understood and their signatures are modeled so that templated searches (matched filtering) can perform optimal filtering for

them in interferometric data. A wealth of astrophysical sources that may potentially emit short-lived gravitational radiation for which we know very little, or close to nothing, on their signal morphologies exist. This includes core-collapse supernovae, neutron star glitches, emission associated with magnetars, other unknown astrophysical systems that power the Fast Radio Bursts, or even topological defects. These transient sources are referred to as unmodeled and they are prime candidates for ML-based anomaly detection approaches.

Gravitational-wave signals are extremely small. A typical gravitational wave on Earth induces a fractional differential arm change of approximately 10^{-22} , denoted strain. With a strain projected onto the km-scale arms of the interferometers, it results in arm displacements thousands of times smaller than a proton's diameter⁴¹. A single interferometer can achieve the sensitivity required to detect gravitational waves. However, it is limited in its capability to distinguish actual signals from large glitches in the detector. These glitches are often unmodeled transient "anomalies"⁴², denoted glitches that originate from events on Earth such as earthquakes or subtle sources of noise. Multiple detectors are required to veto glitches, which occur in a single detector, as opposed to a signal, which will be in both. The first gravitational-wave event was observed within the two detectors that constitute the Laser Interferometer Gravitational-wave Observatory (LIGO) in 2015^{34,43}. Since then, additional gravitational-wave detectors have become operational, allowing for many ways to confirm a signal across multiple instruments and enhancing our ability to identify gravitational-wave events.

The challenge presented here is to find unmodeled astrophysical transients using the two LIGO detectors. To detect such an event, a signal must be observed in both detectors with a time correlation consistent with a wave propagating at the speed of light and a waveform morphology that is correlated across detectors. Since no gravitational-wave event beyond binary black hole and neutron star mergers has been observed, we rely on astrophysical simulations to inject a variety of synthetic signals into the dataset for identification^{7,44–49}. Unmodeled searches have been widely used in the gravitational-wave community and reported in observational papers^{9,50–53}. Furthermore, recent efforts to perform AI-based anomaly detection have emerged^{5,6,54–60}; this is the primary focus of the challenge presented here.

The Anomaly Detection Challenge utilizes data from LIGO's O3a observing run, consisting of calibrated strain time-series recorded by the two LIGO interferometers. The provided dataset has undergone a series of pre-processing steps: whitening, band-passing, and removing from the dataset altogether 1s worth of data around the times of gravitational-wave detections made and published by the LIGO-Virgo-KAGRA collaborations. Additionally, the datasets contain simulated signals injected into the real detector noise, as described in Section 3.1.1. Participants are required to train their models for anomaly detection primarily using the background data and can improve their ability to identify anomalies through injected simulated signals.

3.1.1 Data samples

The dataset used in this challenge was collected by the LIGO Hanford and LIGO Livingston⁶¹. We utilized publicly available data from the beginning of observing run O3a⁶², corresponding to GPS times 1238166018 to 1238170289. The time-series data were sampled at 4096 Hz, and processed to remove and create a separate dataset of transient instrumental artifacts (glitches). We extracted 4 s segments of artifact-free data to serve as the baseline for the injection of signals. We define three data classes representing signals and background signatures:

- **Binaries** – Simulated Binary Black Holes (BBH) signals injected into the real background noise, as shown in Fig. 3 (top).
- **Background** – Background from O3a with excess power glitches and known gravitational-wave events removed, as shown in Fig. 2.
- **Sine-Gaussian (SG)** – Generic ad-hoc signal model used as a proxy to simulate gravitational-wave sources we know little about, as shown in Fig. 3 (bottom).

To generate samples of BBH and SG signals, we perform simulations of the two polarization modes, h_+ and h_\times , which describe the strain-induced distortions in spacetime. We then sample sky localizations uniformly in the sky, generate the polarization modes at the specified sky location, and then inject the projected modes into the two LIGO detectors.

We whiten the data and band-pass them within the frequency range of 30–1500 Hz. After applying these filters, we remove 1 s intervals from each end of the data samples to eliminate edge effects from pre-processing. The remaining 2 s samples, each containing either an injected signal, pure background or an SG, were used to generate training data. To obtain a set of windows suitable for training, we extract 200 data points (total duration of 50 ms sampled at 4096 Hz) from each sample.

3.1.2 Challenge

The challenge is to identify astrophysical signals that are neither background noise nor glitches within the detector. Astrophysical signatures have the characteristic features shown in Figure 3 whereby the two detectors yield inverted signals with roughly the same amplitude. These signals are time-aligned such that the arrival of the signals corresponds to the same gravitational wave propagating from a given location in the sky. The sky location will change the relative amplitudes, along with the knowledge of

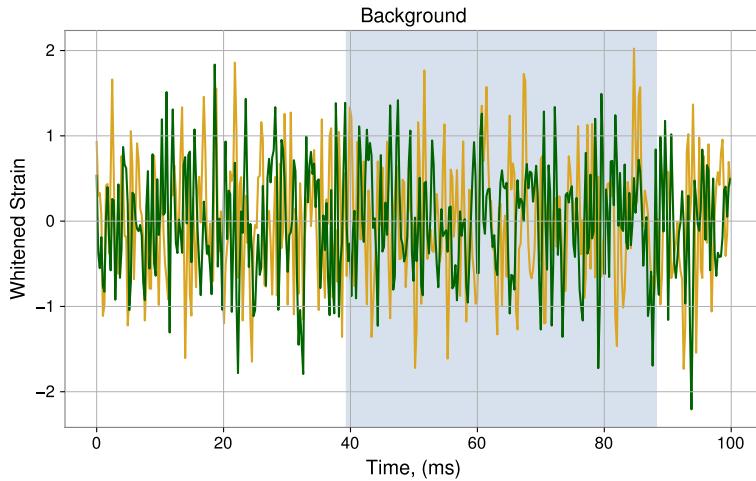


Figure 2. Example of the background strain. The light blue shading highlights an example region that is passed as input to the autoencoders for training. The yellow and green lines indicate the strain from the two LIGO detectors Hanford (yellow) and Livingston (green).

the noise levels at each of the respective detectors making it possible for an asymmetry in amplitudes between the detectors. The signals themselves are limited to frequency ranges from roughly 10 Hz to 1000 Hz.

3.2 Detecting Hybrid Butterflies

The criteria for elevating subspecies to species remain an open question in evolutionary biology. More broadly, these taxonomic distinctions have significant implications for conservation efforts, as they influence which populations receive legal protection under the Endangered Species Act⁶³. A key challenge in resolving species boundaries is our ability to detect and classify subspecies, particularly in cases where hybridization occurs. The question of how hybridization contributes to the evolution and maintenance of new species has intrigued scientists since Charles Darwin, who pondered the variability and stability of hybrid traits. Gregor Mendel later addressed this question through his experiments, demonstrating that hybrid offspring do not always exhibit intermediate traits, instead resembling one parent, following predictable patterns of segregation⁶⁴.

These principles of inheritance remain central to modern hybrid detection, particularly in distinguishing whether hybrids between subspecies exhibit continuous variation or Mendelian inheritance patterns. This challenge is designed to leverage anomaly detection algorithms to detect hybrids produced by parents of hybridizing subspecies. By refining hybrid detection methods, we can improve taxonomic classification and enhance conservation strategies for threatened populations.

3.2.1 Background

H. erato and *H. melpomene* are two **species** of *Heliconius* butterflies (Order: Lepidoptera; Family: Nymphalidae) that diverged approximately 12 million years ago. Both species are chemically defended (unpalatable) and signal their toxicity through bright warning coloration on their wings. Interestingly, they have evolved to resemble each other, forming a classic example of Müllerian mimicry, in which two or more toxic species share similar warning signals to reinforce predator learning⁶⁶. This is distinct from **Batesian** mimicry, where a palatable species mimics an unpalatable one to gain protection. Despite strong predation pressure to maintain similar warning patterns, *Heliconius* species and their color morphs have diversified across Central and South America in what biologists term *adaptive radiation*. That is, distinct geographic regions—sometimes microhabitats within these regions—harbor unique assemblages of *H. erato* and *H. melpomene*, each displaying region-specific wing color patterns that serve as distinguishing characteristics.

The *visually different appearances among subspecies in different regions* and *visually mimicking appearances between species in the same regions* result in large intra-species variation within *H. melpomene* (*H. erato*) and small inter-species variation between *H. melpomene* and *H. erato* (Figure 4). This phenomenon has attracted attention in biology, ecology, computer vision, and machine learning to study the visual traits distinguishing subspecies and species of *Heliconius* butterflies from each other.

3.2.2 Anomaly cases

Heliconius butterflies exhibit strong assortative mating, preferentially choosing mates with the same wing color pattern. This behavior is reinforced by natural selection—**hybrids** resulting from crosses between individuals with different patterns often

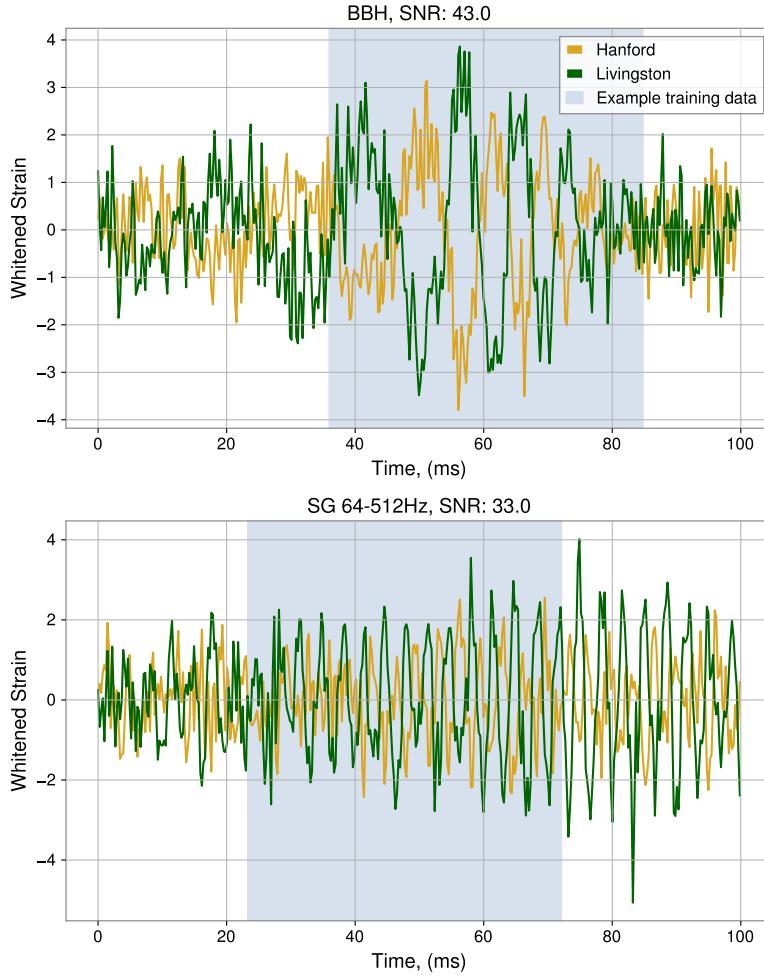


Figure 3. Example of signal-like classes: Binary Black Hole Mergers (top) and sine-gaussian (bottom) strains. The light blue shading highlights an example region that is passed as input to the autoencoders for training.

face a survival disadvantage. Their unfamiliar wing patterns are not recognized by local predators, increasing their likelihood of being removed from the population. However, where subspecies come into contact or overlap, interbreeding still occurs, producing hybrids with diverse phenotypes. The visual appearance of hybrids can vary—some resemble one parent more, while others exhibit an intermediate or novel pattern. Historically, these hybrid forms were labeled as aberrations or given distinct “form names.” If such hybrids become frequent enough to evade predation, they may persist in the population and potentially contribute to speciation. It is these hybrids, aberrations, and forms that we aim to detect.

3.2.3 Scenario

We design our competition to simulate a real-world biological scenario. Suppose a biologist studies *Heliconius melpomene* and *Heliconius erato*, aiming to understand the mimicry phenomenon and the different color patterns of subspecies. One day, she finds that a subset of the butterfly collections looks slightly abnormal in their color patterns. After further investigation, she discovers these samples are hybrids generated by different, rarely observed, subspecies of *H. erato*. She realizes there may also be hybrids among the *H. melpomene* specimens, but has fewer samples. Since in theory, there are quadratically many cases of hybrids and her current collection only covers a small subset of them, she seeks an anomaly detection system to automatically identify (unknown) hybrid cases from future collections of Heliconius butterflies—both *H. erato* and *H. melpomene*.

3.2.4 Data and Splits

Data

The dataset is comprised of a subset of the Heliconius Collection (Cambridge Butterfly)⁶⁷, a compilation of images from Chris Jiggins’ research group at the University of Cambridge^{68–91}. It encompasses two aspects of biological development and

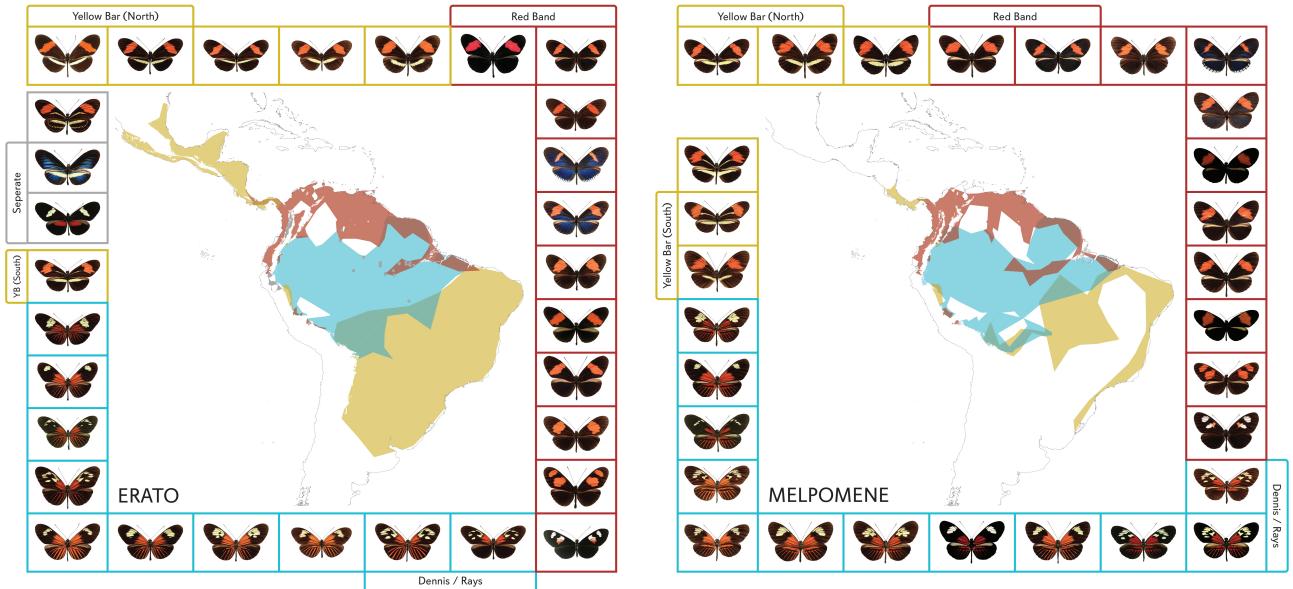


Figure 4. The parallel radiation of mimetic color-pattern complexes in *Heliconius erato* (left) and *Heliconius melpomene* (right). Each region demarcates the boundaries of a different color pattern. See⁶⁵ for more details.

evolutionary change: (1) hybridization—the main theme of this challenge—and (2) mimicry.

1. Hybridization:

Geographic or habitat separation of a species population can lead to species variation developing into different subspecies. The visual appearances (e.g., color patterns on the wings) of these subspecies can be drastically different. In this challenge, offspring produced by the same-subspecies parents (**non-hybrids**) are treated as normal cases because they are far more frequently observed. In contrast, **hybrids** are treated as **anomaly** cases, not only because they are much less frequently observed—with some combinations not yet observed—but also because their visual appearances are much more variant and hardly predictive. We emphasize again that the parents of each hybrid child are from different subspecies of the *same* species (either *H. melpomene* or *H. erato*).

2. Butterfly mimicry

Meanwhile, geographic or habitat overlap can also lead to increased visual similarity between species, known as *mimicry*. This challenge goes beyond developing an anomaly detection algorithm to distinguish between hybrids and non-hybrids for one species, investigating further whether such an algorithm is generalizable to the visually mimetic species.

Splits

The training data comprises images from all the *Heliconius erato* subspecies and the **signal hybrid**: a specific combination of the parent subspecies that has the most images (these hybrids are the most common *within this particular dataset, not necessarily in general*; see Figure 5). All other hybrids, **non-signal hybrids**, are excluded from the training data.

We consider the following two sets of images in the test set, as illustrated in Figure 6.

1. All Species A (*Heliconius erato*) subspecies and their hybrids, including the signal hybrid and the non-signal hybrids.
2. Two subspecies of Species B (*H. melpomene*) and their hybrid. These subspecies of *H. melpomene* are those that mimic the parent subspecies of the signal hybrid of *H. erato*.

3.3 Detecting Anomalous Sea Level Rise Events

The US East Coast is a region particularly vulnerable to coastal flooding and storm surges. Daily tide gauge data from the National Data Buoy Center (NDBC)⁹² provide essential observations, including tidal variations, storm surges, and long-term

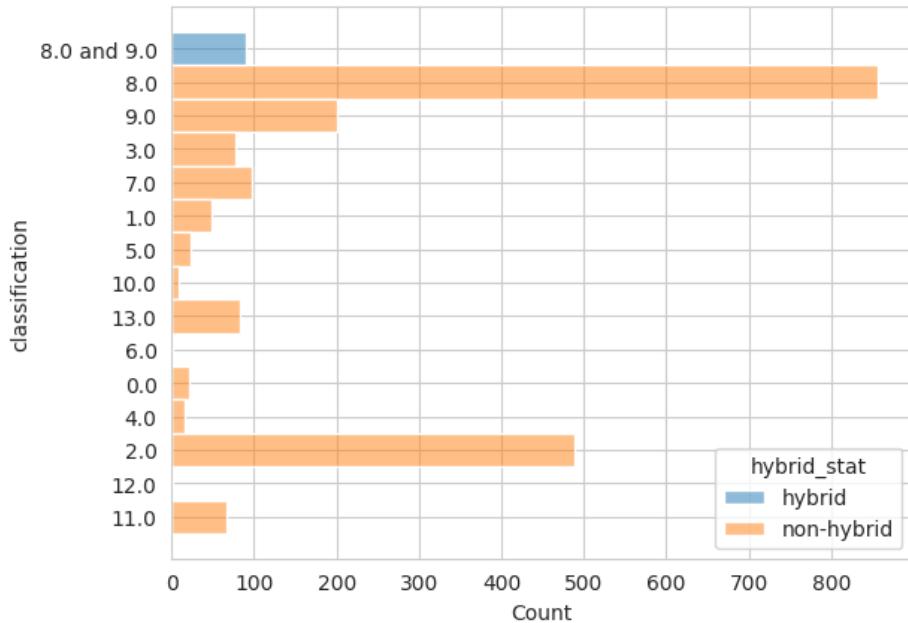


Figure 5. Distribution of anonymized *H. erato* subspecies included in the training data. This figure was generated in a starter Jupyter Noteboook provided to participants in the [sample repository](#).

trends, forming the foundation for this challenge. Additionally, long-term sea level records are invaluable for studying climate change impacts, providing evidence of global warming effects like melting polar ice and thermal expansion of seawater. These observations support the resilience and sustainability of coastal economies, which rely heavily on tourism, fishing, and shipping industries. Predicting sea level anomaly events, such as extreme storm surges or unusually high tides, is challenging along the low-lying US East Coast region due to the complex interplay of atmospheric, oceanic, and climatic factors. These events are influenced by a combination of wind patterns, atmospheric pressure changes, and ocean currents, making accurate forecasting difficult with traditional methods. By leveraging ML to process real-time data from NDBC buoys, historical sea level records, and meteorological information, participants are tasked with detecting and predicting anomalous sea-level rise events caused by factors such as hurricanes, mid-latitude storms, or long-term climatic phenomena like the El Niño Southern Oscillation (ENSO)⁹³. If accurately detected, these anomalies can be vital in enhancing coastal community preparedness and safeguarding lives, infrastructure, and economic activities.

3.3.1 Objective

The objective is to predict anomalous sea-level observations from daily tide gauge data along the US East Coast affected by changes in the sea-level elevation values on the Atlantic Ocean. The challenge leverages a comprehensive training dataset that spans 20 years of daily sea-level measurements from 12 coastal stations along the US East Coast, complemented by regional satellite sea-level elevation data in the North Atlantic. The satellite images are referred to as the Copernicus dataset²⁶, which offers a broader spatial context, enabling participants to incorporate regional oceanic conditions into their models. Therefore, participants are required to develop models that take the satellite maps of spatial information of sea level anomalies over the North Atlantic as input and predict dates when coastal stations along the US East Coast are above a flooding threshold for the subsequent period of 10 years, which is the test dataset. Each submission must specify whether a flooding level anomaly occurred daily at each station. This task tests the participants' ability to build robust predictive models and emphasizes the importance of precision and recall in capturing true anomalies.

3.3.2 Evaluation

Evaluation metrics, including the average true positive rate, average false positive rate, and the F1 score, guide the assessment of model performance. The challenge fosters innovation, collaboration, and the development of scalable solutions with direct societal benefits, ultimately advancing our ability to predict and respond to sea-level anomalies, thereby strengthening the resilience of vulnerable coastal regions.

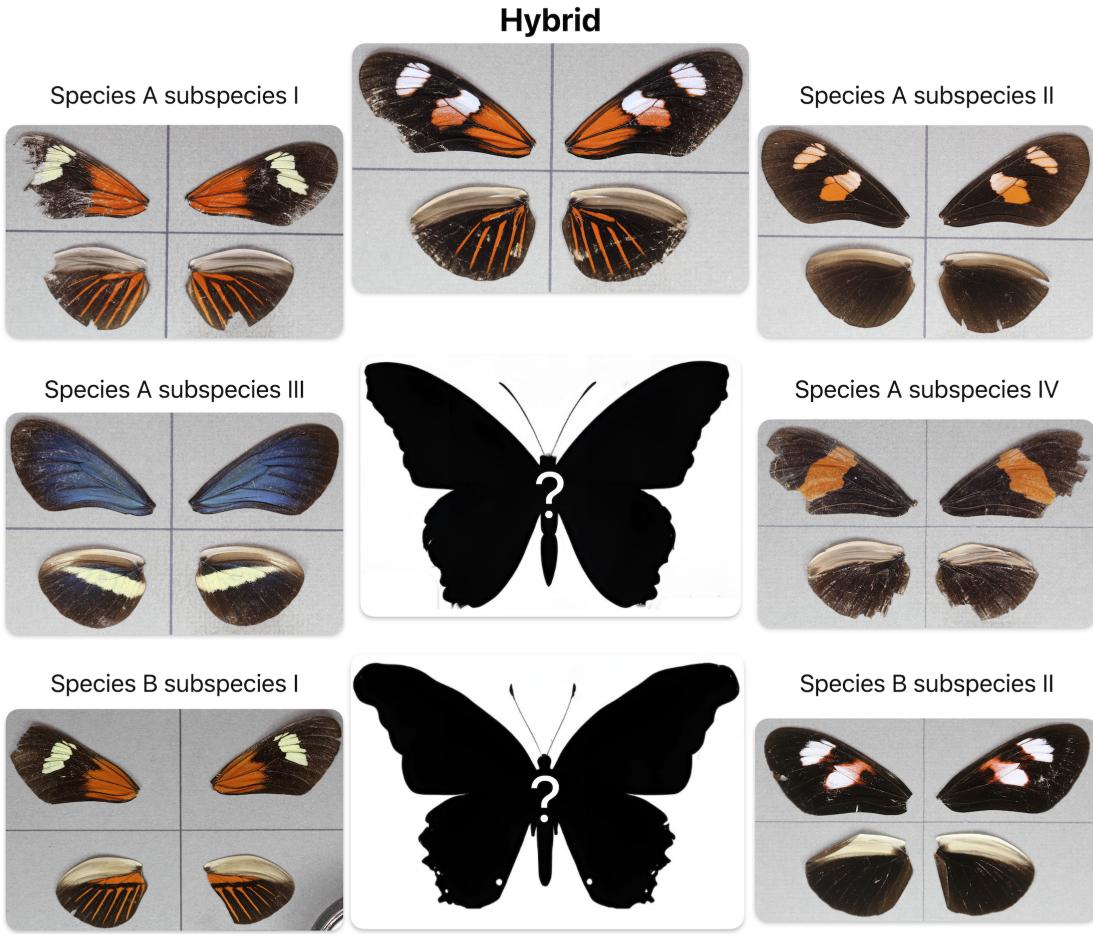


Figure 6. For the challenge, participants were only provided the labels “Species A subspecies #”. These four subspecies of Species A (*H. erato*) are seen in training, along with the signal hybrid (hybrid of *Species A subspecies I & II*), but can we detect the hybrid of *Species A subspecies III & IV* when we do not know what they look like? Additionally, can we distinguish the Species B (*H. melpomene*) subspecies I & II hybrids, considering these subspecies mimic *Species A subspecies I & II*, respectively. Images are from^{68,73,75,78,80}. The hybrid graphics (black butterflies) were generated using Canva Magic Media AI, then manually edited.

3.3.3 Datasets

The core dataset includes hourly sea-level measurements from 12 tide gauge stations from 1993 to 2013 for each station. Gridded Sea Level Anomalies (SLA) computed for a 20-year mean are also provided. These SLA values are estimated using Optimal Interpolation, merging along-track measurements from multiple altimeter missions processed by the DUACS multission altimeter data system. Additional variables, such as Absolute Dynamic Topography and geostrophic currents (both absolute and anomalous), are also included, making the dataset suitable for delayed-time applications and allowing participants to explore the dynamic relationships between sea-level changes and broader oceanic processes.

Each coastal station’s data is represented as a distinct time series, and a separate column identifies the dates associated with known anomalies. This dataset captures critical sea-level fluctuations caused by various climatic and oceanic phenomena⁹⁴. Additionally, the Copernicus sea-level gridded satellite observations dataset²⁶ contributes sea-level elevation values from the Atlantic Ocean for the same period, offering a broader contextual dataset to enhance the predictive capabilities of the models.

The training dataset is a labeled subset of the time series data from the tide gauge stations. This dataset includes both the raw sea-level measurements and the associated anomaly labels. The combination of labeled anomalies and continuous measurements allows for the exploration of various ML approaches, such as supervised learning, time series analysis, and

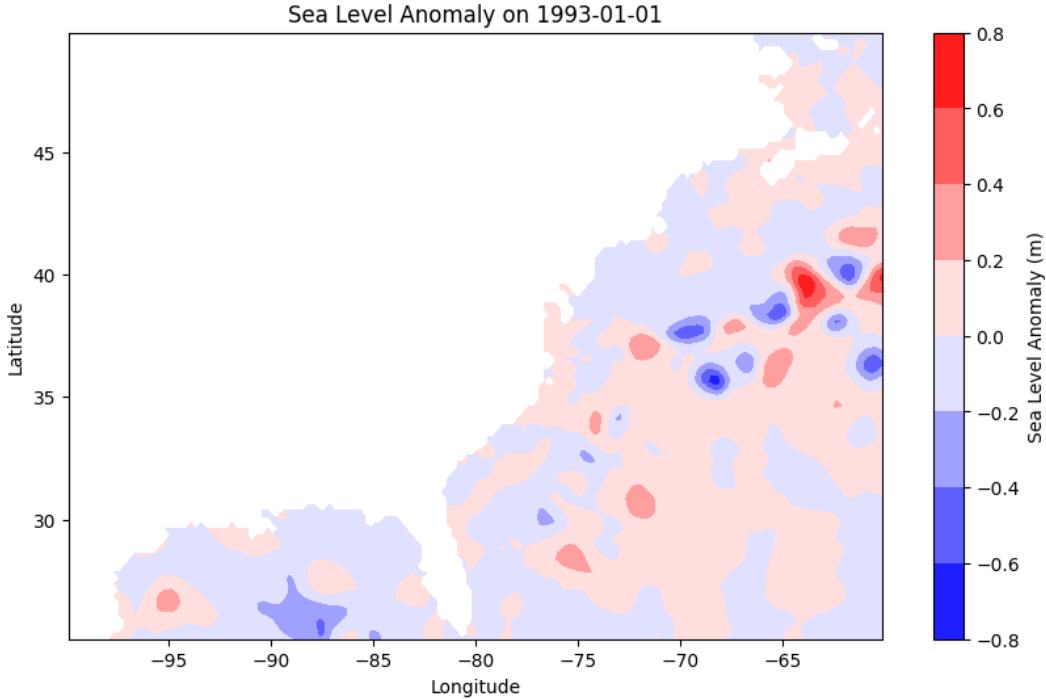


Figure 7. Depiction of the sea levels on the Eastern Seaboard from January 1st, 1993 obtained from satellite images of the US eastern seaboard.

feature engineering.

The public testing dataset includes a subset of the time series data from each station but does not contain labeled anomalies. This dataset allows participants to refine their methodologies. The challenge dataset, on the other hand, includes hidden anomalies and will be used to evaluate the final performance of submitted models.

By offering labeled and unlabeled data across diverse coastal stations, this challenge emphasizes the development of models that can generalize well to unseen data and effectively identify anomalies in complex, multivariate time series. Following FAIR principles ensures that the solutions generated will have practical applications in predicting and responding to coastal sea-level anomalies, ultimately contributing to improved resilience and preparedness in vulnerable regions.

4 Metrics

The evaluation metric is the false positive rate (FPR) at a specified true positive rate (TPR) when detecting anomalies. The specified TPR, which we denote by TP%, varies by challenge, as indicated in Table 1. Given a test set with both normal and anomalous signals/instances, this metric treats the normal cases as positive cases and finds a score threshold τ such that TP% of the normal signals have scores $s(\mathbf{x}) \geq \tau$. It then counts the percentage of anomalous signals whose score $s(\mathbf{x}) \geq \tau$ (i.e., FPR). The higher the FPR is, the poorer the anomaly detection algorithm performs.

| Dataset | True Positive Rate Threshold |
|----------------------------|------------------------------|
| Gravitational-wave signals | 90% |
| Hybrid butterflies | 95% |
| Sea level rise | 95% |

Table 1. True Positive Rate (TPR) threshold chosen for each dataset to calculate the challenge scores.

5 Conclusion

In summary, we have developed three scientific datasets targeting anomaly detection and deployed machine learning challenges. These datasets provide concrete examples of how anomaly detection plays an important role in scientific discovery. Moreover,

the implications of a successful model within any of these challenges would be direct in the scientific domain, and—in some cases—profound. In preparing these challenges, we have emphasized reproducibility partnered with the use of **Findable Accessible Interoperable** and **Reusable** principles. Our current challenge has recently completed, with more than 600 participating teams spread over the three different datasets.

The resulting solutions can have broad impacts in many domains. Time series anomalies within the gravitational-wave problem have direct implications on anomalous neural activity, such as sleep spindles⁹⁵. Equivalently, the hybrid detection problem has direct implications for processing medical images and diagnoses^{96,97}. Lastly, the climate science problem can be applied more broadly to satellite data to predict future catastrophic effects^{98,99}.

When considering future science-based challenges, it is critical to ensure a public codebase that includes the full scoring and example submissions is available. The use of a computing backend with sufficient GPUs has been essential. Though raising the complexity, it increases reproducibility. Going forward, we recommend that future challenges consider these essential elements to expedite the path from a challenging problem to scientific discovery.

References

1. Pang, G., Shen, C., Cao, L. & Hengel, A. V. D. Deep learning for anomaly detection: A review. *ACM Comput. Surv.* **54**, 1–38, DOI: [10.1145/3439950](https://doi.org/10.1145/3439950) (2021).
2. Boniol, P., Liu, Q., Huang, M., Palpanas, T. & Paparrizos, J. Dive into time-series anomaly detection: A decade review (2024). [2412.20512](https://arxiv.org/abs/2412.20512).
3. Blázquez-García, A., Conde, A., Mori, U. & Lozano, J. A. A review on outlier/anomaly detection in time series data (2020). [2002.04236](https://arxiv.org/abs/2002.04236).
4. Krenn, M. *et al.* On scientific understanding with artificial intelligence. *Nat. Rev. Phys.* **4**, 761–769, DOI: [10.1038/s42254-022-00518-3](https://doi.org/10.1038/s42254-022-00518-3) (2022).
5. Raikman, R. *et al.* GWAK: gravitational-wave anomalous knowledge with recurrent autoencoders. *Mach. Learn. Sci. Tech.* **5**, 025020, DOI: [10.1088/2632-2153/ad3a31](https://doi.org/10.1088/2632-2153/ad3a31) (2024). [2309.11537](https://arxiv.org/abs/2309.11537).
6. Raikman, R. *et al.* A neural network-based search for unmodeled transients in ligo-virgo-kagra’s third observing run (2024). [2412.19883](https://arxiv.org/abs/2412.19883).
7. Abbott, B. P. *et al.* Search for transient gravitational-wave signals associated with magnetar bursts during advanced ligo’s second observing run. *Astrophys. J.* **874**, 163, DOI: [10.3847/1538-4357/ab0e15](https://doi.org/10.3847/1538-4357/ab0e15) (2019).
8. Sutton, P. J. *et al.* X-pipeline: an analysis package for autonomous gravitational-wave burst searches. *New J. Phys.* **12**, 053034, DOI: [10.1088/1367-2630/12/5/053034](https://doi.org/10.1088/1367-2630/12/5/053034) (2010).
9. Lynch, R., Vitale, S., Essick, R., Katsavounidis, E. & Robinet, F. Information-theoretic approach to the gravitational-wave burst detection problem. *Phys. Rev. D* **95**, 104046, DOI: [10.1103/PhysRevD.95.104046](https://doi.org/10.1103/PhysRevD.95.104046) (2017).
10. Devnath, M. K., Chakraborty, S. & Janeja, V. P. Cmad: Advancing understanding of geospatial clusters of anomalous melt events in sea ice extent. In *Proceedings of the 32nd ACM International Conference on Advances in Geographic Information Systems*, 282–292 (2024).
11. Ale, T., Janeja, V. P. & Schlegel, N.-J. Harnessing feature clustering for enhanced anomaly detection with variational autoencoder and dynamic threshold. In *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, 8692–8696, DOI: [10.1109/IGARSS53475.2024.10640794](https://doi.org/10.1109/IGARSS53475.2024.10640794) (2024).
12. Souza, F. C. *et al.* New anomaly records in anurans, with a review for brazil. *Herpetol. Notes* **14**, 31–41 (2021).
13. Ito, E. *et al.* Phenotype anomaly detection for biological dynamics data using a deep generative model. In *International Conference on Artificial Neural Networks*, 432–444 (Springer, 2022).
14. Pastore, V. P., Zimmerman, T. G., Biswas, S. K. & Bianco, S. Annotation-free learning of plankton for classification and anomaly detection. *Sci. reports* **10**, 12142 (2020).
15. Hendrycks, D., Mazeika, M. & Dietterich, T. Deep anomaly detection with outlier exposure. In *International Conference on Learning Representations* (2019).
16. Schlegl, T., Seeböck, P., Waldstein, S. M., Schmidt-Erfurth, U. & Langs, G. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In *International conference on information processing in medical imaging*, 146–157 (Springer, 2017).
17. Zong, B. *et al.* Deep autoencoding gaussian mixture model for unsupervised anomaly detection. In *International conference on learning representations* (2018).

18. Yang, J., Zhou, K., Li, Y. & Liu, Z. Generalized out-of-distribution detection: A survey. *Int. J. Comput. Vis.* **132**, 5635–5662 (2024).
19. Galil, I., Dabbah, M. & El-Yaniv, R. A framework for benchmarking class-out-of-distribution detection and its application to imangenet. In *International Conference on Learning Representations* (2023).
20. Salehi, M. *et al.* A unified survey on anomaly, novelty, open-set, and out-of-distribution detection: Solutions and future challenges. *Transactions on Mach. Learn. Res.* (2022).
21. Dillmann, S., Martínez-Galarza, J. R., Soria, R., Stefano, R. D. & Kashyap, V. L. Representation learning for time-domain high-energy astrophysics: Discovery of extragalactic fast x-ray transient xrt 200515. *Mon. Notices Royal Astron. Soc.* **537**, 931–955, DOI: [10.1093/mnras/stae2808](https://doi.org/10.1093/mnras/stae2808) (2024).
22. Cui, Y., Liu, Z. & Lian, S. A survey on unsupervised anomaly detection algorithms for industrial images. *IEEE Access* **11**, 55297–55315, DOI: [10.1109/access.2023.3282993](https://doi.org/10.1109/access.2023.3282993) (2023).
23. Kuusela, M. *et al.* Semi-supervised anomaly detection – towards model-independent searches of new physics. *J. Physics: Conf. Ser.* **368**, 012032, DOI: [10.1088/1742-6596/368/1/012032](https://doi.org/10.1088/1742-6596/368/1/012032) (2012).
24. Jiang, M. *et al.* Weakly supervised anomaly detection: A survey (2023). [2302.04549](#).
25. Consortium, T. H. G. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* **487**, 94–98 (2012).
26. Copernicus Climate Data Store. Sea level gridded data from satellite observations (2024). Accessed: 2024-05-28. Available at: <https://cds.climate.copernicus.eu/datasets/satellite-sea-level-global>.
27. E. A. Huerta, B. Blaiszik, L.C. Brinson, K.E. Bouchard, D. Diaz, C. Doglioni, J.M. Duarte, M. Emani, I. Foster, G. Fox, P. Harris, L. Heinrich, S. Jha, D.S. Katz, V. Kindratenko, C.R. Kirkpatrick, K. Lassila-Perini, R.K. Madduri, M.S. Neubauer, F.E. Psomopoulos, A. Roy, O. Rübel, Z. Zhao, R. Zhu. FAIR for AI: An interdisciplinary and international community building perspective. *Sci. Data* **10** (2023).
28. Zhao, D. *et al.* A taxonomy of challenges to curating fair datasets (2024). [2406.06407](#).
29. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018, DOI: [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18) (2016).
30. Chen, Y. *et al.* A fair and ai-ready higgs boson decay dataset. *Sci. Data* **9**, DOI: [10.1038/s41597-021-01109-0](https://doi.org/10.1038/s41597-021-01109-0) (2022).
31. J. Duarte, H. Li, A. Roy, R. Zhu, E.A. Huerta, D. Diaz, P. Harris, R. Kansal, D.S. Katz, I.H. Kavoori, V. Kindratenko, F. Mokhtar, Mark S. Neubauer, S.E. Park, M. Quinnan, R. Rusack, Z. Zhao. FAIR AI Models in High Energy Physics. *Mach. Learn. Sci. Tech.* **4**, 045062, DOI: [10.1088/2632-2153/ad12e3](https://doi.org/10.1088/2632-2153/ad12e3) (2023). [2212.05081](#).
32. Xu, Z. *et al.* Codabench: Flexible, easy-to-use, and reproducible meta-benchmark platform. *Patterns (New York, N.Y.)* **3**, 100543, DOI: [10.1016/j.patter.2022.100543](https://doi.org/10.1016/j.patter.2022.100543) (2022).
33. Bhimji, W. *et al.* FAIR Universe HiggsML Uncertainty Challenge Competition. *White Pap.* (2024). [2410.02867](#).
34. Abbott, B. P. *et al.* Observation of Gravitational Waves from a Binary Black Hole Merger. *Phys. Rev. Lett.* **116**, 061102, DOI: [10.1103/PhysRevLett.116.061102](https://doi.org/10.1103/PhysRevLett.116.061102) (2016). [1602.03837](#).
35. Acernese, F. *et al.* Advanced Virgo: a second-generation interferometric gravitational wave detector. *Class. Quant. Grav.* **32**, 024001, DOI: [10.1088/0264-9381/32/2/024001](https://doi.org/10.1088/0264-9381/32/2/024001) (2015). [1408.3978](#).
36. Akutsu, T. *et al.* Overview of KAGRA: Detector design and construction history. *PTEP* **2021**, 05A101, DOI: [10.1093/ptep/ptaa125](https://doi.org/10.1093/ptep/ptaa125) (2021). [2005.05574](#).
37. Abbott, R. *et al.* GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run. *Phys. Rev. X* **11**, 021053, DOI: [10.1103/PhysRevX.11.021053](https://doi.org/10.1103/PhysRevX.11.021053) (2021). [2010.14527](#).
38. Abbott, R. *et al.* GWTC-2.1: Deep extended catalog of compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run. *Phys. Rev. D* **109**, 022001, DOI: [10.1103/PhysRevD.109.022001](https://doi.org/10.1103/PhysRevD.109.022001) (2024). [2108.01045](#).
39. Abbott, B. P., Abbott, R. & Abbott, T. D. e. a. Gwtc-1: A gravitational-wave transient catalog of compact binary mergers observed by LIGO and virgo during the first and second observing runs. *Phys. Rev. X* **9**, 031040, DOI: [10.1103/PhysRevX.9.031040](https://doi.org/10.1103/PhysRevX.9.031040) (2019).
40. Abbott, R. *et al.* GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run. *arXiv e-prints* (2021). [2111.03606](#).

41. Buikema, A. *et al.* Sensitivity and performance of the advanced LIGO detectors in the third observing run. *Phys. Rev. D* **102**, 062003, DOI: [10.1103/PhysRevD.102.062003](https://doi.org/10.1103/PhysRevD.102.062003) (2020).
42. Cabero, M. *et al.* Blip glitches in Advanced LIGO data. *Class. Quant. Grav.* **36**, 15, DOI: [10.1088/1361-6382/ab2e14](https://doi.org/10.1088/1361-6382/ab2e14) (2019). [1901.05093](https://arxiv.org/abs/1901.05093).
43. <https://www.nobelprize.org/prizes/physics/2017/press-release/>.
44. Abbott, B. P. *et al.* Optically targeted search for gravitational waves emitted by core-collapse supernovae during the first and second observing runs of advanced LIGO and advanced Virgo. *Phys. Rev. D* **101**, 084002, DOI: [10.1103/PhysRevD.101.084002](https://doi.org/10.1103/PhysRevD.101.084002) (2020).
45. Abadie, J. *et al.* Search for gravitational waves associated with the August 2006 timing glitch of the Vela pulsar. *Phys. Rev. D* **83**, 042001, DOI: [10.1103/PhysRevD.83.042001](https://doi.org/10.1103/PhysRevD.83.042001) (2011).
46. Ebersold, M. & Tiwari, S. Search for nonlinear memory from subsolar mass compact binary mergers. *Phys. Rev. D* **101**, 104041, DOI: [10.1103/PhysRevD.101.104041](https://doi.org/10.1103/PhysRevD.101.104041) (2020).
47. Abbott, R. *et al.* Constraints on cosmic strings using data from the third Advanced LIGO-Virgo observing run. *Phys. Rev. Lett.* **126**, 241102 (2021).
48. Robinet, F. *et al.* Omicron: a tool to characterize transient noise in gravitational-wave detectors. *SoftwareX* **12**, 100620, DOI: [10.1016/j.softx.2020.100620](https://doi.org/10.1016/j.softx.2020.100620) (2020). [2007.11374](https://arxiv.org/abs/2007.11374).
49. Nitz, A. *et al.* gwastro/pycbc: Pycbc release v1.16.9, DOI: [10.5281/zenodo.3993665](https://doi.org/10.5281/zenodo.3993665) (2020).
50. Klimenko, S., Yakushin, I., Mercer, A. & Mitselmakher, G. Coherent method for detection of gravitational wave bursts. *Class. Quant. Grav.* **25**, 114029, DOI: [10.1088/0264-9381/25/11/114029](https://doi.org/10.1088/0264-9381/25/11/114029) (2008). [0802.3232](https://arxiv.org/abs/0802.3232).
51. Macquet, A., Bizouard, M.-A., Christensen, N. & Coughlin, M. Searches for long-duration gravitational wave transients in LIGO and Virgo data. In *55th Rencontres de Moriond on Gravitation* (2021). [2108.10033](https://arxiv.org/abs/2108.10033).
52. The LIGO Scientific Collaboration, t. V. C. & the KAGRA Collaboration. All-sky search for short gravitational-wave bursts in the third advanced LIGO and advanced virgo run. *Phys. Rev. D* **104**, DOI: [10.1103/physrevd.104.122004](https://doi.org/10.1103/physrevd.104.122004) (2021).
53. Abbott, R. *et al.* All-sky search for long-duration gravitational-wave bursts in the third Advanced LIGO and Advanced Virgo run. *Phys. Rev. D* **104**, 102001, DOI: [10.1103/PhysRevD.104.102001](https://doi.org/10.1103/PhysRevD.104.102001) (2021). [2107.13796](https://arxiv.org/abs/2107.13796).
54. Verma, C., Reza, A., Gaur, G., Krishnaswamy, D. & Caudill, S. Detection of gravitational wave signals from precessing binary black hole systems using convolutional neural network (2024). [2206.12673](https://arxiv.org/abs/2206.12673).
55. Baltus, G. *et al.* Convolutional neural networks for the detection of the early inspiral of a gravitational-wave signal. *Phys. Rev. D* **103**, 102003, DOI: [10.1103/PhysRevD.103.102003](https://doi.org/10.1103/PhysRevD.103.102003) (2021).
56. Krastev, P. G. Real-time detection of gravitational waves from binary neutron stars using artificial neural networks. *Phys. Lett. B* **803**, 135330, DOI: [10.1016/j.physletb.2020.135330](https://doi.org/10.1016/j.physletb.2020.135330) (2020).
57. George, D. & Huerta, E. A. Deep neural networks to enable real-time multimessenger astrophysics. *Phys. Rev. D* **97**, 044039, DOI: [10.1103/PhysRevD.97.044039](https://doi.org/10.1103/PhysRevD.97.044039) (2018).
58. Nousi, P. *et al.* Deep residual networks for gravitational wave detection. *Phys. Rev. D* **108**, 024022, DOI: [10.1103/PhysRevD.108.024022](https://doi.org/10.1103/PhysRevD.108.024022) (2023).
59. Skliris, V., Norman, M. R. K. & Sutton, P. J. Toward real-time detection of unmodeled gravitational wave transients using convolutional neural networks. *Phys. Rev. D* **110**, 104034, DOI: [10.1103/PhysRevD.110.104034](https://doi.org/10.1103/PhysRevD.110.104034) (2024). [2009.14611](https://arxiv.org/abs/2009.14611).
60. Que, Z. *et al.* Accelerating Recurrent Neural Networks for Gravitational Wave Experiments. In *32nd IEEE International Conference on Application-specific Systems, Architectures and Processors*, DOI: [10.1109/ASAP52443.2021.00025](https://doi.org/10.1109/ASAP52443.2021.00025) (2021). [2106.14089](https://arxiv.org/abs/2106.14089).
61. Aasi, J. *et al.* Advanced LIGO. *Class. Quant. Grav.* **32**, 074001, DOI: [10.1088/0264-9381/32/7/074001](https://doi.org/10.1088/0264-9381/32/7/074001) (2015). [1411.4547](https://arxiv.org/abs/1411.4547).
62. <https://gwosc.org/>.
63. Zink, R. M. & Klicka, L. B. The taxonomic basis of subspecies listed as threatened and endangered under the endangered species act. *Front. conservation science* **3**, 971280, DOI: [10.3389/fcosc.2022.971280](https://doi.org/10.3389/fcosc.2022.971280) (2022).
64. Mendel, G. Verhandlungen des naturforschenden vereines in brünn. *Abhandlungen* **4**, 3–47 (1865).
65. Hines, H. M. *et al.* Wing patterning gene redefines the mimetic history of heliconius butterflies. *Proc. Natl. Acad. Sci.* **108**, 19666–19671, DOI: [10.1073/pnas.1110096108](https://doi.org/10.1073/pnas.1110096108) (2011).

- 66.** Mallet, J., Beltrán, M., Neukirchen, W. & Linares, M. Natural hybridization in heliconiine butterflies: the species boundary as a continuum. *BMC Evol. Biol.* **7**, DOI: [10.1186/1471-2148-7-28](https://doi.org/10.1186/1471-2148-7-28) (2007).
- 67.** Lawrence, C. & Campolongo, E. G. Heliconius collection (cambridge butterfly), DOI: [10.57967/hf/2668](https://doi.org/10.57967/hf/2668) (2024).
- 68.** Jiggins, C. & Warren, I. Cambridge butterfly wing collection - Chris Jiggins 2001/2 broods batch 1, DOI: [10.5281/zenodo.2549524](https://doi.org/10.5281/zenodo.2549524) (2019).
- 69.** Jiggins, C. & Warren, I. Cambridge butterfly wing collection - Chris Jiggins 2001/2 broods batch 2, DOI: [10.5281/zenodo.2550097](https://doi.org/10.5281/zenodo.2550097) (2019).
- 70.** Jiggins, C., Montejo-Kovacevich, G., Warren, I. & Wiltshire, E. Cambridge butterfly wing collection batch 3, DOI: [10.5281/zenodo.2682458](https://doi.org/10.5281/zenodo.2682458) (2019).
- 71.** Mattila, A., Jiggins, C. & Warren, I. University of Helsinki butterfly collection - Anniina Mattila bred specimens, DOI: [10.5281/zenodo.2555086](https://doi.org/10.5281/zenodo.2555086) (2019).
- 72.** Meier, J. I., Salazar, P., Montejo-Kovacevich, G., Warren, I. & Jiggins, C. Cambridge butterfly wing collection - Patricio Salazar PhD wild specimens batch 3, DOI: [10.5281/zenodo.4153502](https://doi.org/10.5281/zenodo.4153502) (2020).
- 73.** Montejo-Kovacevich, G., Jiggins, C. & Warren, I. Cambridge butterfly wing collection batch 2, DOI: [10.5281/zenodo.2677821](https://doi.org/10.5281/zenodo.2677821) (2019).
- 74.** Montejo-Kovacevich, G., Jiggins, C., Warren, I. & Wiltshire, E. Cambridge butterfly wing collection batch 5, DOI: [10.5281/zenodo.2684906](https://doi.org/10.5281/zenodo.2684906) (2019).
- 75.** Montejo-Kovacevich, G., Jiggins, C., Warren, I. & Wiltshire, E. Cambridge butterfly wing collection batch 6, DOI: [10.5281/zenodo.2686762](https://doi.org/10.5281/zenodo.2686762) (2019).
- 76.** Montejo-Kovacevich, G., Jiggins, C., Warren, I. & Wiltshire, E. Cambridge butterfly wing collection batch 7, DOI: [10.5281/zenodo.2702457](https://doi.org/10.5281/zenodo.2702457) (2019).
- 77.** Montejo-Kovacevich, G., Jiggins, C., Warren, I. & Wiltshire, E. Cambridge butterfly wing collection batch 8, DOI: [10.5281/zenodo.2707828](https://doi.org/10.5281/zenodo.2707828) (2019).
- 78.** Montejo-Kovacevich, G., Jiggins, C., Warren, I., Wiltshire, E. & Gavins, I. Cambridge butterfly wing collection batch 9, DOI: [10.5281/zenodo.2714333](https://doi.org/10.5281/zenodo.2714333) (2019).
- 79.** Montejo-Kovacevich, G. *et al.* Cambridge and collaborators butterfly wing collection batch 10, DOI: [10.5281/zenodo.2813153](https://doi.org/10.5281/zenodo.2813153) (2019).
- 80.** Montejo-Kovacevich, G., Jiggins, C. & Warren, I. Cambridge butterfly wing collection batch 1- version 2, DOI: [10.5281/zenodo.3082688](https://doi.org/10.5281/zenodo.3082688) (2019).
- 81.** Montejo-Kovacevich, G., Paynter, Q. & Ghane, A. Heliconius erato cyrbia, Cook Islands (New Zealand) 2016, 2019, 2021, DOI: [10.5281/zenodo.5526257](https://doi.org/10.5281/zenodo.5526257) (2021).
- 82.** Montejo-Kovacevich, G., van der Heijden, E., Nadeau, N. & Jiggins, C. Cambridge butterfly wing collection batch 10, DOI: [10.5281/zenodo.4289223](https://doi.org/10.5281/zenodo.4289223) (2020).
- 83.** Montejo-Kovacevich, G., van der Heijden, E. & Jiggins, C. Cambridge butterfly collection - GMK Broods Ikiam 2018, DOI: [10.5281/zenodo.4291095](https://doi.org/10.5281/zenodo.4291095) (2020).
- 84.** Salazar, P. A., Nadeau, N., Montejo-Kovacevich, G. & Jiggins, C. Sheffield butterfly wing collection - Patricio Salazar, Nicola Nadeau, Ikiam broods batch 1 and 2, DOI: [10.5281/zenodo.4288311](https://doi.org/10.5281/zenodo.4288311) (2020).
- 85.** Pinheiro de Castro, E. *et al.* Brazilian Butterflies Collected December 2020 to January 2021, DOI: [10.5281/zenodo.5561246](https://doi.org/10.5281/zenodo.5561246) (2022).
- 86.** Salazar, P., Montejo-Kovacevich, G., Warren, I. & Jiggins, C. Cambridge butterfly wing collection - Patricio Salazar PhD wild and bred specimens batch 1, DOI: [10.5281/zenodo.1748277](https://doi.org/10.5281/zenodo.1748277) (2018).
- 87.** Salazar, P., Montejo-Kovacevich, G., Warren, I. & Jiggins, C. Cambridge butterfly wing collection - Patricio Salazar PhD wild and bred specimens batch 2, DOI: [10.5281/zenodo.2548678](https://doi.org/10.5281/zenodo.2548678) (2019).
- 88.** Salazar, C., Montejo-Kovacevich, G., Jiggins, C., Warren, I. & Gavins, I. Camilo Salazar and Cambridge butterfly wing collection batch 1, DOI: [10.5281/zenodo.2735056](https://doi.org/10.5281/zenodo.2735056) (2019).
- 89.** Warren, I. & Jiggins, C. Miscellaneous Heliconius wing photographs (2001-2019) Part 1, DOI: [10.5281/zenodo.2552371](https://doi.org/10.5281/zenodo.2552371) (2019).

90. Warren, I. & Jiggins, C. Miscellaneous Heliconius wing photographs (2001-2019) Part 2, DOI: [10.5281/zenodo.2553501](https://doi.org/10.5281/zenodo.2553501) (2019).
91. Warren, I. & Jiggins, C. Miscellaneous Heliconius wing photographs (2001-2019) Part 3, DOI: [10.5281/zenodo.2553977](https://doi.org/10.5281/zenodo.2553977) (2019).
92. <https://www.ndbc.noaa.gov/>.
93. Ward, P. J. *et al.* Strong influence of el niño southern oscillation on flood risk around the world. *Proc. Natl. Acad. Sci.* **111**, 15659–15664 (2014).
94. Ghosh, S. *et al.* Towards kriging-informed conditional diffusion for regional sea-level data downscaling: A summary of results. In *Proceedings of the 32nd ACM International Conference on Advances in Geographic Information Systems*, 372–383 (2024).
95. Tapia-Rivas, N. I., Estevez, P. A. & Cortes-Briones, J. A. A robust deep learning detector for sleep spindles and k-complexes: towards population norms. *Sci. Reports* **14** (2024).
96. Carloni, G., Berti, A., Iacconi, C., Pascali, M. A. & Colantonio, S. On the applicability of prototypical part learning in medical images: Breast masses classification using protopnet. In *International Conference on Pattern Recognition*, 539–557 (Springer, 2022).
97. Chen, C. *et al.* This looks like that: deep learning for interpretable image recognition. *Adv. neural information processing systems* **32** (2019).
98. Yang, J. *et al.* The role of satellite remote sensing in climate change studies. *Nat. climate change* **3**, 875–883 (2013).
99. Yang, M. *et al.* Climate smart computing: A perspective. *Pervasive Mob. Comput.* **108**, 102019, DOI: <https://doi.org/10.1016/j.pmcj.2025.102019> (2025).