MACHINE LEARNING
Science and Technology

CHALLENGES

Learning to detect continuous gravitational waves: an open data-analysis competition

R Tenorio[1,2,3,*], M J Williams[4,5], J Bayley[5], C Messenger[5], M Demkin[6], W Reade[6], J Koda[7], Y Yamakawa[7], T Yamaguchi[7], K Abe[7], C Achard[7], H S T Bukhari[8], M V Shugaev[9], G Sokolov[7], V Debout[12], S Goulet[12], J-L Tastet[13,14], I Timiryasov[15], O Ruchayskiy[15], D Kanonik[7], S Seferbekov[7], S Saito[7], R Sato[7], S Segawa[7], A Zhyvalkouski[7], Y Uchida[7], S Yokoi[7], A Sayed[7], R-Q Xing[7], I Yamashita[7] and Z Wang[7]

1 Dipartimento di Fisica 'G. Occhialini', Università degli Studi di Milano-Bicocca, Piazza della Scienza 3, 20126 Milano, Italy
2 INFN, Sezione di Milano-Bicocca, Piazza della Scienza 3, 20126 Milano, Italy
3 Departament de Física, Universitat de les Illes Balears, IAC3-IEEC, Ctra. Valldemossa km 7.5, E-07122 Palma, Spain
4 University of Portsmouth, Portsmouth PO1 3FX, United Kingdom
5 SUPA, School of Physics and Astronomy, University of Glasgow, Glasgow G12 8QQ, United Kingdom
6 Kaggle/Google, Mountain View, CA, United States of America
7 Independent researcher
8 Vergesense, Mountain View 94041, CA, United States of America
9 Department of Materials Science and Engineering, University of Virginia, Charlottesville, VA 22904-4745, United States of America
10 Aillis Inc., Tokyo, Japan
11 Department of Health Policy and Public Health, Graduate School of Pharmaceutical Sciences, The University of Tokyo, Tokyo, Japan
12 CS Group, Tolouse 31500, France
13 Departamento de Física Teórica and Instituto de Física Teórica UAM/CSIC, Universidad Autónoma de Madrid, Cantoblanco 28049, Madrid, Spain
14 Department of Computer Science, University of Copenhagen, Universitetsparken 1, DK-2100 Copenhagen, Denmark
15 Niels Bohr Institute, University of Copenhagen, Jagtvej 155A, DK-2200 Copenhagen, Denmark
* Author to whom any correspondence should be addressed.

E-mail: rodrigo.tenorio@unimib.it

Keywords: gravitational waves, continuous gravitational waves, gravitational-wave data analysis, machine-learning applications

OPEN ACCESS

RECEIVED
25 September 2025

REVISED
9 December 2025

ACCEPTED FOR PUBLICATION
16 December 2025

PUBLISHED
30 December 2025

Abstract

We report results of a public data-analysis challenge, hosted on the open data-science platform Kaggle, to detect simulated continuous gravitational-wave signals (CWs). These are weak signals from rapidly spinning neutron stars that remain undetected despite extensive searches. The competition dataset consisted of a population of CW signals using both simulated and real LIGO detector data matching the conditions of actual CW searches. The competition attracted more than 1000 participants to develop realistic CW search algorithms. We describe the top 10 approaches and discuss their applicability as a pre-processing step compared to standard CW-search approaches. For the competition's dataset, we find that top approaches can reduce the computing cost by 1 to 3 orders of magnitude at a false-dismissal probability comparable to standard CW searches. Additionally, the competition drove the development of new GPU-accelerated detection pipelines, which facilitated their adoption in other areas of gravitational-wave data analysis. We release the associated dataset, which constitutes the first open standardized benchmark for CW detection, to enable reproducible method comparisons and to encourage further developments toward the first detection of these elusive signals.

1. Introduction

Continuous gravitational waves (CWs) are long-duration gravitational-wave (GW) signals expected from rapidly-spinning non-axisymmetric neutron stars (NSs) [1]. Although currently undetected, their detection would shed light on the extreme physics of NSs [2], allow us to understand their population in our galaxy [3, 4], or even establish a new probe to direct and indirect detection of dark matter [5].

CW searches can be classified according to their assumptions on the expected source [6]. Targeted searches, for example, assume the CW emission of a pulsar is phase-locked to its electromagnetic emission; this allows for very sensitive searches using matched filtering. Blind searches, on the other hand, attempt to detect a CW signal from unknown sources; due to the sheer size of their parameter-space, these kinds of searches cannot make use of matched filtering.

A standard CW search operates by filtering data against a set of signal models, usually referred to as *templates*, which are parameterized by their amplitude and phase-evolution parameters (see section 3.3). The result of filtering data against a specific signal model is a *detection statistic*, which is then used to establish the significance of a candidate. The specific detection statistic in use depends on search assumptions and available computing budget, and has a direct impact on sensitivity [7].

Fully-coherent searches typically use the $\mathcal{F}$-statistic [8, 9], which tracks the phase of a CW signal accounting for possible amplitude and frequency modulations due to the orbital motion of GW detectors. This approach is unfeasible for broad parameter spaces as the number of required signal templates quickly becomes computationally prohibitive [10]. Semicoherent detection statistics, on the other hand, follow the frequency evolution of a signal and only track its phase evolution along short segments. Effectively, a semiocherent statistic accumulates coherent statistics (such as segment-wise $\mathcal{F}$-statistics, power spectra [11, 12], or the binarized power [13, 14]) following the signal's frequency evolution. These statistics impose milder constraints and require less templates to cover a given parameter-space region, reducing the computing cost of a search, but also reduce the sensitivity as background noise is more likely to fit a signal template [15]. Because of this, semicoherent searches are robust to unmodeled physics such as NS glitches [16] or spin-wandering [17, 18].

The current trend in blind CW searches is to explore relatively narrow parameter-space regions to aim for astrophysically plausible CW sources [19–22], or to extend the parameter space towards binary systems [23–25], which may emit detectable CW signals more easily due to accretion [26–28]. The most sensitive of these searches are based on highly-efficient implementations of the semicoherent $\mathcal{F}$-statistic, such as the global correlation transform (`GCT`) [29], `Weave` [30], or `BinarySkyHou`$\mathcal{F}$ [31]. These searches require significant computing budgets to be successfully deployed. Alternatively, short-coherence searches followed by hierarchical follow-ups [32–35], or signal-agnostic searches such as `SOAP` [11], require a much more modest budget, albeit with a corresponding reduction in sensitivity. All in all, the sensitivity of a CW search is primarily limited by the available computing resources [15].

Machine learning applications [36] have demonstrated significant accelerations in search [37–40] and parameter-estimation [41–44] workflows targeting compact binary coalescences compared to standard approaches [45–47]. Due to the comparatively broader parameter space, applications to blind CW signals have reported limited success [48–51], with most successful applications focusing on accelerating specific computational steps [52] or increasing the robustness of a search to unmodeled physics [53, 54].

In this work, we present the results of an open data-analysis competition to spark new approaches and algorithmic developments in CW searches [55].

We constructed a dataset using the sensitivity of flagship searches such as `GCT` or `Weave` as a baseline reference to maximize the impact of winning solutions, and compared their performance to an agnostic CW search (`SOAP`) to gauge the impact of design choices made by different teams. This competition delivered a comparison of methods beyond what within-domain mock data challenges [56–59] are able to prospect and allowed us to understand the complications and computational advantages achievable by off-the-shelf approaches from beyond the community. We release the dataset of the competition [60] to provide the first publicly-available benchmark targeted towards the detection of CW searches. Tools and tutorials to interact with this data are publicly available in Kaggle [61] and the `PyFstat` repository [62].

The paper is structured as follows: In section 2 we introduce the rationale behind this CW-search data-analysis competition. In section 3 we introduce the competition's dataset. In section 4 we discuss the top-scoring solutions of the competition. In section 5 we interpret these solutions in the context of standard CW searches. We finish by discussing these results in section 6. We also summarize the results of the first Kaggle competition aimed at detecting GW signals from compact binary coalescences [63] in appendix.

2. Leveraging open competitions for GW research

The development of CW search methods is often constrained by the expertise and computational perspective available within the CW research community. The complexity of detecting a signal in noisy data, however, represents a challenge that may benefit, if properly supervised and tested, from a fresh perspective as provided by the broader data science community.

Open data-analysis competitions have demonstrated significant value in advancing scientific research across multiple domains, such as medical imaging [64–66], time-series forecasting [67], or particle physics [68], often producing solutions that outperform existing standards. These successes stem from competitions' ability to attract diverse problem-solving approaches, encourage rapid iteration, and provide standardized benchmarks for method comparison [69].

The GW community has already embraced crowd-sourcing approaches through platforms, such as Zooniverse [70], to understand transient noise in interferometric detectors [71–73], or Kaggle [74] to search for GW signals from compact binary coalescences [63]. This approach can be understood as an evolution of the well-established mock data challenges conducted by the GW community [56–59], where the aim has shifted from comparing existing pipelines to exploring new strategies altogether. The overall experience in GW citizen-science so far suggests non-experts, when given a proper training framework, can deliver solutions beneficial to the GW community [75, 76]. For example, the results of the first Kaggle competition [63] served as a basis to push forward our understanding of GW searches using machine learning [40].

The problem of detecting CW signals presents an ideal case for an open competition: First, the problem can be framed as a well-defined classification task [77, 78], which provides a simple criterion to rank different strategies against each other. Second, synthetic data can be easily generated, either from first principles or through well-established software packages [79–81]. Third, since large-scale searches can be understood as the aggregation of many small-scale searches [12], competitors can tackle a simulation of the real problem using standard hardware. Finally, despite the existing experience in the field, large portions of the search-pipeline configuration space where novel improvements may be found remain yet to be explored.

Building on previous positive results [63], we released our competition on Kaggle [74], a well-established data-analysis platform powered by Google with proven track record in scientific applications and more than a million active users spread across vast domains of data science and engineering.

3. Dataset description

We seek a data-analysis strategy compatible with a blind CW search; that is, a pipeline which classifies narrow frequency bands of LIGO/Virgo/KAGRA [82–85] data according to whether they contain a signal or not. We take the sensitivity of deep searches conducted using `GCT` [86] and `Weave` [87] as a reference for our competition. These pipelines use highly-efficient implementations of the semicoherent $\mathcal{F}$-statistic [29, 30, 88] to deliver the most sensitive all-sky search results across their explored parameter space. Their sensitivity is effectively bounded by the available computing resources, which amount to about 100 million CPU hours or 1 million GPU hours, depending on the specific implementation. By framing CW detection as a binary classification problem, we aim to discover whether alternative approaches could compete with or complement these well-established strategies. Note that the high computing cost of these searches complicates a direct comparison of methods; instead, we use the sensitivity of these pipelines to construct a signal distribution for the challenge, and after the competition discuss the advantages of the methods presented by the competitors.

3.1. Data format

We constructed a dataset using short Fourier transforms [89] (SFTs), a data format used by actual blind CW searches. This choice provides the participants with the same information available to CW data analysts and simplifies the transition of a successful solution into a production stage.

Given a set of N time-domain data samples $x_\alpha[j], j = 0, \ldots, N-1$ starting at an epoch t_α sampled with a period Δt, we define an SFT following [89]:

$$\tilde{x}_\alpha[k] = \Delta t \sum_{j=0}^{N-1} x_\alpha[j]\, \mathrm{e}^{-\frac{-2i\pi jk}{N}}. \tag{1}$$

Here the index $k = 0, \ldots, N/2$ labels the positive frequencies $f_k = k/T_{\mathrm{SFT}}$, where $T_{\mathrm{SFT}} = N\Delta t$ is the duration of an SFT. These frequencies physically correspond to $F_0 + f_k$, where the fundamental frequency f_0 is always extracted through a homodyne filter before computing the SFT.

The competition dataset consisted of 8000 test samples divided into signal and noise samples in equal parts. Additionally, 600 training samples were provided to the competitors as a quick entrypoint to the competition. The dataset was made available on the Kaggle platform, and adds up to $\sim$0.5 TB. Additionally, we supplied the tools to generate simulated CW data through the `PyFstat` package [81], which provides a Python interface to the LIGO/Virgo/KAGRA Algorithm Library [79, 80].

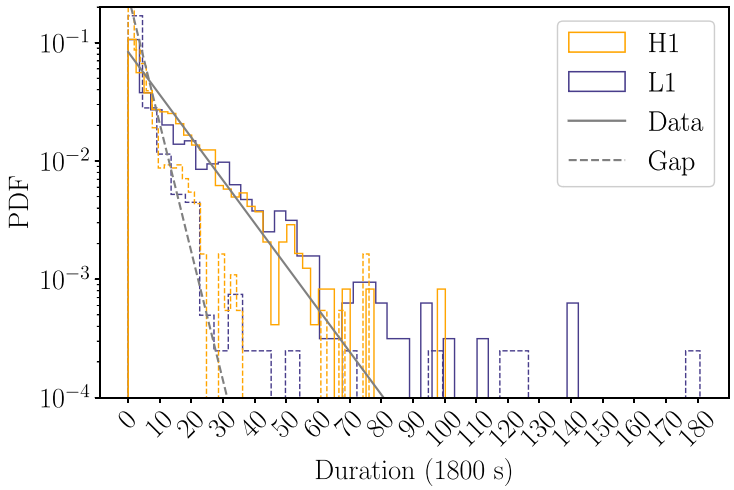

Figure 1. Distribution of observing-mode (solid line) and gap (dashed line) duration throughout the O3 observing run. Different colors refer to different detectors. Gray lines are an exponential fit as discussed in the text.

The dataset contains both noise-only samples and software-simulated CW signals, using both simulated Gaussian noise and data slices from the third observing run of the Advanced LIGO detectors [90–92]. The chosen frequency band (0.2 Hz or 0.2 Hz $\times T_{\mathrm{SFT}} = 360$ frequency bins) is compatible with those analyzed by current blind searches. The sample duration, on the other hand, falls on the shorter end, as most searches analyze at least 6 months of data. This choice is due to the limited storage available on the Kaggle platform. Regardless, the duration is long enough to contain the main distinctive features of CW signals, namely a Doppler shift on the time-scale of months combined with a daily amplitude modulation.

3.2. Noise distribution

All samples of the competition contain background noise. This was either simulated Gaussian noise or real Advanced LIGO noise from the third LVK observing run (O3) [90–92].

Each data sample in the competition's dataset consists of a two sets of SFTs, representing the H1 (Hanford) and L1 (Livingston) LIGO detectors, spanning a total time of $T_{\mathrm{span}} = 4$ month and a frequency band of 0.2 Hz with $T_{\mathrm{SFT}} = 1800$ s. To better simulate the conditions of a real search, we introduce gaps in the data. As a result, SFTs are not necessarily contiguous (i. e. $t_{\alpha+1} - t_\alpha \geq T_{\mathrm{SFT}}$).

We model the duration of observing-quality ('science-mode') segments and gaps during the O3 run using an exponential distribution, as shown in figure 1. We compute the duration of science-mode segments and gaps by generating 1800 s-SFTs from publicly-available O3 data using the timestamps provided in [93]. The average science-mode segment (gap) duration is compatible with $\tau_{\mathrm{data}} = 2.16 \times 10^4$ s ($\tau_{\mathrm{gap}} = 7.2 \times 10^3$ s), which corresponds to $\sim$12 ($\sim$4) SFTs.

For each interferometer, we draw gap/science-mode durations alternatively from the distributions in figure 1 until completing a span of T_{span}. The first segment is sampled from the gap distribution and added to the starting time of O3 to avoid fixing the starting time. As a result, the SFTs of different detectors have, in general, different starting timestamps. By construction, the resulting data samples for both detectors have a 'duty cycle' (fraction of science-mode data) of

$$\mathrm{dc} = \frac{\tau_{\mathrm{data}}}{\tau_{\mathrm{data}} + \tau_{\mathrm{gap}}} \simeq 0.78\,, \tag{2}$$

which is comparable with realistic observing runs [94, 95].

Gaussian-noise data sample are generate by drawing from a Gaussian distribution with zero mean and (single-sided) amplitude spectral density (ASD) $\sqrt{S_n} = 5 \times 10^{-24} Hz^{\frac{-11}{2}}$. Equivalently, the real and imaginary parts of each complex SFT bin $\tilde{x}_\alpha[k]$ are drawn from a Gaussian distribution with zero mean and standard deviation $\frac{1}{2}\sqrt{T_{\mathrm{SFT}} S_{\mathrm{n}}}$. This setup is comparable with mock analyzes conducted in current CW searches (see e.g. [12] and references therein).

To generate a data sample containing real noise, we instead sample SFTs from LVK O3 data. To prevent participants from identifying noise realizations using publicly available data, we shift the frequency labels by a random amount within ± 4 Hz and add Gaussian noise with 0.1% of the data's ASD to mangle the numerical noise values but maintain real-noise features. This value was chosen empirically.

Table 1. Distribution of signal parameters used to generate the competition's challenge. The Uniform distributions are defined by their lower and upper limits; the gamma distribution (Γ) is expressed using the shape-scale parameterization.

Parameter	Unit	Symbol	Distribution
Frequency	Hz	f_0	$\text{Uniform}(50, 500)$
Spindown	Hz^2	f_1	$-10^{-\text{Uniform}(8,12)}$ (90%) $+10^{-\text{Uniform}(9,12)}$ (10%)
Right Ascension	rad	α	$\text{Uniform}(0, 2\pi)$
(sin) Declination	—	$\sin\delta$	$\text{Uniform}(-1, 1)$
(cos) Inclination angle	—	$\cos\iota$	$\text{Uniform}(-1, 1)$
Polarization angle	rad	ψ	$\text{Uniform}(-\pi/4, \pi/4)$
Initial phase	rad	ϕ_0	$\text{Uniform}(0, 2\pi)$
SNR	—	ρ	$\Gamma(5, 9)$

3.3. Signal distribution

CW signals from isolated NSs are typically described in terms of the amplitude parameters $\mathcal{A}$, comprising the GW's nominal amplitude h_0, the (cosine of the) inclination angle with respect to the line of sight $\cos\iota$, the polarisation angle ψ, and the initial phase ϕ_0; and the phase-evolution parameters λ, which include the signal's frequency f_0 and frequency derivative (spindown) f_1 at a fiducial reference time t_{ref}, and the sky location $\hat{n}$. We take the reference time at the start of the O3 run and parameterize the sky position using equatorial coordinates (right ascension α and declination δ).

We simulate a source population compatible with that assumed in blind searches, as summarized in table 1. To generate a signal sample, we draw a set of parameters $(\mathcal{A}, \lambda)$ from the population, we simulate noise as described in section 3.2 for both the H1 and L1 LIGO detectors, and use `MakeFakeData_v5` [79] to add a simulated signal into the noise. This software simulates a CW signal as observed by a ground-based interferometric detector, adds Gaussian noise with a given ASD, and returns a the corresponding SFTs with the specified T_{SFT}.

The sky position $\hat{n}$ is drawn uniformly across the celestial sphere; the distribution over the initial phase is uniform along $[0, 2\pi)$; orientation angles $(\cos\iota, \psi)$ are drawn following an isotropically-oriented population. The spindown f_1 is drawn from a log-uniform distribution along $[-10^{-8}, -10^{-12}]\,\text{Hz}\,\text{s}^{-1}$ with 90% probability and $[10^{-12}, 10^{-9}]\,\text{Hz}\,\text{s}^{-1}$ with a 10% probability. This split accounts for possible spin-ups due to accretion, proper motion, or the evaporation of a boson cloud [8, 17, 96]

The frequency f_0 is sampled uniformly along $[50, 500]\,\text{Hz}$, which is a frequency range commonly explored by blind CW searches as it contains the most sensitive band of the Advanced LIGO detectors [94]. The bandwidth of a CW signal depends on its frequency and sky location, as it is largely dominated by the Doppler effect due to the orbital motion of the detector around the Sun. For the parameter space considered in this challenge, the maximum bandwidth is about 0.05 Hz. To complete a band of 0.2 Hz, we attach noise data to both sides of the signal's spectrum so it is contained completely within the data sample. Also, to avoid biasing the dataset, the proportion of data attached to each side of the signal's frequency band is chosen by drawing a uniform random number between 0 and 1 so that the signal is not systematically centered.

We describe the amplitude of a CW signal using the signal-to-noise ratio (SNR)

$$\rho(h_0, \psi, \cos\iota, \hat{n}) = \sqrt{\frac{T_{\text{data}}}{S_{\text{n}}}}\, g(\psi, \cos\iota, \hat{n})\, h_0 \tag{3}$$

where $g(\psi, \cos\iota, \hat{n})$ is a known function [97] encoding the dependency on the orientation angles and sky position, and $T_{\text{data}} = T_{\text{SFT}} N_{\text{SFT}}$, where N_{SFT} refers to the *total* number of SFTs (counting each detector separately). The SNR is directly related to the detectability of a signal [8], and is proportional to h_0 given $(\psi, \cos\iota, \hat{n})$ and a dataset. This implies the SNR distribution will determine the difficulty of the challenge, as high SNR signals ($\rho > 80$ in our specific configuration) tend to display recognizable features in the data, making them easy to detect.

Blind CW searches usually quote their results in terms of population-based upper limits at a certain false-dismissal probability (usually 5% – 10%) on a constant-h_0 population [6, 7]. The false-alarm probability of a search is usually not computed. This complicates the comparison of Kaggle solutions to realistic searches on equal footing. Instead, we will base our comparisons on computing cost; concretely, we will assess the capability of Kaggle solution to discard broad parameter-space regions in such a way that signals detectable by a blind CW search are not discarded. To do so, it will suffice to re-scale the sensitivity estimates reported by the two reference searches to the dataset duration used in the competition.

Table 2. Summary of sensitivity estimates reported by the two searches considered in this work. These are computed by averaging the reported sensitivity estimates across the frequency band of the search.

	$T_{\rm coh}$ (h)	$\langle h_0 \rangle$	$\sqrt{\langle \rho^2 \rangle_{\vec{n},\cos\iota,\psi}}$	$\sqrt{\langle \rho^2 \rangle^{\rm K}_{\vec{n},\cos\iota,\psi}}$
`GCT` [86]	60	1.3×10^{-25}	47	61
`Weave` [87]	240	1.0×10^{-25}	36	47

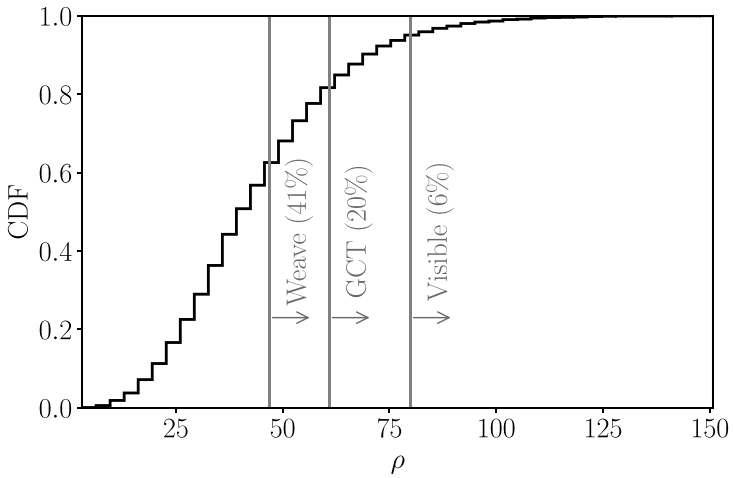

Figure 2. Empirical cumulative distribution of SNR $\rho \sim \Gamma(5, 9)$. Vertical lines denote the re-scaled sensitivity of different search pipelines as discussed in the main text.

The two reference searches we use in this challenge, based on the `GCT` [86] and `Weave` [87] pipelines, report their sensitivity estimates in terms of the average amplitude $\langle h_0 \rangle$ at which 90% of an isotropically-oriented uniformly-sky-distributed signals would be detected. We relate this quantity to the corresponding averaged SNR as [98]

$$\sqrt{\langle \rho^2 \rangle_{\vec{n},\cos\iota,\psi}} = \frac{2}{5}\sqrt{\frac{T_{\rm data}}{S_{\rm n}}}\langle h_0 \rangle \,. \tag{4}$$

These pipelines split the dataset into segments with a duration $T_{\rm coh}$ which are analyzed independently using the $\mathcal{F}$-statistic, and then the results are combined to compute the semicoherent $\mathcal{F}$-statistic.

We re-scale the sensitivity estimates using $T_{\rm data}$, which corresponds to increasing $T_{\rm coh}$ so that the number of coherent segments is unchanged. The original searches used O2 Advanced LIGO data, with a nominal duration of 9 months and about 60%–65% duty cycle [99]: $T^{\rm O2}_{\rm data} \approx 2.6 \times 10^7$ s. Our challenge covers a nominal duration of 4 months and a duty cycle of 78%: $T^{\rm K}_{\rm data} \approx 1.6 \times 10^7$ s. The sensitivity of these searches in the competition's dataset corresponds to scaling their reported sensitivity up by $\sqrt{T^{\rm O2}_{\rm data}/T^{\rm K}_{\rm data}} \approx 1.3$, as shown in table 2.

With this information, we choose to describe the SNR distribution using a Gamma distribution with shape parameter $k = 5$ and scale parameter $\theta = 9$, shown in figure 2. This distribution is such that 20% of the signals are above the sensitivity of `GCT` and about 41% of the signals are above the sensitivity of `Weave`. 6% of the signals display visible features ($\rho > 80$). We show an example CW signal at different SNR levels in figure 3.

4. Competition results

The Kaggle competition was held from October 4th 2022 to January 4th 2023, and accumulated a total number of ∼30 000 submissions amongst 936 teams. Competitors were provided with a dataset as described in section 3, and the task was to submit a ranking statistic for each sample in order to classify them as signal or noise.

Each submission was ranked according to the area under the receiver operating characteristic (ROC) curve (AUC score). That is, the ranking statistics of a submission where used as thresholds t to compute the false-dismissal probability (fraction of signal samples classified as noise, $p_{\rm FD}(t)$) at a given level of

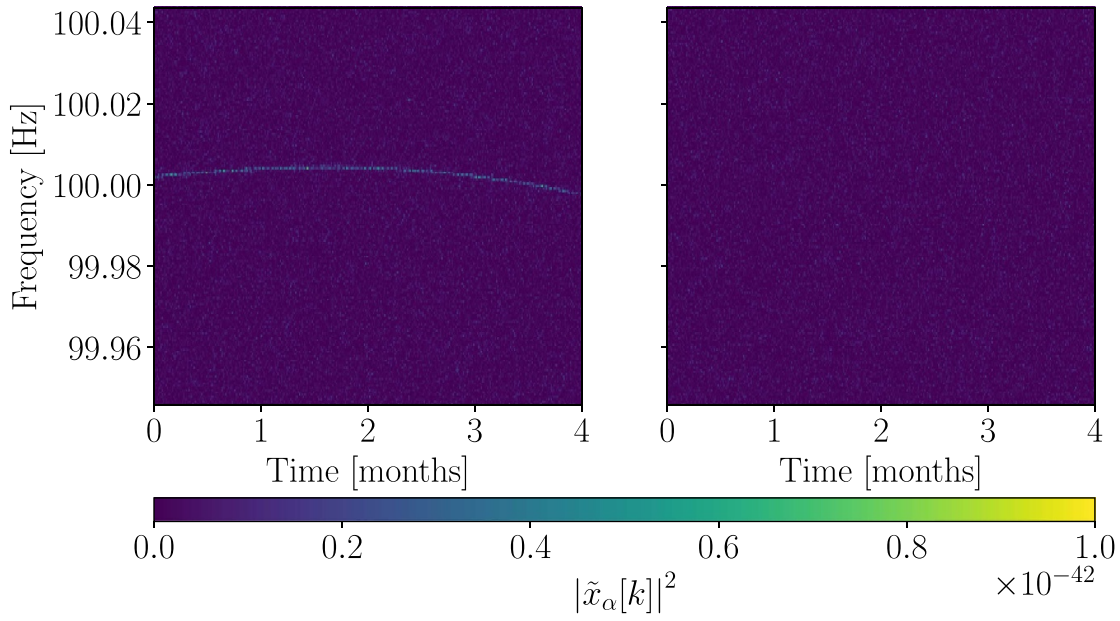

Figure 3. Spectrograms of a sample containing 4 months of simulated Gaussian noise with a fiducial CW signal as observed by the H1 detector. The CW amplitude in the left (right) spectrogram is such that $\rho = 350$ ($\rho = 35$); the signal parameters are otherwise identical. As a result, the signal is visible in the left spectrogram but not in the right one. This data was generated without gaps for the sake of visualization.

false-positive probability (fraction of noise samples classified as signals, $p_{\mathrm{FP}}(t)$). The AUC was then computed by integrating the area under the curve described by $(p_{\mathrm{FP}}(t), p_{\mathrm{FD}}(t))$. This ranking metric was the most suitable for our data-analysis problem given the available selection in the Kaggle platform.

The challenge actually makes use of two leaderboards, namely *public* leaderboard, which is visible throughout the duration of the competition, using a small fraction of the competition's dataset to compute the AUC score, and the *private* leaderboard, which computes the AUC score using the complete dataset. This design allows participants to obtain a continuous assessment of their performance against other competitors, which is beneficial for a rapid iteration of the solution's design, and minimizes the chances of winning solutions ovefitting to a specific dataset. Additionally, the use of a random subset of the dataset minimizes the probability of competitors successfully inferring the true classification of test data samples through repeated submission [100–102].

We show in figure 4 the evolution of the leaderboard scores for the top 10 teams of the competition. Overall, one can distinguish two periods in the competition: From the start of the competition to about 10 days before the end the leaderboard was dominated by Space Coders. Ten days before the end, three more teams (`JunKoda`, `PreferredWave`, and `BearWaves`) managed to surpass their score. Overall, most teams converged rapidly into a sensible solution whose sensitivity remained broadly unchanged throughout the challenge.

4.1. Top-scoring submissions

We divide the solutions proposed by the top-10 scoring teams into three broad classes: Track-Statistic, Track-Fitting, and Machine-Learning.

4.1.1. Track-Statistic

The first class contains solutions implementing a track-based detection statistic similar to those used in blind CW searches with short coherence times. Specifically, these methods assume a closed-form frequency evolution (i.e. a 'frequency track' in the spectrogram) parameterized by a small number of parameters (such as frequency, spindown, and sky position) along which a detection statistic is accumulated. The search then consists in finding the parameters that maximize said detection statistic. This class includes the teams `JunKoda` [103] (1st), `PreferredWave` [104] (2nd), `HiddenNeuralLayers` [105] (5th), and `Shun_PI` [106] (6th). These methods present two main degrees of freedom, namely the detection statistic and strategy to cover the parameter space, which we summarize in table 3. To address non-Gaussianities arising in real-data samples, all teams masked frequency bands in the spectrogram containing large deviations from Gaussian noise.

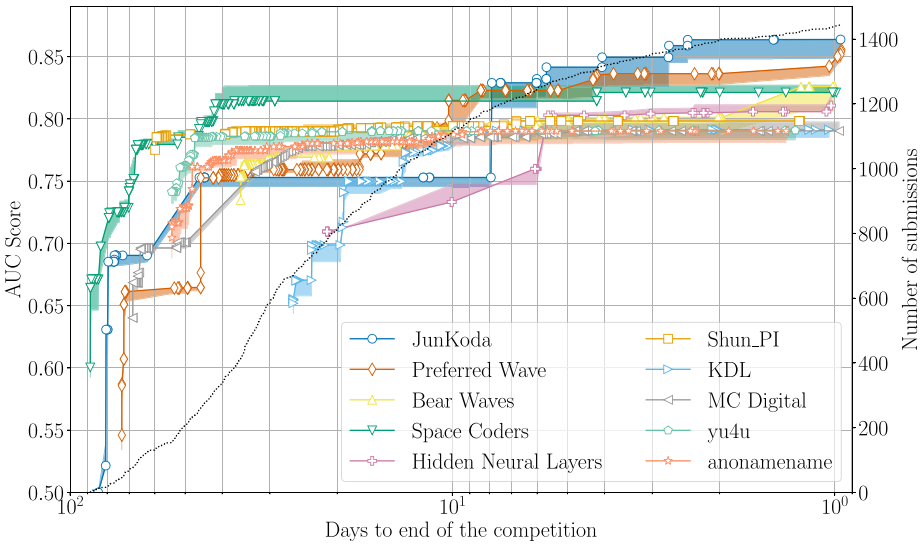

Figure 4. Evolution of the top-10 teams through the competition in terms the AUC score. The markers correspond to the private leaderboard score, while the shaded regions extend to the public leaderboard score. The cumulative number of submissions by these 10 teams is shown as a dotted line.

Table 3. Design choices of the winning solutions presented by the teams in the Track-Statistic class of solutions.

	GPU	Statistic	Template placement
`JunKoda` (1st)	Yes	Maximum Dirichlet-weighed power [107]	Square grid (3.5×10^8 templates)
`PreferredWave` (2nd)	Yes	Average power	Uniform sampling (3.3×10^6 templates)
`HiddenNeuralLayers` (5th)	No	Maximum power	Differential evolution (100 steps) [108]
`Shun_PI` (6th)	No	Maximum power	Simulated-annealing [109].

The relative placement of these four teams in the leaderboard is consistent with their strategies. `HiddenNeuralLayers` and `Shun_PI` place templates in a data driven manner. While such a strategy has been proven effective in local analyzes [16, 33–35], the breadth of the CW parameter-space makes them unsuitable for broad searches such as the ones here considered [16], hence their relatively lower placement. The difference between `JunKoda` and `PreferredWave`, on the other hand, can be ascribed to the significant difference in the number of evaluated templates (10^8 in `JunKoda` versus 10^6 in `PreferredWave`) and the fact that `JunKoda` takes the maximum rather than the average power in the template bank [12].

4.1.2. Track-Finding

The second class introduces an extra degree of freedom with respect to section 4.1.1, namely that the frequency evolution of the signal is not parameterized but inferred from the data. This approach is used by team `SpaceCoders` (4th), and bears some similarities to CW searches based on the Viterbi algorithm [11].

Specifically, the method uses dynamic programming to find a track maximizing the accumulated SFT power. The track is constrained to contain a single inflection point, given the duration of the dataset and the expected signal morphology. Upon identifying a track, the signal's parameters are estimated using a parametric fit in terms of frequency, spindown, and sky position. The optimization is carried out using the Nelder–Mead algorithm [110], in a similar manner to the methods presented in section 4.1.1.

The relative loss of `SpaceCoders` with respect to other track-based methods can be justified as follows: First, the number of tracks explored by the `SpaceCoders` search is exponentially greater than a closed-form track search [6]; this shifts the expected background distribution upwards [111] and, consequently, lowers the detection probability at a fixed amplitude. Second, the track partially relies on the signal displaying strong features within short timescales to clearly identify it. Due to the idiosyncrasy of CWs, this is only true for relatively strong signals.

4.1.3. Machine-Learning

The final class includes solutions using machine-learning strategies; that is, search pipelines in which some component is *trained* before producing a submission. This class comprises teams `BearWaves` [112] (3rd), `KDL`[16] (7th), `MCDigital` [113] (8th), `yu4u` [114] (9th), and `anonamename` [115] (10th).

`anonamename` implemented a classifier based on a convolutional neural network (CNN). Data samples are pre-processed following [115] and two outputs are produced: a detection statistic, and the estimated location of the signal in the spectrogram as a mask. To account for real data, line-like features were generated and included during training.

`yu4u` [114] implemented a similar strategy using a U-net architecture to estimate the location of the signal in the spectrogram and compute a detection statistic. As in the previous case, real data noise features were addressed by generating data containing line-like features during training.

`MC Digital` [113] combines two detection statistics using a CNN and the $\mathcal{F}$-statistic as computed by `PyFstat` [81]. Gaussian and real-data are addressed separately by training two different networks. The combination of CNN and $\mathcal{F}$-statistic results is done manually: samples with high $\mathcal{F}$-statistic values are classified according to the $\mathcal{F}$-statistic; samples with low $\mathcal{F}$-statistic values are classified according to the CNN.

Finally, `BearWaves` [112] presented a more involved approach. First, training data was simulated on-the-fly using `PyFstat` to avoid overfitting. To simplify the task of combining data from two detectors, data samples were time-averaged to obtain a common set of timsetamps. Data samples were analyzed by an ensemble of neural networks applying different processing strategies.

Based on their relative leaderboard positions, image-based machine-learning strategies tend to deliver less sensitive pipelines compared to track-based statistics unless a significant amount of resources is invested in data pre- and post-processing, as shown by `BearWaves` [112].

4.2. Reference submission: `SOAP`

We now describe the application of `SOAP` [11] to the competition's dataset. `SOAP` is a fast, model-agnostic blind CW search capable of detecting a broad class of CW sources, even if they are affected by complicated physics such as spin-wandering, glitches, or timing noise [18]. This robustness, however, limits the sensitivity of `SOAP` to well-modeled isolated NSs due the breadth of its parameter space [6].

Contrary to most blind CW searches, which identify interesting parameter-space templates or regions where a signal may be present [33], `SOAP`'s main search stage assigns detection statistics to frequency *bands* by returning the most significant track constructed via the Viterbi algorithm [116]. This matches the submission format of the Kaggle competition and makes `SOAP` a natural baseline for comparing against Kaggle solutions.

The `SOAP` analysis begins 1800 s SFT power spectra, which are summed over one-day periods (48 SFTs). This daily summation process effectively average out antenna pattern modulation and increasing the SNR in each frequency bin. The search uses the line-aware statistic [11], a multi-detector approach that computes the Bayesian odds ratio between the signal hypothesis and the noise hypothesis including spectral lines [117]. This statistic specifically penalizes instrumental line-like features in spectrograms, helping to distinguish genuine astrophysical signals from detector artifacts. We use the same `SOAP` configuration as that used in [22]. For each frequency band in the test dataset described in section 3, a value of the line-aware statistic is produced for each band.

Note that `SOAP`'s signal model only requires continuity in the frequency-evolution track across consecutive days. No restrictions are set on the coherence of the signal, or the specific functional form of the frequency and amplitude modulation due to the detector's motion. This freedom gives rise to the large parameter space covered by this search. As a result, if a Kaggle solution surpasses the performance of `SOAP`, it suggests an enhanced capability to identify structures in the data that are characteristic of CW signals, likely because the method incorporates more informative signal priors in its design.

4.3. Comparison

The performance of the competition's top-10 solutions is summarized using a ROC curve in Figure 5, which shows the false-dismissal probability p_{FD} of a pipeline (or, equivalently, the detection probability $1 - p_{\mathrm{FD}}$) as a function of false-positive probability p_{FP}. The false-positive probability p_{FP} is defined as the fraction of noise samples that get classified as a signal; the false-dismissal probability p_{FD} corresponds to the fraction of signal samples that get classified as noise.

[16] No available information about this team.

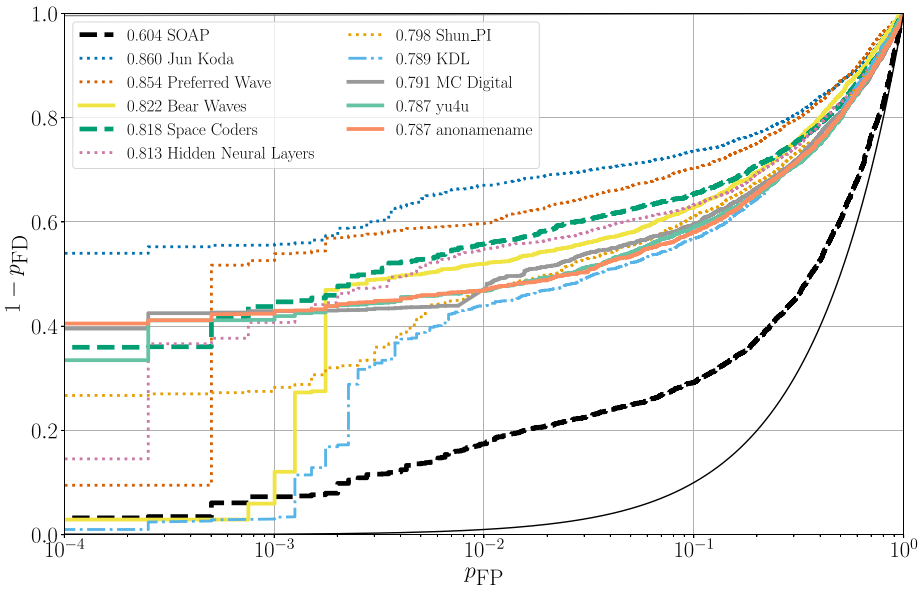

Figure 5. ROC curves corresponding to the top 10 solutions of the competition. The legend lists the AUC scores corresponding to each solution. The thin black line corresponds to a non-discerning classifier with an AUC score of 0.5.

All solutions in the top 10 leaderboard produce an AUC above the value reported by the `SOAP` reference submission in section 4.2. This indicates that all pipelines introduce cogent information into their assumptions when compared to an agnostic search, as previously discussed.

The AUC score, however, aggregates results computed from the complete dataset, which contains signals with different SNRS, across multiple false-positive probabilities. As a result, there is no single solution that dominates across all false-positive probabilities. To more directly compare to standard CW search methods, we present results as a function of SNR in section 5.

5. Comparison to CW searches

Blind CW searches operate by identifying interesting regions in a broad parameter-space to then follow them up with a more sensitive method or using a different dataset. The result of a search is a candidate (or collection of candidates) with a relatively-low associated significance that will be increased through follow-up steps if they are caused by CW signals. The solutions in this competition, due to its design, operate on data samples directly and return an associated false-positive probability p_{FP}; that is, the significance they return is not associated to a specific template. Hence, these two approaches are not directly comparable.

On the other hand, the false-dismissal probability p_{FD}, which is defined as the fraction of signals not detected by a method, is equivalent in both Kaggle solutions and CW searches, as in both cases it refers to the fraction of bands containing a signal that ended up labeled as noise.

This equivalence allows us to interpret Kaggle solutions as one-shot CW searches. Specifically, we consider the situation in which a Kaggle solution is configured to run a CW search at a relatively small p_{FD} so that the Kaggle solution discards a fraction of about $(1 - p_{\mathrm{FP}})$ of the initial dataset, reducing the cost of the CW search, at a negligible sensitivity impact. The result of this first step would then be followed-up using other methods, depending on the intent of the search.

The cost of a standard CW search can be modeled simply as

$$\mathcal{C}_{\mathrm{CW}} = c_{\mathrm{t}} N_{\mathrm{t}} \tag{5}$$

where c_t is the average computing cost per template and N_{t} is the number of templates evaluated by the search. N_{t} itself can be further expressed as

$$N_{\mathrm{t}} = n_{\mathrm{t}} N_{\mathrm{bands}} \tag{6}$$

where n_{t} is the average number of templates per band. The cost of a search including a Kaggle-based pre-processing step, on the other hand, can be expressed as

$$\mathcal{C}_{\mathrm{K+CW}} = c_{\mathrm{K}} N_{\mathrm{bands}} + p_{\mathrm{FP}} \mathcal{C}_{\mathrm{CW}} \tag{7}$$

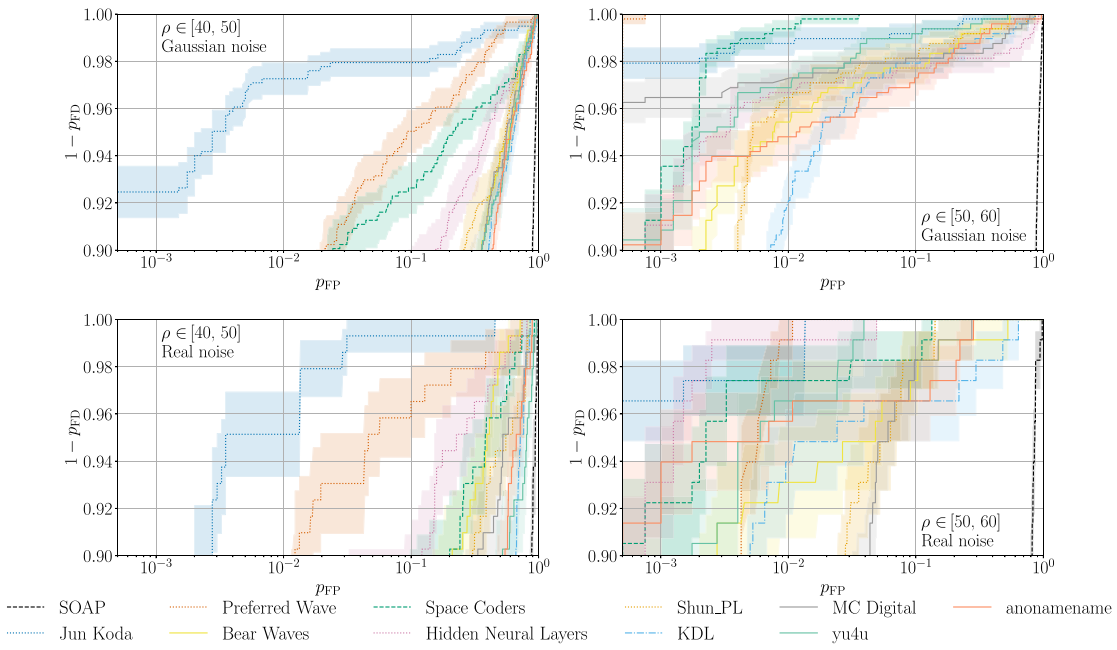

Figure 6. False dismissal probability as a function of false-positive probability for a restricted range of SNR. The upper row shows the results using Gaussian noise samples, while the lower row uses only real-data samples. The shaded regions show binominal errorbars $\delta p_{\mathrm{FD}} = \sqrt{p_{\mathrm{FD}}(1-p_{\mathrm{FD}})/N_{\mathrm{s}}}$, where N_{s} is the number of signals in said SNR range.

where c_{K} is the cost of running the model in a band and p_{FP} represents the fraction of bands flagged as containing a signal by the method. The relative cost ends up as

$$\frac{\mathcal{C}_{\mathrm{K+CW}}}{\mathcal{C}_{\mathrm{CW}}} = p_{\mathrm{FP}} + q_{\mathrm{K}}. \tag{8}$$

where $q_{\mathrm{K}} = c_{\mathrm{K}}/(c_{\mathrm{t}} n_{\mathrm{t}})$. As a result, Kaggle solutions will provide an advantage if they are computationally cheaper ($q_{\mathrm{K}} < 1$) and can reach a sufficiently low p_{FP} at a low p_{FD} level.

Strictly speaking, c_t and n_t depend on the specific implementation of a CW search [6]. We here instead take an educated guess based on the latest result in the literature. The latest `GCT` searches conducted in LIGO O2 and O3 data [21, 86] use $n_{\mathrm{t}} \sim 10^{12} - 10^{15}$, depending on the frequency band. Similarly, the computing cost of the $\mathcal{F}$-statistic is about $c_{\mathrm{t}} \sim (10 - 100)\,\mathrm{ns}$ [21, 86–88]. Thus, $c_{\mathrm{t}} n_{\mathrm{t}} \sim (10^4 - 10^7)\,\mathrm{s}$.

The computing cost per-band is reported by some of the participants: `JunKoda` reports a cost of 5 days on a GPU to evaluate the full dataset ($c_{\mathrm{K}} \approx 50\,\mathrm{s}$); `PreferredWave` reports 2 days on a GPU ($c_{\mathrm{K}} \approx 30\,\mathrm{s}$); `Hidden Neural Layers` reports 10 hours using 8 CPU cores ($c_{\mathrm{K}} \approx 5\,\mathrm{s}$); `Shun_PI` reports about 1 day using 1 CPU ($c_{\mathrm{K}} \approx 1\,\mathrm{s}$). No computing cost estimation is available for the rest of the teams. For algorithms based on numerical optimization we conservatively assume $c_{\mathrm{K}} \approx (1 - 10)\,\mathrm{s}$, while those based on neural networks can reach $c_{\mathrm{K}} \approx (0.1 - 1)\,\mathrm{s}$ if evaluations are batched on a GPU. Overall, we find $q_{\mathrm{K}} \lesssim 10^{-3}$.

We finally evaluate the performance of each Kaggle solution in restricted SNR ranges matching the two reference searches previously discussed, namely $\rho \in [40, 50]$ corresponding to `Weave` and $\rho \in [50, 60]$ corresponding to `GCT`. The results are shown in figure 6 for Gaussian and real-data samples. The number of signal samples within said ranges limits the resolution of false dismissal probabilities to about $(1-2)\%$, and consequently we evaluate p_{FP} at $1 - p_{\mathrm{FD}} = 0.98$. Similarly, the precision on p_{FP} is about 10^{-3}; given the computed value of q_{K}, this implies the computational gains are dominated by p_{FP}^{-1}.

For the `Weave` range ($\rho \in [40, 50]$), in the case of Gaussian noise, `JunKoda` reaches a false dismissal probability $p_{\mathrm{FD}} = 0.02$ at $p_{\mathrm{FP}} \sim (10^{-2} - 10^{-1})$, while `Preferred Wave` performs at $p_{\mathrm{FP}} \sim (0.2 - 0.4)$. All other solutions achieved the allowed false dismissal probability at $p_{\mathrm{FP}} \gtrsim 0.5$. The potential gain in this case is from a factor of about 5 (`Preferred Wave`) to one to two orders of magnitude (`JunKoda`). The situation in the case of real noise remains comparable within the uncertainty.

For the `GCT` range ($\rho \in [50, 60]$), in Gaussian noise, `JunKoda`, `Preferred Wave` and `Space Coders` reach $p_{\mathrm{FP}} \sim 10^{-3}$ at the specified false dismissal probability level. The next relevant team is `yu4u` (9th), at $p_{\mathrm{FP}} \sim 3 \times 10^{-2}$. The remaining solutions about $p_{\mathrm{FP}} \sim 10^{-1}$.

Despite the increased uncertainty, the situation in real-data samples changes slightly: `JunKoda` stays at $p_{\mathrm{FP}} \sim 10^{-3}$; `Preferred Wave` worsens up to $p_{\mathrm{FP}} \sim 7 \times 10^{-3}$, and `Space Coders` loses an order of magnitude up to $p_{\mathrm{FP}} \sim 3 \times 10^{-2}$. `yu4u` maintains a comparable performance to the Gaussian case. The relative loss of `PreferredWave` with respect to `JunKoda` can be ascribed to the use of average power in a template bank: In non-Gaussian data, the average template-bank power tends to rise due to the presence of extreme values in the background, even if most of the problematic bins have been removed. Similarly, the loss of `Space Coders` can be related to the fact that their method builds-up a frequency-evolution track in a data-driven manner, an thus can get biased due to local maxima in the spectrogram. Finally, we highlight the performance of `Hidden Neural Layers` (5th), which reaches $p_{\mathrm{FP}} \sim 10^{-3}$ surpassing all other teams except `JunKoda`. This is likely a consequence of their approach to isolate and remove potential glitches [105].

6. Discussion

CW signals are one of the potential next discoveries of GW astronomy. Currently their detection is complicated by our ignorance on the specific emission mechanism at play, which gives rise to a broad parameter-space that requires significant computing budgets to be effectively covered. Fundamentally, the sensitivity of CW searches for unknown sources is limited by the available computing [6, 15, 30].

In this work, we explored the use of data-analysis competitions to discover new search strategies. Specifically, we framed CW searches as a classification problem and deployed a Kaggle competition to find the most effective strategy to identify data samples harboring simulated CW signals. Kaggle [74] is a online data-analysis competition platform powered by Google hosting a broad community of data-analysis practitioners with a wide variety of backgrounds. By tapping into a new, broad community of experts, we intended to assess the performance of data-analysis methods from other field when applied to the problem of detecting CW signals. This approach is similar to mock data challenges previously developed for other GW use cases [56–59].

The challenge lasted for 3 months and attracted more than 1000 participants. The specific task was to classify 8000 data samples (containing Gaussian noise, real detector data, and possibly simulated CW signals). Each submission was ranked according to the Area under the ROC curve metric. Upon completion, we contacted the top 10 solutions to understand their approach and implementation.

Overall, the competition was dominated by pipelines based on semicoherent matched filtering, in a similar fashion to [13, 14]. The proposed solutions, however, introduced several improvements with respect to well established methods.

First, the evaluation of detection statistics was vectorized over multiple waveforms using GPUs without using any signal approximation. This contrasts with most GW GPU-accelerated methods so far, which either parallelize the evaluation of a single template [88, 118] or compute an approximated version of the detection statistic [117, 119]. The potential improvement in computational cost and complexity of search pipelines inspired the development of `fasttracks` [12], a unified, GPU-accelerated engine to compute CW detection statistics using SFTs for general CW waveform models.

Second, the winning solution [103] made use of alternative filtering strategies, similar to those previously outlined in [107]. While these specific statistics have not been applied to short-coherence CW searches, their efficiency allowed us to extend the use-case of efficient SFT-based detection statistics to the search for binary NSs and massive black-hole binaries [120] in next-generation ground-based detectors (Einstein Telescope [121], Cosmic Explorer [122]). Similar strategies have been proposed to analyze GW signals in the LISA [123] frequency band, such as galactic binaries [124, 125], which are modeled in a similar way to LIGO CW sources, and more recently a broader class of systems such as stellar-mass binary black holes [126] and extreme mass-ratio inspirals [127].

Solutions accounting for the presence of non-Gaussian data did so in a relatively simple manner, either by masking problematic regions in the spectrogram [103–106] or by generating non-Gaussian training data [112–115]. This is comparable to the practical approach taken in CW searches, where non-Gaussianities can be identified as sudden spikes in the accumulated power [12, 117].

Our analysis suggests that, for datasets compared to those used in this competition, the use of Kaggle solutions could reduce the number of frequency bands to be evaluated by one to two orders of magnitude.

The relevance of approaches based on alternative techniques, such as machine learning, becomes apparent when analyzing specific subsets of SNR. For example, for moderately strong signals $\rho \in [50, 60]$, we find that U-net architectures [114] are able to compete with track-based statistics and show a negligible performance degradation when applied to non-Gaussian data.

Despite the limited duration of this challenge, the obtained solutions have the potential to affect future CW searches not only by improving the efficiency and deployment of track-based searches [12, 120], but also by spurring approaches prescinding from template banks all together. This latter point is particularly interesting, as the computing cost of a CW search is dominated by the evaluation of statistics on a template bank to numerically marginalize with respect to the signal parameters to compute a Bayes factor [77, 78, 128]; this Kaggle competition, on the other hand, allows for pipelines operating directly on data samples, so that the method itself constitutes an approximation of the marginalized Bayes factor. The solutions here collected, together with those proposed from within the community [48–50, 52–54, 129], bear witness of the potential of alternative approaches to revolutionize the search for blind CW sources.

Although the dataset is shorter than typical production searches, the chosen duration (4 months) remains a valid regime for candidate generation. Given the expected cadence of forthcoming LVK observing runs, these results support a strategy based on analyzing each run shallowly to produce candidate lists, with significance subsequently accumulated across runs, akin to the 'fresh-data mode' discussed in [130]. Such implementation will be presented elsewhere.

This Kaggle competition is the first public, large-scale exposure of the CW detection problem to a wide community of data-analysis practitioners, and a logical step beyond established closed-door mock-data challenges [56–59] to foster cross-pollination across fields. Its impact is already evident in new GPU-accelerated methods [12] and in applying CW-inspired detection statistics to other domains, such as compact binary coalescences in next-generation GW detectors [120, 126, 127]. The materials developed for this competition, including tutorials for generating and pre-processing CW data with `PyFstat` [81], and the dataset itself [60], are publicly available to support further development of alternative methods aimed at the first CW detection.

Data availability statement

The data that support the findings of this study will be openly available following an embargo at the following URL/DOI: https://zenodo.org/records/17 060 457.

Acknowledgments

We are grateful to all Kaggle competitors for their invaluable input and engagement throughout the competition. We thank Rafel Jaume, David Keitel, Narenraju Nagarajan, Alicia M Sintes, and Daniel Williams for support and discussions, and Christopher Berry, Lorenzo Mirasola, and Karl Wette for comments on the manuscript. We thank Kaggle/Google for contributing to the competition prize, and acknowledge support from COST Action CA17137 (G2Net). RT thanks Graham Woan and the Institute for Gravitational Research at the University of Glasgow for their hospitality during the development of this work. RT is supported by ERC Starting Grant No. 945155-GWmining; Cariplo Foundation Grant No. 2021-0555; MUR PRIN Grant No. 2022-Z9X4XS; MUR Grant 'Progetto Dipartimenti di Eccellenza 2023-2027' (BiCoQ); Italian-French University (UIF/UFI) Grant No. 2025-C3-386; the ICSC National Research Centre funded by NextGenerationEU; the Universitat de les Illes Balears (UIB); the Spanish Agencia Estatal de Investigación Grants PID2022-138626NB-I00, RED2024-153978-E, RED2024-153735-E, funded by MICIU/AEI/10.13039/501100011033 and the ERDF/EU; and the Comunitat Autónoma de les Illes Balears through the Conselleria d'Educació i Universitats with funds from the European Union—NextGenerationEU/PRTR-C17.I1 (SINCO2022/6719) and from the European Union—European Regional Development Fund (ERDF) (SINCO2022/18146). MJW is supported by the Science and Technology Facilities Council [2285031, ST/X002225/1, ST/Y004876/1], and the University of Portsmouth. JB and CM are supported by STFC Grant ST/V005634/1. Computational work was performed at MareNostrum5 with technical support provided by the Barcelona Supercomputing Center (RES-FI-2024-3-0013, RES-FI-2025-1-0022), at CINECA with allocations through INFN and Bicocca. This research has made use of data or software obtained from the Gravitational Wave Open Science Center (gw-openscience.org), a service of LIGO Laboratory, the LIGO Scientific Collaboration, the Virgo Collaboration, and KAGRA. This paper has been assigned Document Number LIGO-P2200295.

Appendix. First Kaggle competition

The G2Net Kaggle challenge (G2Net Gravitational Wave Detection), launched on 30 June 2021, marked the first public data competition in gravitational-wave (GW) astrophysics aimed specifically at the

machine learning community. Its goal was to explore the capabilities of machine learning classifiers for detecting transient GW signals from stellar-mass binary black hole (BBH) mergers. Competitors were tasked with predicting the presence or absence of a BBH signal within short time-series recordings of simulated detector data. Similarly to the CW challenge discussed in this paper, the performance was also evaluated using the AUC, computed from participants' probability estimates of signal presence in the test set.

The dataset comprised simulated measurements from the three-detector network of advanced interferometers: LIGO-Hanford, LIGO-Livingston, and Virgo. Each training example consisted of three time series, corresponding to the three detectors, covering two seconds sampled at 2048 Hz. Some samples contained only Gaussian noise, while others were augmented with simulated GW signals. The simulated BBH signals were generated using `IMRPhenomPv2` [131] and varied across 15 astrophysically relevant parameters, including component masses, spins, binary orientation, sky location, distance, polarization, merger phase, and arrival time. Although the distributions of these parameters reflected realistic astrophysical populations, the overall signal amplitudes were adjusted to produce a mix of challenging and easier detection cases. Most signals were designed to lie near the typical sensitivity threshold of matched-filter searches, around an SNR of 8.

The challenge attracted broad participation, with 1501 individuals forming 1219 teams. Competitors applied a variety of machine learning approaches, from convolutional and recurrent neural networks to hybrid architectures, often incorporating data augmentation and ensemble methods to improve generalization. The competition validated claims in the literature [37, 132, 133] that deep learning methods could achieve competitive detection performance on realistically simulated multi-detector GW datasets. Moreover, the challenge provided a benchmark for future efforts in applying modern machine learning techniques to GW detection, highlighting both the potential and limitations of data-driven approaches in this domain.

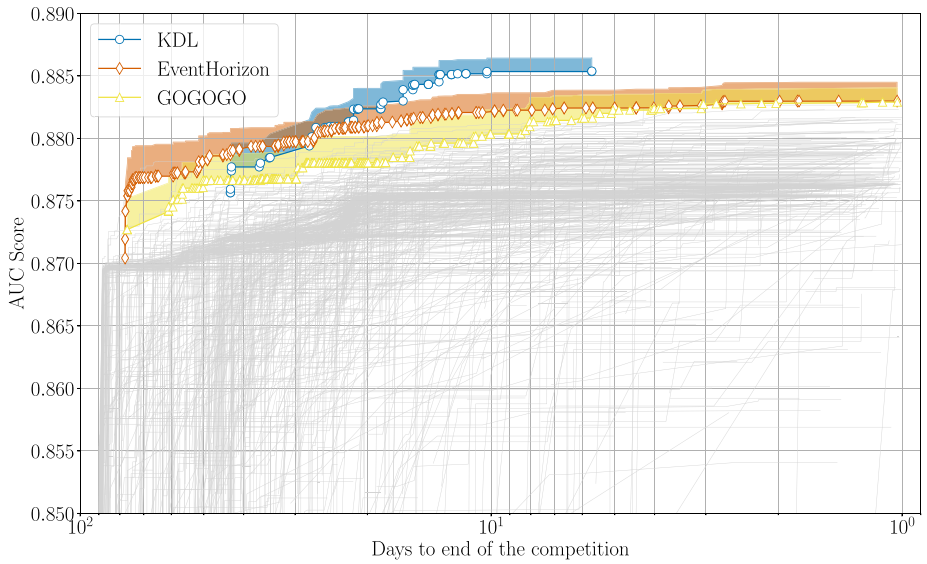

Figure A1. Evolution of the top-3 teams throughout the competition in terms the AUC score. The markers correspond to the private leaderboard score, while the shaded regions extend to the public leaderboard score. The evolution of the private leaderboard score for other teams is shown in grey.

In figure A1, we show the evolution of the AUC scores achieved by all participating teams over the course of the competition. Most final solutions clustered tightly around an AUC of ~ 0.88. The highlighted curves for the top three solutions indicate that their initial approaches could readily achieve AUCs of ~ 0.875. The ROC curves in figure A2 provide additional insight into the relative performance of the top three solutions. Of particular note is that the winners, `KDL`, were able to gain an advantage over 2nd place `EventHorizon` and 3rd place `GOGOGO`, achieving marginally higher overall probability of detection but, importantly, substantially better performance at the lowest false-positive rates. Figure A2 also provides the detection efficiencies of each of the top three teams as a function of network SNR at a fixed false positive probability equivalent to 1 per day. It is interesting to see clear distinction in sensitivity between the teams at efficiencies ~ 0.5 but the KDL team achieve efficiencies of 1 at significantly

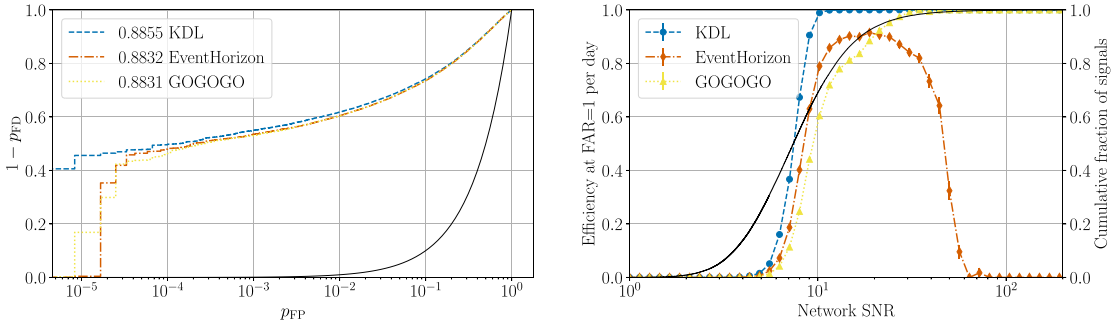

Figure A2. Left: ROC curves corresponding to the top-3 solutions of the first Kaggle competition. The legend lists the AUC scores corresponding to each solution. The black line corresponds to a non-discerning classifier with an AUC score of 0.5. Right: Efficiency as a function of network matched SNR (computed using the injection parameters) at a FAR of 1 per day for the top-3 solutions of the first Kaggle competition. We use 50 bins with logarithmic spacing between SNR 1 and 100 to produce the curve. The black line shows the cumulative fraction of signals in the original dataset as a function of SNR.

lower SNR. We also note that the 2nd place efficiency curve indicates a potential dataset bias or extrapolation failure where under-representation in training can lead to extremely loud signals being outside the learned feature distribution and hence classified incorrectly as noise.

The top three teams share several key strategies that contributed to their performance. They emphasized careful preprocessing of the waveform, including whitening, bandpass filtering, and mitigation of edge effects, while selectively using augmentation techniques such as Gaussian noise, waveform flipping, and time shifts to improve generalization. Each team used 1D convolutional architectures as primary frontends, often in combination with 2D CNN backends or hybrid models, and recognized the importance of channel-specific processing—sharing weights for similar detectors while handling others separately. Synthetic data generation and pre-training played a critical role in avoiding overfitting and providing access to auxiliary targets, and all groups used ensembles of multiple models to further boost robustness. Finally, iterative approaches like pseudo-labeling, test-time augmentation, and progressive refinement were widely employed to exploit unlabeled data and maximize leaderboard performance.

The winning team (KDL) overcame the main competition challenge—overfitting—by generating large amounts of realistic synthetic data for pre-training, which also allowed the use of hidden parameters as auxiliary targets. Their model design included learnable frontends for each channel, transformation of 1D data into richer time-frequency or feature spaces, and lean encoders to limit overfitting. Key innovations included a custom 1D CNN with large kernels, recognition that 1D models outperform 2D CNNs, and extensive use of synthetic data for pre-training. They also observed fundamental limitations: that about 30% of true signals are inherently undetectable, some noise samples are always misclassified as signals, and training on difficult samples can invert model predictions.

ORCID iDs

R Tenorio 0000-0002-3582-2587
M J Williams 0000-0003-2198-2974
J Bayley 0000-0003-2306-4106
C Messenger 0000-0001-7488-5022
J Koda 0009-0005-5764-0417
Y Yamakawa 0009-0009-5319-9782
H S T Bukhari 0009-0003-6174-7909
M V Shugaev 0000-0002-1841-3677
V Debout 0009-0003-7425-8514
J-L Tastet 0000-0003-1763-378X
I Timiryasov 0000-0001-9547-1347
O Ruchayskiy 0000-0001-8073-3068
D Kanonik 0009-0009-6102-3736
R Sato 0000-0001-8914-4905
A Zhyvalkouski 0009-0004-6905-440X
Y Uchida 0000-0002-6932-1465

S Yokoi 0009-0008-8770-4778
A Sayed 0009-0007-5384-0416
Z Wang 0009-0002-0374-311X

References

[1] Riles K 2023 Searches for continuous-wave gravitational radiation *Living Rev. Relativ.* **26** 3
[2] Haskell B and Bejger M 2023 Astrophysics with continuous gravitational waves *Nat. Astron.* **7** 1160–70
[3] Pagliaro G, Papa M A, Ming J, Lian J, Tsuna D, Maraston C and Thomas D 2023 Continuous gravitational waves from galactic neutron stars: demography, detectability and prospects *Astrophys. J.* **952** 123
[4] Hua Y, Wette K, Scott S M and Pitkin M D 2023 Population synthesis and parameter estimation of neutron stars with continuous gravitational waves and third-generation detectors *Mon. Not. R. Astron. Soc.* **527** 10 564–10 574
[5] Miller A L 2025 Gravitational wave probes of particle dark matter: a review vol 3 (arXiv: 2503.02607 [astro-ph.HE])
[6] Wette K 2023 Searches for continuous gravitational waves from neutron stars: a twenty-year retrospective *Astropart. Phys.* **153** 102880
[7] Tenorio R, Keitel D and Sintes A M 2021 Search methods for continuous gravitational-wave signals from unknown sources in the advanced-detector era *Universe* **7** 474
[8] Jaranowski P, Krolak A and Schutz B F 1998 Data analysis of gravitational - wave signals from spinning neutron stars: I. The Signal and its detection *Phys. Rev. D* **58** 063001
[9] Cutler C and Schutz B F 2005 The generalized $\mathcal{F}$-statistic: multiple detectors and multiple GW pulsars *Phys. Rev. D* **72** 063006
[10] Wette K 2015 Parameter-space metric for all-sky semicoherent searches for gravitational-wave pulsars *Phys. Rev. D* **92** 082003
[11] Bayley J, Messenger C and Woan G 2019 Generalized application of the Viterbi algorithm to searches for continuous gravitational-wave signals *Phys. Rev. D* **100** 023006
[12] Tenorio R, Mérou J-R and Sintes A M 2025 One-stop strategy to search for long-duration gravitational-wave signals *Phys. Rev. D* **111** 104002
[13] Krishnan B, Sintes A M, Papa M A, Schutz B F, Frasca S and Palomba C 2004 The Hough transform search for continuous gravitational waves *Phys. Rev. D* **70** 082001
[14] Astone P, Colla A, D'Antonio S, Frasca S and Palomba C 2014 Method for all-sky searches of continuous gravitational wave signals using the frequency-Hough transform *Phys. Rev. D* **90** 042002
[15] Prix R and Shaltev M 2012 Search for continuous gravitational waves: optimal stackslide method at fixed computing cost *Phys. Rev. D* **85** 084010
[16] Ashton G, Prix R and Jones D I 2017 Statistical characterization of pulsar glitches and their potential impact on searches for continuous gravitational waves *Phys. Rev. D* **96** 063004
[17] Mukherjee A, Messenger C and Riles K 2018 Accretion-induced spin-wandering effects on the neutron star in Scorpius X-1: Implications for continuous gravitational wave searches *Phys. Rev. D* **97** 043016
[18] Carlin J B and Melatos A 2025 How much spin wandering can continuous gravitational wave search algorithms handle? *Phys. Rev. D* **111** 083016
[19] Dergachev V and Papa M A 2025 Early release of low-frequency atlas of continuous gravitational waves (arXiv: 2507.12161) vol 7
[20] Dergachev V and Papa M A 2025 Expanded atlas of the sky in continuous gravitational waves (arXiv: 2503.11512) vol 3
[21] Steltner B, Papa M A, Eggenstein H B, Prix R, Bensch M, Allen B and Machenschalk B 2023 Deep Einstein@Home all-sky search for continuous gravitational waves in LIGO O3 public data *Astrophys. J.* **952** 55
[22] Abbott R *et al* 2022 All-sky search for continuous gravitational waves from isolated neutron stars using Advanced LIGO and Advanced Virgo O3 data *Phys. Rev. D* **106** 102008
[23] Abbott R *et al* 2021 All-sky search in early O3 LIGO data for continuous gravitational-wave signals from unknown neutron stars in binary systems *Phys. Rev. D* **103** 064017
Abbott R *et al* 2023 All-sky search in early O3 LIGO data for continuous gravitational-wave signals from unknown neutron stars in binary systems *Phys.Rev. D* **108** 069901 (erratum)
[24] Covas P B, Papa M A, Prix R and Owen B J 2022 Constraints on r-modes and mountains on millisecond neutron stars in binary systems *Astrophys. J. Lett.* **929** L19
[25] Covas P B, Papa M A and Prix R 2025 Search for continuous gravitational waves from unknown neutron stars in binary systems with long orbital periods in O3 data *Astrophys. J.* **985** 192
[26] Ushomirsky G, Cutler C and Bildsten L 2000 Deformations of accreting neutron star crusts and gravitational wave emission *Mon. Not. R. Astron. Soc.* **319** 902
[27] Haskell B, Priymak M, Patruno A, Oppenoorth M, Melatos A and Lasky P D 2015 Detecting gravitational waves from mountains on neutron stars in the advanced detector Era *Mon. Not. R. Astron. Soc.* **450** 2393–403
[28] Hutchins T J and Jones D I 2023 Gravitational radiation from thermal mountains on accreting neutron stars: sources of temperature non-axisymmetry *Mon. Not. R. Astron. Soc.* **522** 226–51
[29] Pletsch H J and Allen B 2009 Exploiting global correlations to detect continuous gravitational waves *Phys. Rev. Lett.* **103** 181102
[30] Wette K, Walsh S, Prix R and Papa M A 2018 Implementing a semicoherent search for continuous gravitational waves using optimally-constructed template banks *Phys. Rev. D* **97** 123016
[31] Covas P B and Prix R 2022 Improved all-sky search method for continuous gravitational waves from unknown neutron stars in binary systems *Phys. Rev. D* **106** 084035
[32] Ashton G and Prix R 2018 Hierarchical multistage MCMC follow-up of continuous gravitational wave candidates *Phys. Rev. D* **97** 103020
[33] Tenorio R, Keitel D and Sintes A M 2021 Application of a hierarchical MCMC follow-up to Advanced LIGO continuous gravitational-wave candidates *Phys. Rev. D* **104** 084012
[34] Covas P B, Prix R and Martins J 2024 New framework to follow up candidates from continuous gravitational-wave searches *Phys. Rev. D* **110** 024053
[35] Mirasola L and Tenorio R 2024 Toward a computationally efficient follow-up pipeline for blind continuous gravitational-wave searches *Phys. Rev. D* **110** 124049
[36] Cuoco E, Cavaglià M, Heng I S, Keitel D and Messenger C 2025 Applications of machine learning in gravitational-wave research with current interferometric detectors *Living Rev. Relativ.* **28** 2

[37] Gabbard H, Williams M, Hayes F and Messenger C 2018 Matching matched filtering with deep networks for gravitational-wave astronomy *Phys. Rev. Lett.* **120** 141103
[38] Chua A J K and Vallisneri M 2020 Learning Bayesian posteriors with neural networks for gravitational-wave inference *Phys. Rev. Lett.* **124** 041102
[39] Koloniari A E, Koursoumpa E C, Nousi P, Lampropoulos P, Passalis N, Tefas A and Stergioulas N 2025 New gravitational wave discoveries enabled by machine learning *Mach. Learn. Sci. Technol.* **6** 015054
[40] Nagarajan N and Messenger C 2025 Identifying and mitigating machine learning biases for the gravitational-wave detection problem (arXiv: 2501.13846) [gr-qc]
[41] Gabbard H, Messenger C, Heng I S, Tonolini F and Murray-Smith R 2022 Bayesian parameter estimation using conditional variational autoencoders for gravitational-wave astronomy *Nat. Phys.* **18** 112–7
[42] Dax M, Green S R, Gair J, Macke J H, Buonanno A and Schölkopf B 2021 Real-time gravitational wave science with neural posterior estimation *Phys. Rev. Lett.* **127** 241103
[43] Williams M J, Veitch J and Messenger C 2021 Nested sampling with normalizing flows for gravitational-wave inference *Phys. Rev. D* **103** 103006
[44] Williams M J, Veitch J and Messenger C 2023 Importance nested sampling with normalising flows *Mach. Learn. Sci. Technol.* **4** 035011
[45] Usman S A *et al* 2016 The PyCBC search for gravitational waves from compact binary coalescence *Class. Quantum Grav.* **33** 215004
[46] Sachdev S *et al* 2019 The GstLAL search analysis methods for compact binary mergers in advanced LIGO's second and advanced Virgo's first observing runs
[47] Cornish N J, Littenberg T B, Bécsy B, Chatziioannou K, Clark J A, Ghonge S and Millhouse M 2021 BayesWave analysis pipeline in the era of gravitational wave observations *Phys. Rev. D* **103** 044006
[48] Dreissigacker C, Sharma R, Messenger C, Zhao R and Prix R 2019 Deep-learning continuous gravitational waves *Phys. Rev. D* **100** 044009
[49] Yamamoto T S and Tanaka T 2021 Use of an excess power method and a convolutional neural network in an all-sky search for continuous gravitational waves *Phys. Rev. D* **103** 084049
[50] Joshi P M and Prix R 2025 Transformer networks for continuous gravitational-wave searches vol 9 (arXiv: 2509.10912)
[51] Cheung D H T 2025 Attention U-Net for all-sky continuous gravitational wave searches vol 9 (arXiv: 2509.19838)
[52] Modafferi L M, Tenorio R and Keitel D 2023 Convolutional neural network search for long-duration transient gravitational waves from glitching pulsars *Phys. Rev. D* **108** 023005
[53] Bayley J, Messenger C and Woan G 2020 Robust machine learning algorithm to search for continuous gravitational waves *Phys. Rev. D* **102** 083024
[54] Bayley J, Messenger C and Woan G 2022 Rapid parameter estimation for an all-sky continuous gravitational wave search using conditional varitational auto-encoders *Phys. Rev. D* **106** 083022
[55] G2net detecting continuous gravitational waves 2022 (available at: https://kaggle.com/competitions/g2net-detecting-continuous-gravitational-waves)
[56] Messenger C *et al* 2015 Gravitational waves from Scorpius X-1: a comparison of search methods and prospects for detection with advanced detectors *Phys. Rev. D* **92** 023006
[57] Walsh S *et al* 2016 Comparison of methods for the detection of gravitational waves from unknown neutron stars *Phys. Rev. D* **94** 124010
[58] Schäfer M B *et al* 2023 First machine learning gravitational-wave search mock data challenge *Phys. Rev. D* **107** 023021
[59] Baghi Q 2022 The LISA data challenges *56th Rencontres de Moriond on Gravitation*
[60] Tenorio R, Williams M J, Bayley J, Messenger C, Demkin M and Reade W 2025 (available at: https://zenodo.org/records/17060457)
[61] (Available at: www.kaggle.com/competitions/g2net-detecting-continuous-gravitational-waves/discussion/347052)
[62] Ashton G, Keitel D, Prix R, Tenorio R and Ferrer M-A PyFstat *zenodo* (available at: https://doi.org/10.5281/zenodo.3967045)
[63] G2net gravitational wave detection 2021 (available at: www.kaggle.com/competitions/g2net-gravitational-wave-detection)
[64] Pan I, Cadrin-Chênevert A and Cheng P M 2019 Tackling the radiological society of North America pneumonia detection challenge *American Journal of Roentgenology* **213** 568–74
[65] Baid U *et al* 2021 The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification (arXiv: 2107.02314)
[66] Wayment-Steele H K *et al* 2021 Deep learning models for predicting RNA degradation via dual crowdsourcing (arXiv: 2110.07531)
[67] Bojer C S and Meldgaard J P 2021 Kaggle forecasting competitions: an overlooked learning opportunity *Int. J. Forecast.* **37** 587–603
[68] Calafiura P *et al* 2018 TrackML: a high energy physics particle tracking challenge *14th Int. Conf. on e-Science* p 344
[69] Sculley D *et al* 2025 Position: AI competitions provide the gold standard for empirical rigor in GenAI evaluation (arXiv: 2505.00612)
[70] Zooniverse (available at: www.zooniverse.org/about/resources)
[71] Zevin M *et al* 2017 Gravity Spy: integrating advanced LIGO detector characterization, machine learning and citizen science *Class. Quantum Grav.* **34** 064003
[72] Zevin M *et al* 2024 Gravity Spy: lessons learned and a path forward *Eur. Phys. J. Plus* **139** 100
[73] Razzano M, Di Renzo F, Fidecaro F, Hemming G and Katsanevas S 2023 GWitchHunters: machine learning and citizen science to improve the performance of gravitational wave detector *Nucl. Instrum. Meth. A* **1048** 167959
[74] (Available at: www.kaggle.com/)
[75] Soni S *et al* 2021 Discovering features in gravitational-wave data through detector characterization, citizen science and machine learning *Class. Quantum Grav.* **38** 195016
[76] Mackenzie E, Berry C P L, Niklasch G, Téglás B, Unsworth C, Crowston K, Davis D and Katsaggelos A K 2025 Hunting for new glitches in LIGO data using community science *24th Int. Conf. on General Relativity and Gravitation (GR24) and 16th Edoardo Amaldi Conf. on Gravitational (Amaldi16) Waves*
[77] Prix R and Krishnan B 2009 Targeted search for continuous gravitational waves: Bayesian versus maximum-likelihood statistics *Class. Quantum Grav.* **26** 204013

[78] Searle A C 2008 Monte-Carlo and Bayesian techniques in gravitational wave burst data analysis *12th Gravitational Wave Data Analysis Workshop*
[79] LIGO Scientific Collaboration, Virgo Collaboration and KAGRA Collaboration 2018 LVK Algorithm Library - LALSuite *Free software (GPL)*
[80] Wette K 2020 SWIGLAL: Python and Octave interfaces to the LALSuite gravitational-wave data analysis libraries *SoftwareX* **12** 100634
[81] Keitel D, Tenorio R, Ashton G and Prix R 2021 PyFstat: a Python package for continuous gravitational-wave data analysis *J. Open Source Softw.* **6** 3000
[82] Aasi J *et al* 2015 Advanced LIGO *Class. Quantum Grav.* **32** 074001
[83] Acernese F *et al* 2015 Advanced Virgo: a second-generation interferometric gravitational wave detector *Class. Quantum Grav.* **32** 024001
[84] Akutsu T *et al* 2019 KAGRA: 2.5 generation interferometric gravitational wave detector *Nat. Astron.* **3** 35–40
[85] Akutsu T *et al* 2021 Overview of KAGRA: detector design and construction history *Progr. Theor. Exper. Phys.* **2021** 05A101
[86] Steltner B *et al* 2021 Einstein@Home all-sky search for continuous gravitational waves in LIGO O2 public data *Astrophys. J.* **909** 79
[87] Wette K, Dunn L, Clearwater P and Melatos A 2021 Deep exploration for continuous gravitational waves at 171–172 Hz in LIGO second observing run data *Phys. Rev. D* **103** 083020
[88] Dunn L, Clearwater P, Melatos A and Wette K 2022 Graphics processing unit implementation of the $\mathcal{F}$-statistic for continuous gravitational wave searches *Class. Quantum Grav.* **39** 045003
[89] Bruce A *et al* 2004 SFT (Short-Time Fourier Transform) data format version 2 specification (available at: https://dcc.ligo.org/LIGO-T040164/public)
[90] Abbott R *et al* 2023 Open data from the third observing run of LIGO, Virgo, KAGRA and GEO *Astrophys. J. Suppl.* **267** 29
[91] LIGO – Virgo Collaboration 2021 O3a data release (available at: https://doi.org/10.7935/nfnt-hm34)
[92] Abbott R *et al* 2021 O3b data release (available at: https://doi.org/10.7935/pr1e-j706)
[93] Goetz E and Riles K 2023 Segments used for creating standard SFTs in O3 data (available at: https://dcc.ligo.org/LIGO-T2300068)
[94] Buikema A *et al* 2020 Sensitivity and performance of the advanced ligo detectors in the third observing run *Phys. Rev. D* **102** 062003
[95] Davis D *et al* 2021 LIGO detector characterization in the second and third observing runs *Class. Quantum Grav.* **38** 135014
[96] Zhu S J, Baryakhtar M, Papa M A, Tsuna D, Kawanaka N and Eggenstein H-B 2020 Characterizing the continuous gravitational-wave signal from boson clouds around Galactic isolated black holes *Phys. Rev. D* **102** 063020
[97] Dreissigacker C, Prix R and Wette K 2018 Fast and accurate sensitivity estimation for continuous-gravitational-wave searches *Phys. Rev. D* **98** 084058
[98] Wette K 2012 Estimating the sensitivity of wide-parameter-space searches for gravitational-wave pulsars *Phys. Rev. D* **85** 042003
[99] Abbott R *et al* 2021 Open data from the first and second observing runs of Advanced LIGO and Advanced Virgo *SoftwareX* **13** 100658
[100] Whitehill J 2015 Exploiting an oracle that reports AUC scores in machine learning contests (arXiv: 1506.01339)
[101] Whitehill J 2017 Climbing the Kaggle leaderboard by exploiting the log-loss oracle (arXiv: 1707.01825)
[102] Whitehill J 2017 How does knowledge of the auc constrain the set of possible ground-truth labelings? (arXiv: 1709.02418)
[103] Koda J 2023 (available at: www.kaggle.com/competitions/g2net-detecting-continuous-gravitational-waves/discussion/375910)
[104] PreferredWave 2023 (available at: www.kaggle.com/competitions/g2net-detecting-continuous-gravitational-waves/discussion/376504)
[105] Hidden Neural Layers 2023 (available at: www.kaggle.com/competitions/g2net-detecting-continuous-gravitational-waves/discussion/376022)
[106] Shun_PI 2023 (available at: www.kaggle.com/competitions/g2net-detecting-continuous-gravitational-waves/discussion/375923)
[107] Allen B, Papa M A and Schutz B F 2002 Optimal strategies for sinusoidal signal detection *Phys. Rev. D* **66** 102003
[108] Storn R and Price K 1997 Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces *J. Global Optim.* **11** 341–59
[109] Kirkpatrick S, Gelatt C D and Vecchi M P 1983 Optimization by simulated annealing *Science* **220** 671–80
[110] Nelder J A and Mead R 1965 A simplex method for function minimization *Comput. J.* **7** 308–13
[111] Tenorio R, Modafferi L M, Keitel D and Sintes A M 2022 Empirically estimating the distribution of the loudest candidate from a gravitational-wave search *Phys. Rev. D* **105** 044029
[112] Bear Waves 2023 (available at: www.kaggle.com/competitions/g2net-detecting-continuous-gravitational-waves/discussion/376233)
[113] MC Digital 2023 (available at: www.kaggle.com/competitions/g2net-detecting-continuous-gravitational-waves/discussion/376253)
[114] yu4u 2023 (available at: www.kaggle.com/competitions/g2net-detecting-continuous-gravitational-waves/discussion/375897)
[115] anonamename 2023 (available at: www.kaggle.com/competitions/g2net-detecting-continuous-gravitational-waves/discussion/376052)
[116] Viterbi A 1967 Error bounds for convolutional codes and an asymptotically optimum decoding algorithm *IEEE Trans. Inf. Theory* **13** 260–9
[117] Covas P B and Sintes A M 2019 New method to search for continuous gravitational waves from unknown neutron stars in binary systems *Phys. Rev. D* **99** 124019
[118] García-Quirós C, Tiwari S and Babak S 2025 GPU-accelerated LISA parameter estimation with full time domain response (arXiv: 2501.08261)
[119] Rosa I L, Astone P, D'Antonio S, Frasca S, Leaci P, Miller A L, Palomba C, Piccinni O J, Pierini L and Regimbau T 2021 Continuous gravitational-wave data analysis with general purpose computing on graphic processing units *Universe* **7** 218
[120] Tenorio R and Gerosa D 2025 Scalable data-analysis framework for long-duration gravitational waves from compact binaries using short Fourier transforms *Phys. Rev. D* **111** 104044
[121] Abac A *et al* 2025 The science of the Einstein telescope vol 3
[122] Reitze D *et al* 2019 Cosmic explorer: the U.S. contribution to gravitational-wave astronomy beyond LIGO *Bull. Am. Astron. Soc.* **51** 035
[123] Colpi M *et al* 2024 LISA definition study report

[124] Prix R and Whelan J T 2007 $\mathcal{F}$-statistic search for white-dwarf binaries in the first Mock LISA data challenge *Class. Quantum Grav.* **24** S565–74
[125] Blaut A, Babak S and Krolak A 2010 Mock LISA Data Challenge for the galactic white dwarf binaries *Phys. Rev. D* **81** 063008
[126] Bandopadhyay D, Chapman-Bird C E A and Vecchio A 2025 Global time-frequency search for stellar-mass binary black holes in LISA (arXiv: 2510.19047)
[127] Speri L, Tenorio R, Chapman-Bird C and Gerosa D 2025 Ab uno disce omnes: single-harmonic search for extreme mass-ratio inspirals (arXiv: 2510.20891)
[128] Prix R, Giampanis S and Messenger C 2011 Search method for long-duration gravitational-wave transients from neutron stars *Phys. Rev. D* **84** 023007
[129] Miller A L *et al* 2019 How effective is machine learning to detect long transient gravitational waves from neutron stars in a real search? *Phys. Rev. D* **100** 062005
[130] Cutler C, Gholami I and Krishnan B 2005 Improved stack-slide searches for gravitational-wave pulsars *Phys. Rev. D* **72** 042004
[131] Hannam M, Schmidt P, Bohé A, Haegel L, Husa S, Ohme F, Pratten G and Pürrer M 2014 Simple model of complete precessing black-hole-binary gravitational waveforms *Phys. Rev. Lett.* **113** 151101
[132] George D and Huerta E A 2018 Deep neural networks to enable real-time multimessenger astrophysics *Phys. Rev. D* **97** 044039
[133] Gebhard T D, Kilbertus N, Harry I and Schölkopf B 2019 Convolutional neural networks: a magic bullet for gravitational-wave detection? *Phys. Rev. D* **100** 063015
[134] Tenorio R, Williams M J, Bayley J, Demkin M and Reade W 2025 G2Net Detecting Continuous Gravitational Waves *Zenodo* (https://doi.org/10.5281/zenodo.17060457)