

A Neural Network Approach to Preferred Event Selection for Low-Latency Gravitational-Wave Alerts

Pratyusava Baral,* Cody Messick, and Patrick Brady

University of Wisconsin-Milwaukee, Milwaukee, WI 53201, USA

Abstract

The LIGO-Virgo-KAGRA collaboration uses multiple independent search pipelines to detect gravitational waves, often resulting in multiple triggers (g-events) for a single astrophysical source. These triggers are grouped into superevents, raising a critical question for multimessenger astronomy: which g-event provides the most accurate sky localization for electromagnetic follow-up? Currently, the g-event with the highest signal-to-noise ratio (SNR) is selected, under the assumption that it should provide the best estimators of the source’s parameters, including its location on the sky. Analysis of simulated signals reveals systematic deviations from this expectation. In particular, a false-alarm rate (FAR)-based selector performs slightly better than the SNR-based method, but introduces pipeline biases. We present a neural network-based selector trained on simulated signals to identify the g-event with the minimum searched area – a metric quantifying localization accuracy. The network uses information (detector SNRs, FAR, and chirp mass) from all of the triggers associated with each astrophysical source and is designed to be pipeline-agnostic. Our results show that the neural network outperforms both traditional selectors, achieving a mean searched area $\sim 2\%$ smaller than the SNR-based selector. Unlike FAR-based selection, the neural network preserves the underlying distribution of pipeline contributions, avoiding systematic biases toward specific pipelines. The network can be trained in approximately one minute on a few thousand events and performs event selection instantaneously, making it suitable for low-latency applications. These results demonstrate that machine learning can enhance multimessenger astronomy capabilities while maintaining fairness across detection pipelines. We recommend implementing this approach for future observing runs.

* pbaral@uwm.edu

I. INTRODUCTION

The LIGO-Virgo-KAGRA (LVK) collaboration has detected gravitational waves (GWs) [1–4] from the merger of binary black holes (BBHs) [5], binary neutron stars (BNSs) [6], and neutron-star-black-hole mergers (NSBHs) [7]. The fourth observing run (O4) ended on Nov 18, 2025 [8]. Building upon the ~ 100 confident detections from the first three observing runs (O1–O3), the fourth observing run has so far yielded ~ 250 additional gravitational wave candidates. The first BNS merger GW170817 [6], was detected in low-latency during the second observing run in coincidence with GRB170817A [9–11] and the optical counterpart AT2017gfo [12–14]. Such coincident multimessenger observations enable measurements of the NS equation-of-state (EoS) [15–25], the Hubble constant [24, 26–29], and r -process nucleosynthesis [13, 30–35].

There are two types of low-latency search pipelines in the current observing run [4, 36, 37], namely compact-binary-coalescence (CBC) searches that use matched filtering based on pre-computed waveforms and weakly-modeled burst searches. The CBC searches [38] target primarily BNS, NSBH, and BBH mergers. Weakly-modeled burst searches [39] look for consistent signal morphologies from various detectors and are capable of detecting a variety of astrophysical sources like core-collapse supernova, magnetic starquakes, in addition to standard CBC sources [39]. Multiple independent search pipelines process strain data to identify GW triggers. Each pipeline employs distinct methods for trigger generation, background estimation, and significance estimation, resulting in slightly different sensitivities [40–46] across the parameter space. This multi-pipeline approach maximizes detection efficiency [4], enables cross-validation of results, and provides operational redundancy. For this paper, we only consider triggers from CBC pipelines.

The triggers from low-latency pipelines are fed into an event orchestrator called GWCelery [47] [48], which annotates the triggers with data-quality information, computes a number of astrophysical attributes, and issues public alerts when certain pre-defined conditions are satisfied. The Gravitational-wave Candidate Event Database (GraceDB) [49] stores information about the candidate events and provides a web interface to view them. The low-latency alert infrastructure is described in detail in [50], and the LVK Alert User Guide [51] is used to communicate changes to this system to the broader astrophysical community.

Each GW trigger is assigned a Grace ID and is referred to as a g-event. We often see

multiple g-events corresponding to the same candidate event. This occurs both because we employ multiple search pipelines, and because a single pipeline may upload several g-events for a given astrophysical event. These g-events are grouped when they are believed to originate from the same astrophysical source, forming what is known as a superevent. A natural question that arises is, among all the g-events in a superevent, which one most closely represents the properties of the astrophysical source? This work proposes a neural network-based selector trained to select the event with the *best* skymap in low-latency. The *best* skymap is chosen using a metric called searched area, which is described in the following section.

The paper is organized as follows. In Section 2, we describe the method we use currently for selecting the preferred event. Section 3 describes the data used for the paper. Section 4 describes the details of the neural network-based selector developed in this paper. Section 5 presents our key results, and Section 6 is reserved for discussions.

II. PREFERRED EVENT SELECTION

A key motivation for low-latency GW detection is multimessenger astronomy. Each CBC g-event in a superevent has a skymap generated in low-latency by an algorithm called BAYESian TriAngulation and Rapid localization (BAYESTAR) [52]. The skymap is a two-dimensional map of the posterior probabilities of right ascension and declination, or a three-dimensional map that also includes the luminosity distance to the source. Three-dimensional skymaps generated using BAYESTAR are used for public alerts. Our goal is to identify the event with the *best* skymap for multimessenger follow-up. A key parameter associated with CBC triggers is the signal-to-noise ratio, which represents the ratio of the power contained in the signal to the noise. Under the assumption of a stationary Gaussian noise model, the highest SNR event is expected to be the best match to the observed data. It also yields the most constrained skymap, with the smallest 90% credible region, making it a natural choice for multimessenger follow-up. However, real detector noise is not perfectly Gaussian, and these desirable properties are not guaranteed to hold. This leads us to ask the question: *How can we select the best skymap for electromagnetic follow-up?*

The primary constraint for electromagnetic follow-up observations is accurate localization. In simulations, the searched area, defined as the sky area encompassed by pixels in the

localization skymap with posterior probability densities exceeding those of the true source location, can be used as a metric to identify the best sky map. Better localization results in decreased telescope time allocation and enhanced probability of transient detection. Under the assumption of stationary Gaussian noise, the trigger with maximum SNR is also the maximum likelihood estimator of the signal [53], and is expected to yield the minimum searched area. We evaluate this hypothesis by injecting simulated GW signals into realistic detector noise realizations. Given our knowledge of the injected source positions, we compute the corresponding searched areas and compare the performance of events ranked by minimum searched area versus maximum SNR. Our analysis reveals a systematic deviation from expectations, which might be due to non-Gaussianities in detector noise.

Selecting the g-event with the lowest FAR has been proposed as an alternative to selecting the highest-SNR g-event. A key complication arises from the fact that each pipeline defines a FAR in a self-consistent way, and as a result FAR values are not directly comparable across pipelines [4]. Nevertheless based on GWTC-4 results, SNR systematics across pipelines are also non-negligible [4].

III. DATA

We use simulated GW signals in real detector data for our study. A total of 50,000 simulated CBC waveforms, or injections, were added to a stretch of data collected by the LVK instruments between Jan 05, 2020 and Feb 14, 2020 as a part of the third observing run (O3) [50]. Each injection, simulated using IMRPhenomPv2_NRTidalv2 waveform family, has an optimal network SNR greater than 4, ensuring that signals are realistically detectable by search pipelines and avoiding those that would be impossible to recover. The SLy EoS [54] is used to map from component masses to the tidal deformability parameters. The maximum NS mass allowed by the SLy EoS is $\sim 2.05M_{\odot}$, and all objects above this mass are classified as BHs with the tidal parameters set to 0. The injection set is comprised of 40.9% BNS, 35.8% NSBH, and 23.3% BBH mergers. These simulated events are distributed uniformly in comoving volume, assuming a flat Λ CDM cosmology based on Planck 2018 results [55]. The rate of injections is artificially high and not astrophysical.

The simulated strain is projected onto the detectors, timeshifted to fit within the 40-day window, and streamed in 1-second segments to enable low-latency analysis by search

and other annotation pipelines. This exercise is repeated in several cycles, even during the observation run, to benchmark analysis, continuously track improvements in the alert infrastructure, and provide avenues for pipelines to test their changes. It serves as an end-to-end testbed for data ingestion, execution of online search pipelines, event annotation, and the issuance of alerts to the astrophysics community.

This study is based on the 11th iteration of the datastream, referred to as Mock Data Challenge (MDC) 11, which was conducted before the start of O4. The MDC 11 data were analyzed by four compact binary coalescence (CBC) search pipelines: Multi-Band Template Analysis (MBTA) [41, 56, 57], GStreamer LIGO Algorithm Library (GstLAL) [40, 58–66], the Python-based CBC search software (PyCBC) [45, 67–71], and Summed Parallel Infinite Impulse Response (SPIIR) [42, 43]. In addition to the CBC searches, two burst searches—Coherent WaveBurst (cWB) [72–74] and Omicron LALInferenceBurst (oLIB) [75]—and an external coincidence search, Rapid On-source VOEvent Coincident Monitor (RAVEN) [46, 76, 77], were also operational during the challenge. However, we focus exclusively on the CBC search pipeline uploads in this work.

A. Data Preparation

A g-event from a CBC pipeline is labeled as significant if it is associated with a FAR below 3.9×10^{-7} Hz (one per month). At the time of the MDC, four distinct CBC search pipelines, plus RAVEN, operated independently in low latency. While these pipelines are not strictly independent (as they are all searching for the same underlying signals), they often respond differently to detector noise. To account for the use of multiple pipelines, we take the number of trials to equal the number of pipelines. This likely leads to an overestimate of the false-alarm-rate, but we adopt this approach to avoid overestimating significance. Accounting for the trials factor, we mark g-events from CBC searches as significant when $\text{FAR} < 7.7 \times 10^{-8}$ Hz (one per 5 months).

For this study, g-events were classified into two distinct categories based on the FAR. Events with a $\text{FAR} < 7.7 \times 10^{-8}$ Hz are referred to as significant g-events. In a low-latency search, such events would correspond to astrophysical events [78]. A superevent which has at least one significant g-event is referred to as a significant event henceforth. The other category mimics the set of events made public during O4 and is comprised of the set of

superevents that have at least a g-event recovered with $\text{FAR} < 2.3 \times 10^{-5}$ Hz (approximately two false alarms per day). This category is referred to as all events henceforth. This dataset typically has triggers of both astrophysical and non-astrophysical origin. However, for this study, we retain only the injected astrophysical triggers and exclude all non-astrophysical triggers from consideration because we need to compute the searched area.

Each g-event is characterized by five observational features: the SNRs from the H1, L1, and V1 detectors, the logarithm of the FAR, and the chirp mass. We impose a maximum limit of 22 g-events per superevent. This corresponds to the maximum number of g-events corresponding to a superevent in MDC11. While 22 serves as a practical threshold in this study, larger values can be adopted in future analyses without loss of generality. Superevents containing fewer than 22 constituent g-events are zero-padded to maintain uniform input dimensionality. Consequently, each superevent is encoded as a fixed-dimension matrix of 22×5 floating-point values. The g-events within each superevent are ordered by ascending chirp mass and indexed from 0 to 21, with index 0 corresponding to the g-event with the lowest chirp mass. This sorting procedure ensures that the neural network predictions remain invariant to the temporal sequence of g-event detection. Alternative sorting parameters may be substituted. The sorting does not change the outcome of the neural network-based selector as shown in the appendix.

To assess model performance and evaluate robustness against data partitioning, the complete superevent dataset is split into training and testing sets using a 9:1 ratio. Identical superevent allocations are maintained across both high-significance and low-significance filtered datasets, ensuring consistency. This partitioning procedure is iterated 10 times, yielding 10 training-testing pairs. The 10 test pairs do not overlap and are independent datasets. This approach enables us to assess the stability and predictive uncertainty of the neural network across different data configurations, helping ensure that performance is not overly sensitive to a particular split.

IV. METHODS

We present a neural network-based approach to identify the preferred event based solely on the outputs of the search pipelines, without relying on any knowledge of the true sky location of the injected signal. The goal of this work is to assess whether a neural network

can be developed that takes as input the collective outputs from multiple low-latency search pipelines and successfully selects the event with the *best* skymap for multimessenger follow-up. For training purposes, we use the skymap with the lowest searched area as the *best* skymap. Our network is designed to be pipeline-agnostic, a deliberate choice to mitigate systematic biases towards any specific search algorithm. By treating all pipelines equally and learning only from the aggregate features they report, we aim to achieve robust generalization across diverse pipeline-specific characteristics.

A. Network Architecture

We implement a fully connected feedforward architecture using PyTorch, comprising of an input layer, two hidden layers, and an output layer.

- Input layer: A linear transformation from 110 (flattened from the 22×5 matrix representing pipeline outputs) input features to 256 hidden units. The inputs of the layer are normalized using one-dimensional batch normalization [79], followed by a Rectified Linear Unit (ReLU) activation function to introduce non-linearity.
- Hidden layer: Two successive hidden layers progressively reduce dimensionality from 256 to 128, then to 64 units. Each transformation incorporates batch normalization and ReLU activation to maintain stable training dynamics.
- Output layer: The final layer projects the hidden representation consisting of 64 units to 22 output units, corresponding to the 22 possible g-event indices within each superevent. These outputs represent raw scores (logits), which are fed into a categorical cross-entropy loss function during training to guide the model toward predicting the index of the g-event with the lowest searched area. We use a categorical cross-entropy loss function [80], regularized by the square of the sum of the scores of the zero-padded events, weighed by a factor of 0.01. The position of the highest raw score is chosen to be the index of the preferred event.

To prevent overfitting and ensure robust generalization to unseen GW events, we implemented several regularization techniques throughout the network architecture. Batch normalization, introduced in 2015 [79], provides fast and stable training by scaling and

re-centering the inputs to each layer. Dropout regularization [81] with a rate of 30% was employed between hidden layers to prevent co-adaptation of neurons and improve model robustness. This essentially randomly ignores some nodes, which helps to improve performance on unseen data by reducing overfitting. L2 weight decay with a coefficient of 1×10^{-4} was applied during optimization to keep the weights small, which prevents the gradients from exploding during training. Network weights were initialized using Xavier uniform initialization [82] to ensure stable gradient flow during early training phases. These comprehensive regularization strategies were chosen due to the limited size of the dataset relative to the model complexity.

To optimize the network, we use the Adam optimizer [83] with an initial learning rate of 0.001. We employ cross-entropy loss as the objective function, which is well-suited for the multi-class classification task of selecting the preferred event from candidate sets. Training is performed with mini-batches of 64 samples to balance computational efficiency with stable gradient estimation.

We train using the entire training dataset and use the test dataset for validation. Early stopping is implemented with a patience of 30 epochs based on validation loss to prevent overfitting, complemented by learning rate reduction on plateau scheduling (factor=0.5, patience=20 epochs) to fine-tune convergence. Training is allowed to proceed for a maximum of 2000 epochs, though early stopping typically terminates training significantly earlier.

All computations are performed on Mac M3 Pro hardware utilizing Metal Performance Shaders (MPS) for GPU acceleration, with models operating in float32 precision to ensure MPS compatibility. Training the neural network is computationally inexpensive, taking around a minute on an Apple M3 Pro processor (with MPS). After training, the step involving identifying the preferred event can be performed instantaneously using the models stored on disk. This efficiency makes the neural network-based preferred event selector particularly well-suited for low-latency applications. We also expect the neural network to have two additional properties:

1. The event selected by the neural network should not be worse than the event selected using the present SNR-based algorithm.
2. The per pipeline fraction of events selected by the neural network should closely follow the distribution of events with the lowest searched area per pipeline.

In a realistic deployment scenario, we envision training the neural network on a set of simulated signals injected into real detector data, ideally from a stretch of time that closely mimics the noise and conditions of the observational run. The neural network, trained to recognize which events tend to have the lowest searched area, could then be applied to low-latency data during actual observations to select the most promising candidate for EM follow-up.

However, for the scope of this study, we evaluate the neural network on the MDC11 dataset, which comprises simulated CBC injections into archival detector data. As outlined in earlier sections, we split the dataset into training and testing subsets using a 9:1 ratio, and repeat this procedure across 10 different random seeds to quantify the variability and robustness of the model’s predictions. Each 9:1 split is referred to as an iteration in this work. This enables a thorough assessment of the model’s accuracy, generalization, and sensitivity to data splits, establishing a strong foundation for future deployment on real observational data.

V. RESULTS

We train 10 models on the 10 iterations of the significant dataset and 10 other models on 10 iterations of the datasets containing all triggers recovered at $\text{FAR} < 2/\text{day}$ (referred to as all triggers henceforth). Including all triggers increases the size of the training set by roughly 47%. The performance of each model is then evaluated on both datasets. Training models on both significant-only and all-event datasets allows for a direct comparison of their performance. The neural network takes ~ 40 iterations to learn sufficiently with a training loss ≤ 1.5 (≤ 1.25) and a validation loss of ≤ 1.8 (≤ 1.4) (see Figure 1) when trained on all (significant) events.

The absolute count of correctly identified g-events is not a reliable metric for evaluating performance. For superevents with only one associated g-event, all preferred event selectors are guaranteed to make the correct selection, thus inflating this number. To better assess the performance of various selection schemes, we examine the mean searched area for events obtained by various selectors (see left panel of Figure 2). Based on the means, we find that the SNR-based selector (mean searched area of 987 sq. deg. for significant events and 1135 sq. deg. for all events) performs worse than the FAR-based selector (mean searched

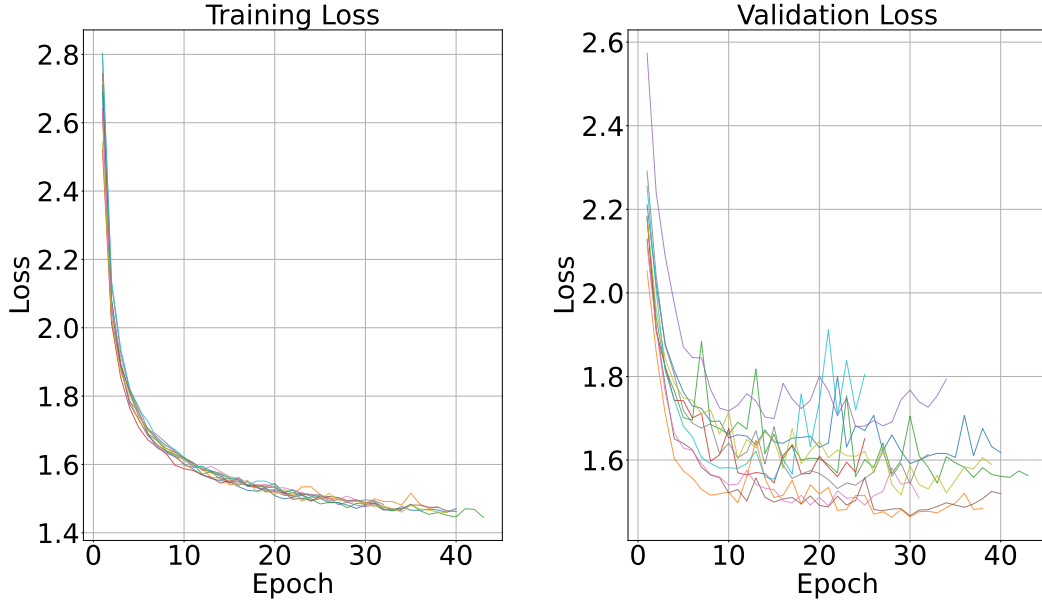


Figure 1: Left panel shows the training loss (trained on 90% of the events recovered in MDC11 at a FAR $< 2/\text{day}$) while the right panel shows the validation loss (evaluated on the remaining 10% of the events).

area of 973 sq. deg. for significant events and 1133 sq. deg. for all events). The FARs are computed using the SNR and additional information regarding the nature of the noise. Thus, the FAR-based selector has more information than the SNR-based selector and is expected to perform better. For comparison a selector randomly picking a g-event from the set of superevents has a mean searched area of 991 sq. deg. for significant events and 1141 sq. deg. for all events. It is worth noting that this random selector is biased towards pipelines that send out more g-events per superevent. The errorbars have been computed from the standard deviation of the means obtained from the 10 smaller datasets used to test the neural network's performance. As described earlier, we train two neural networks: one only on significant events and the other on all events. The neural network trained on significant events (mean 963 sq. deg.) performs better than traditional selectors when tested on significant events based on the mean. The neural network trained on all events (mean 1106 sq. deg.) performs better than traditional selectors when tested on significant events based on the mean. The fraction of events versus searched areas for various selectors are shown in Figure 3.

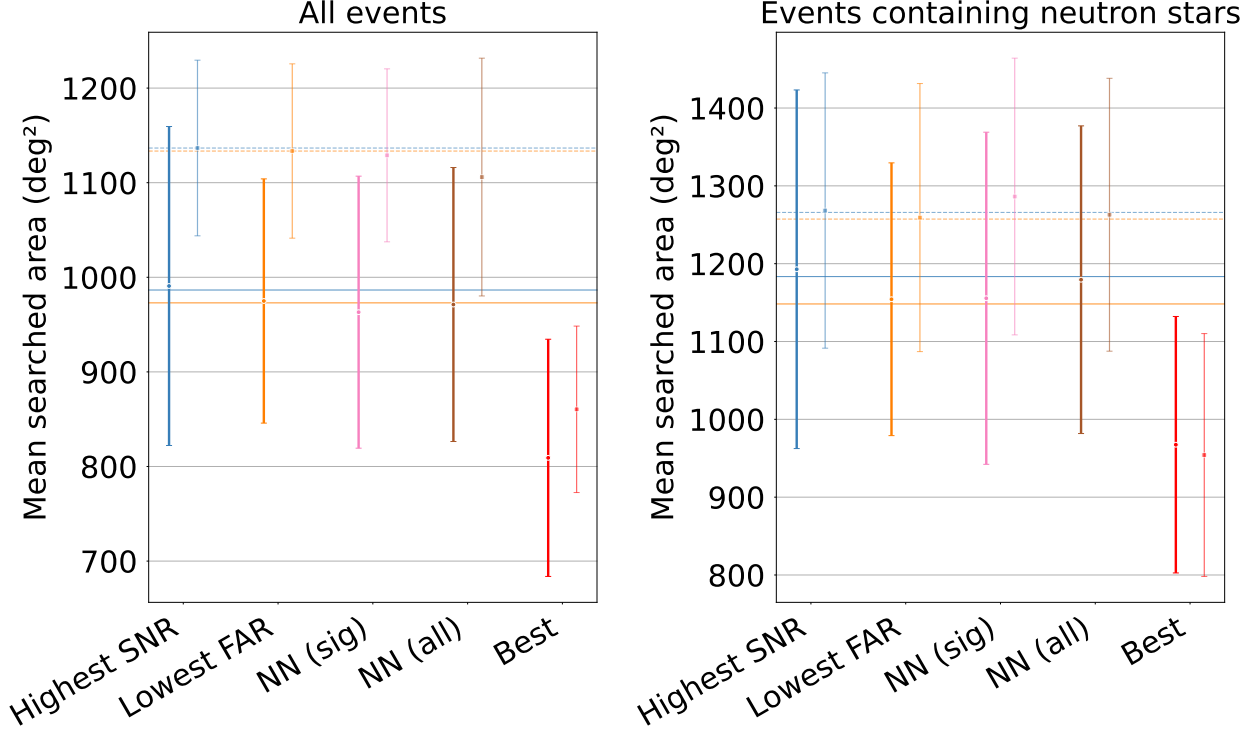


Figure 2: Plots showing the mean searched areas of the skymaps chosen for significant (darker and thicker lines) and all (lighter and thinner lines) events by different selection algorithms. NN (sig) refers to the neural network (NN) trained on significant events, and NN (all) refers to the neural network trained on all events. Best refers to the event with lowest searched area. The spread (1σ) in the mean searched area is due to the 10 test sets. The solid (dashed) line denotes the mean searched areas for the skymaps of significant (all) events chosen by highest SNR (blue) and lowest FAR (orange).

We further investigate the subset of events that have at least one component with an injected mass less than $2.05 M_{\odot}$ or sources with neutron stars. As seen from Figure 2, the trends are similar when we look specifically for events containing neutron stars. For an illustrative example, we look at S230309en an event containing a neutron star at MDC11. The highest SNR g-event, G927743, is recovered at a network SNR of 12.17 and has a searched area of 776 sq. deg. (see Figure 4a). G-event, G927745 is recovered at an SNR of 11.90 and has the lowest searched area (53 sq. deg.) (see Figure 4b). Both the neural networks choose G927726, which is recovered at an SNR of 11.43 and a searched area of 197 sq. deg. (see Figure 4c).

The developed neural network-based selector, however, stands out in its representation

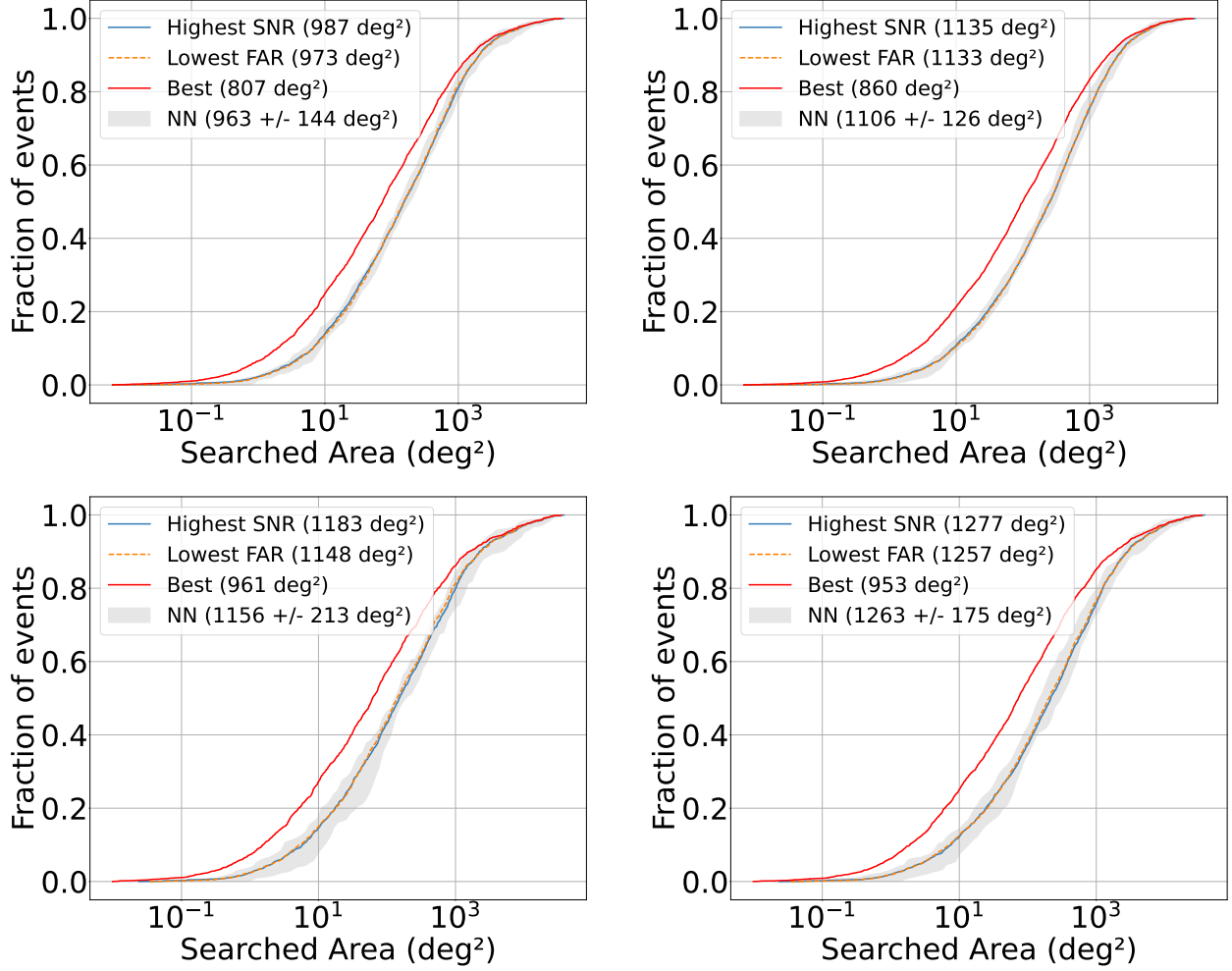
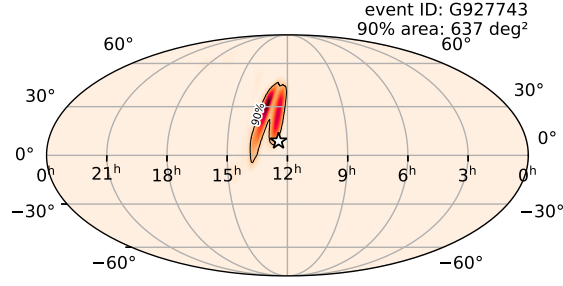
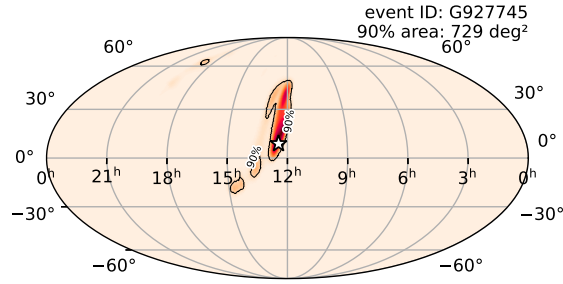


Figure 3: The left (right) column comprises of the cumulative distribution of searched areas for significant (all) g-event skymaps selected by lowest FAR, by highest SNR, by lowest searched area, and the neural network (NN) trained on significant (all) events. The median searched areas have been reported in parentheses. The bands have been obtained by evaluating different preferred event selectors on the 10 test datasets. The first panel comprises of all source types and the second panel comprises of sources with neutron stars.

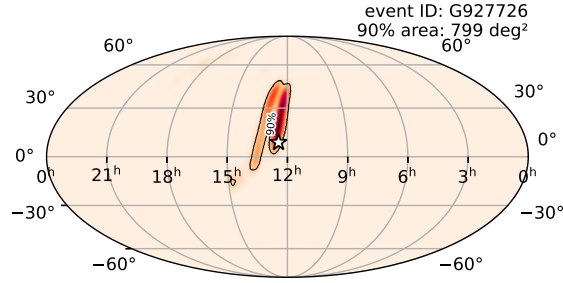
of pipelines across the selected events as shown in Figure 5. Since MDC11 injections have a known ground truth, we can quantify how often each pipeline recovers the lowest searched-area event. Both the highest-SNR and lowest-FAR selectors exhibit strong biases toward specific pipelines, deviating from the true underlying distribution. The neural network-based selector, which is pipeline-agnostic by design, in contrast, closely follows the underlying distribution while selecting events. This is crucial in realistic applications to ensure equitable



(a) The highest SNR event



(b) The lowest searched area event



(c) The NN selected event

Figure 4: Skymaps associated with various g-events corresponding to S230309en. The highest SNR event, G927743, has an SNR of 12.17 and a searched area of 776 sq. deg. G927726, selected by a neural network, has an SNR of 11.43 and a smaller searched area of 197 sq. deg. The event with the best localization, G927745, is recovered at an SNR of 11.90 and has a searched area of 53 sq. deg.

treatment of all low-latency pipelines.

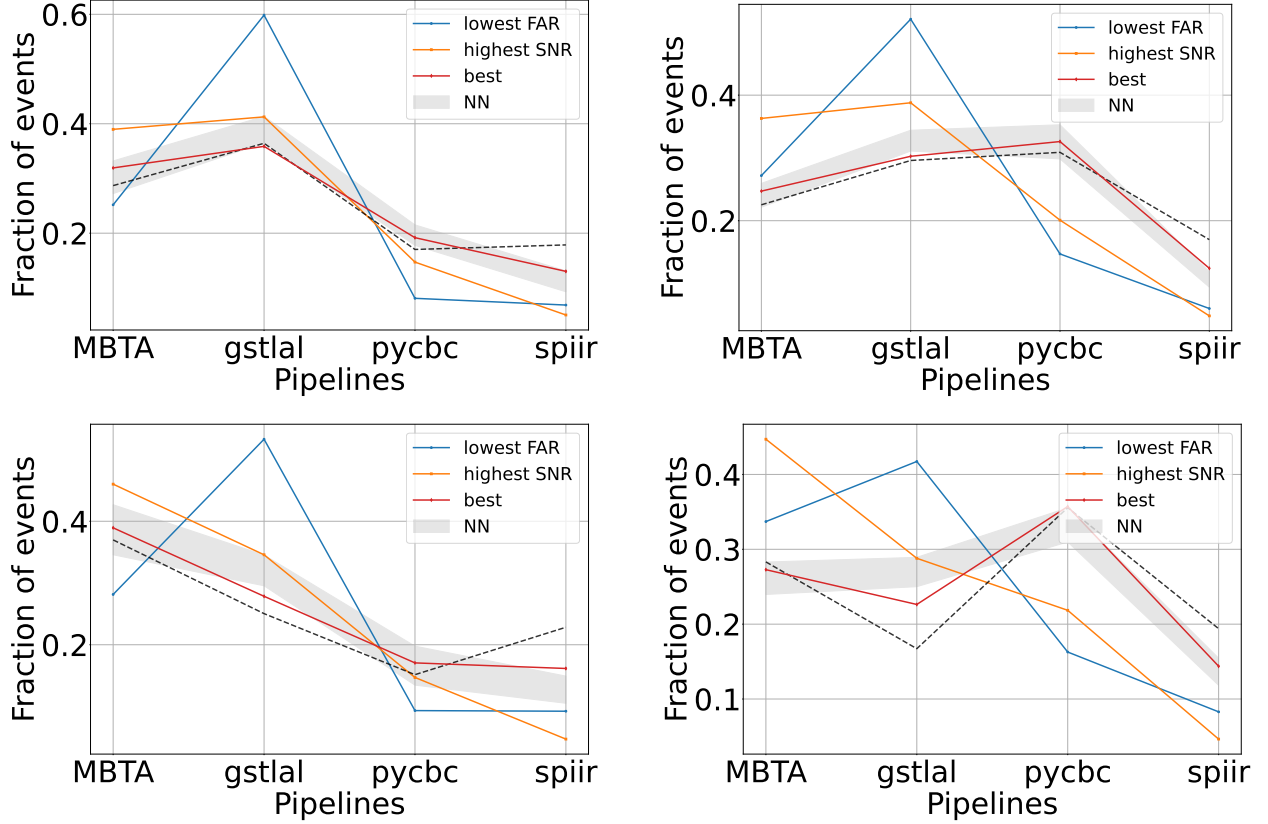


Figure 5: Fraction of selected g-events per pipeline for different selection choices. Those labeled best correspond to the g-event with the lowest searched area determined, i.e. the ground truth. The left column has significant events, while the right column has all events.

The first panel comprises of all source types and the second panel comprises of sources with neutron stars. The dashed line shows the average fraction of g-events per superevent recovered by each pipeline. This is equivalent to picking g-events at random from a superevent.

To summarize, we use MDC11 injections to demonstrate that a lowest-FAR-based selector performs slightly better than a highest SNR-based selector when all source classes are combined for MDC11 injections. Consequently, we develop a neural network-based selector that outperforms the traditional highest-SNR method based on the means. The neural network also preserves the distribution of pipeline contributions with respect to ground-truth injected events. Importantly, the neural network is computationally inexpensive to train and can select the preferred event instantaneously, making it well-suited for low-latency applications.

VI. DISCUSSIONS

This work shows that a neural network-based preferred event selector can serve as a more effective alternative to the current strategies employed in low-latency gravitational-wave follow-up. Our selector outperforms the traditional highest-SNR method, while preserving the distribution of pipeline contributions based on ground-truth injected events. Once trained, the neural network can identify the preferred event almost instantaneously, making it highly suitable for low-latency applications. Given these benefits, we conclude that a neural network-based approach is worthy of serious consideration for the next observing run (O5).

The conclusions in our study are entirely based on MDC11, which was conducted before the start of the current observing run. Since then, all pipelines have evolved significantly, with new features such as the SNR optimizers [65] being introduced, and additional pipelines like Aframe [84] and cwb-bbh [37] joining the search. Our results remain relevant, as the goal is not to compare pipeline performance but to demonstrate, as a proof of concept, that a neural network-based event selector can be deployed in low latency. Nonetheless, substantial work is still required before deployment in an observing run.

To enable realistic deployment, it is essential to train the neural network on data that more accurately reflects the state of the interferometers and the low-latency pipelines during an observing run. Fortunately, the lightweight nature of our neural network model makes retraining on a more representative dataset both straightforward and computationally inexpensive. Moreover, there is scope to enhance the model's performance by incorporating additional features. For sources observed in multiple detectors, incorporating additional information like the time delay and phase difference between detectors could potentially enhance the selection of the event with a lower searched area. These features capture aspects of the SNR time series that are sensitive to sky localization and may help the network more reliably identify the preferred event. Exploring such physically motivated features is left for future work.

Our neural network architecture arranges the g-events within each superevent in ascending order based on chirp mass. This is done to ensure that the network is invariant to the order in which g-events arrive. This setup ensures that the preferred event is determined only by detection pipeline outputs (SNR, FAR, and chirp mass). Nevertheless, this

restriction is not required. We observe that key statistical properties of the selection, including searched-area performance and pipeline representation, remain stable even when the g-events are ordered randomly. However, we would like our network to be permutation-invariant for reasons of reproducibility. This design choice reduces the effect of extraneous factors such as data transmission latencies, computational resource allocation, or cluster processing schedules.

Importantly, our neural network is pipeline-agnostic, intentionally omitting the identity of the pipeline that produced each g-event. Although one might expect that including such information could enhance performance, given the inherent differences in FAR and SNR distributions across pipelines, our experiments found no significant performance gains from incorporating pipeline identity. This suggests that our decision to exclude pipeline-specific information does not compromise the science case of this paper.

One question that naturally arises here is how the neural network preserves the per-pipeline fraction of preferred events. It is worth noting that the minimum FAR selector and the maximum SNR selector are biased towards specific pipelines. The neural network is pipeline-agnostic, and so it tends to choose randomly between g-events from different pipelines when it assesses their skymaps to be comparably accurate. This makes the neural network follow some combination of the random selector (represented by dashed lines in Figure 5) and the lowest searched area selector (represented by a solid red line in Figure 5).

Finally, our analysis underscores the limitations of the current SNR-based preferred event selection strategy. This shortcoming might arise from the presence of real, non-Gaussian noise in the data, which can distort the SNRs. This may also arise from differences in how the pipelines compute SNRs internally or estimate the noise power spectral density. While selecting events based on the lowest FAR yields better performance in terms of identifying the event with the lowest searched area, it introduces a bias in pipeline representation, favoring some pipelines over others. Our neural network-based approach achieves a better balance: it is similar in performance to a lowest FAR selector in terms of searched area while representing all pipelines and offering low-latency applicability.

Data and code availability : The codes for this project are hosted in https://git.ligo.org/pratyusava.baral/preferred_event_ml. The data used for this project will be made available via Zenodo.

VII. ACKNOWLEDGEMENT

The authors would like to thank Deep Chatterjee for providing valuable feedback during the internal LIGO review. PB would like to thank Gaurav Waratkar, Andrew Toivonen, and Sushant Sharma-Chaudhary for their assistance in calculating the searched area for skymaps corresponding to each g-event in MDC11. PB is also grateful to Soichiro Morisaki, Leo Singer, Leo Tsukada, Carl Haster, Jolien Creighton, Lucy Thomas, and Adrian Cornell for their valuable comments and suggestions, which helped improve this work. This material is based upon work supported by NSF’s LIGO Laboratory which is a major facility fully funded by the National Science Foundation. This work was supported partially by the National Science Foundation (NSF) awards NSF PHY-2513124, PHY-2207728, PHY-2110594 and PHY-2513358 and partially by Wisconsin Space Grant Consortium Awards RFP25_9-0. This work carries LIGO Document No. LIGO-P2500665.

Appendix A: Results without sorting by chirp mass

We train and test the same neural network on g-events that are randomly organized, rather than being sorted by chirp mass. The results remain broadly consistent (see Figure 6 and Figure 7), confirming that the chirp mass sorting primarily serves to preserve permutation invariance and does not impact the scientific outcomes.

-
- [1] LIGO Scientific Collaboration, Classical and Quantum Gravity **32**, 074001 (2015), arXiv:1411.4547 [gr-qc].
 - [2] F. Acernese, M. Agathos, K. Agatsuma, *et al.*, Classical and Quantum Gravity **32**, 024001 (2015), arXiv:1408.3978 [gr-qc].
 - [3] T. Akutsu, M. Ando, K. Arai, *et al.*, Progress of Theoretical and Experimental Physics **2021**, 05A101 (2021), arXiv:2005.05574 [physics.ins-det].
 - [4] The LVK Collaboration, arXiv e-prints , arXiv:2508.18082 (2025), arXiv:2508.18082 [gr-qc].
 - [5] B. P. Abbott, R. Abbott, T. D. Abbott, *et al.*, ApJ **826**, L13 (2016), arXiv:1602.08492 [astro-ph.HE].

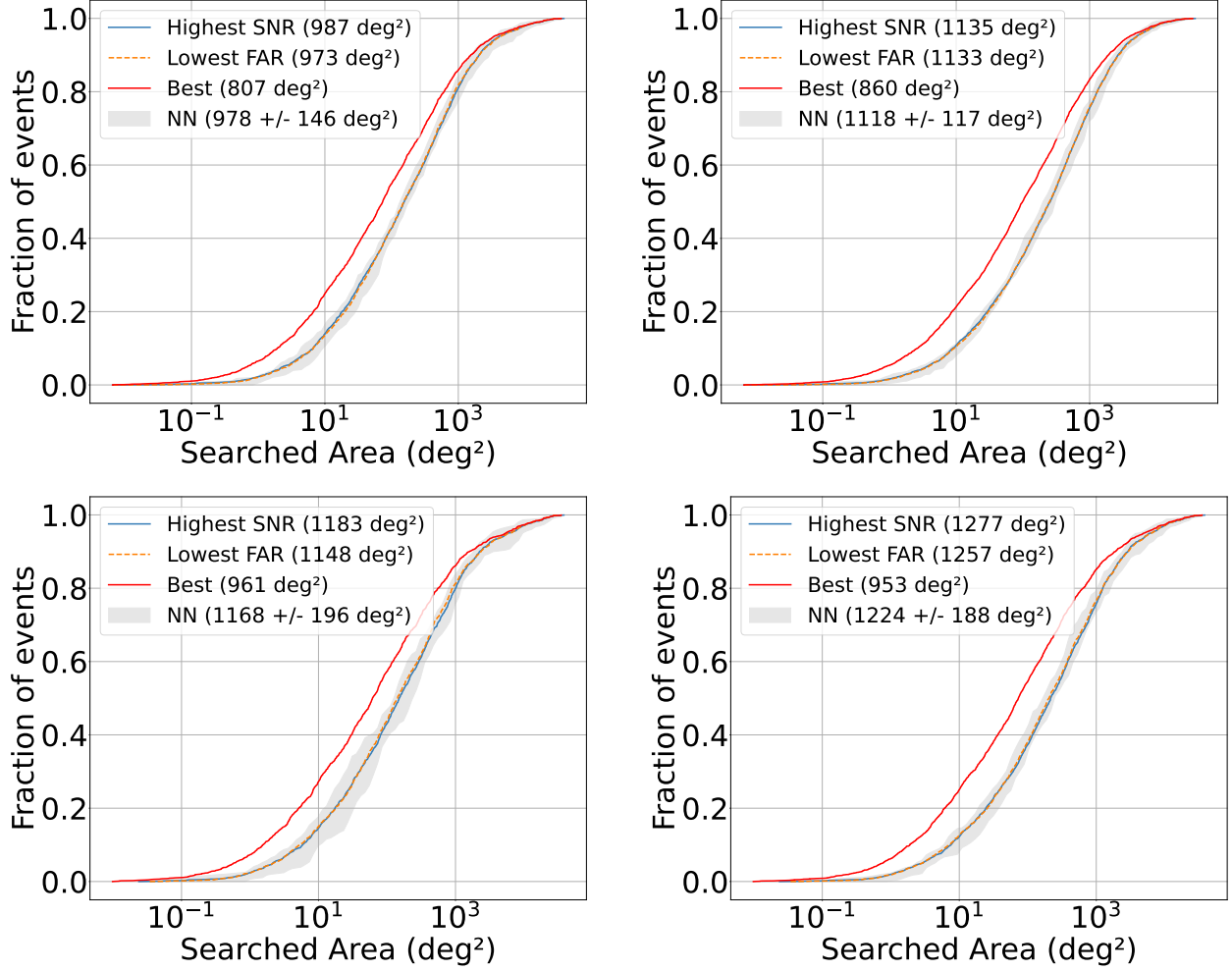


Figure 6: The left (right) column comprises of the cumulative distribution of searched areas for significant (all) g-event skymaps selected by lowest FAR, by highest SNR, by lowest searched area, and the neural network (NN) trained on significant (all) events.

Unlike figure 3 we do not sort by chirp mass. The median searched areas have been reported in parentheses. The bands have been obtained by evaluating different preferred event selectors on the 10 test datasets. The first panel comprises of all source types and the second panel comprises of sources with neutron stars.

[6] B. P. Abbott, R. Abbott, T. D. Abbott, *et al.*, Phys. Rev. Lett. **119**, 161101 (2017), arXiv:1710.05832 [gr-qc].

[7] R. Abbott, T. D. Abbott, S. Abraham, *et al.*, ApJ **915**, L5 (2021), arXiv:2106.15163 [astro-ph.HE].

[8] <https://observing.docs.ligo.org/plan>.

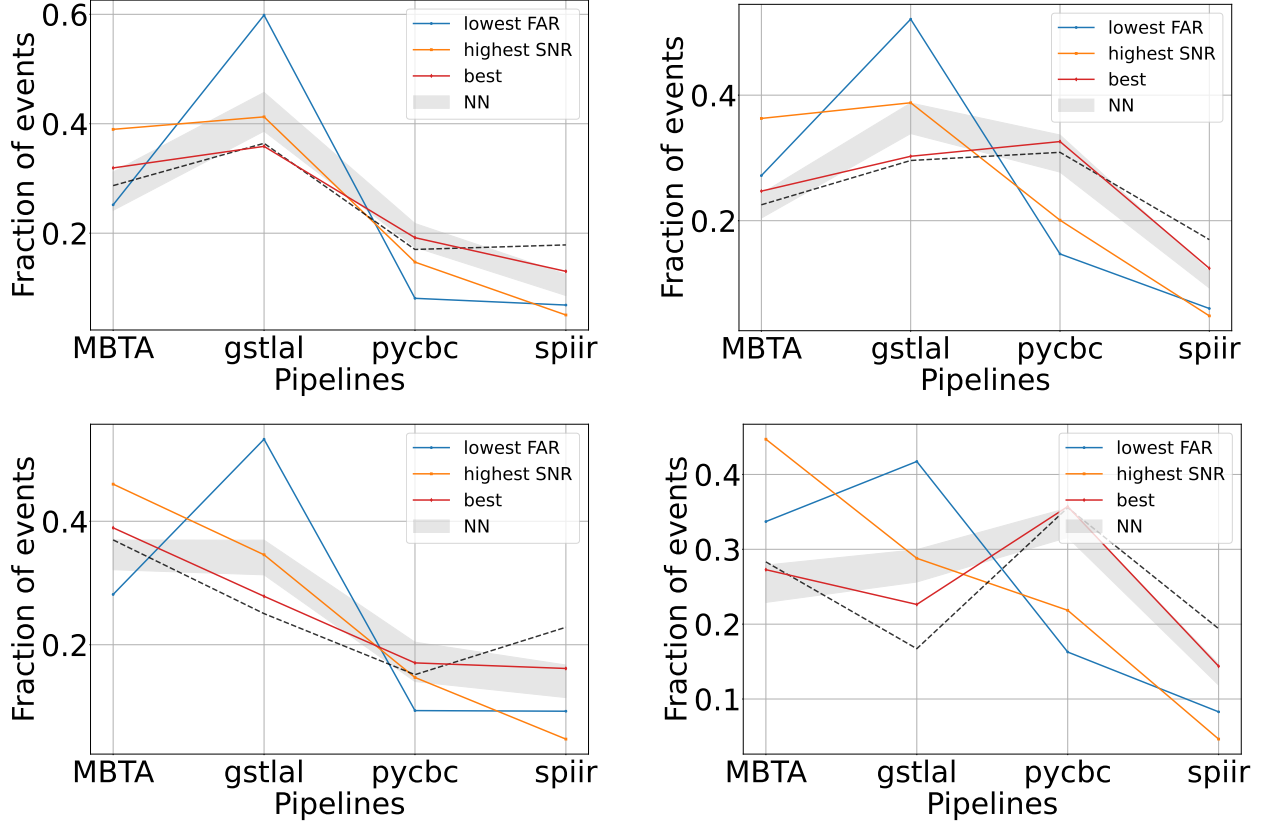


Figure 7: Fraction of selected g-events per pipeline for different selection choices. Unlike figure 5 we do not sort by chirp mass. Those labeled best correspond to the g-event with the lowest searched area determined, i.e. the ground truth. The left column has significant events, while the right column has all events. The first panel comprises of all source types and the second panel comprises of sources with neutron stars. The dashed line shows the average fraction of g-events per superevent recovered by each pipeline. This is equivalent to picking g-events at random from a superevent.

- [9] A. Goldstein, P. Veres, E. Burns, *et al.*, ApJ **848**, L14 (2017), arXiv:1710.05446 [astro-ph.HE].
- [10] V. Savchenko, C. Ferrigno, E. Kuulkers, *et al.*, ApJ **848**, L15 (2017), arXiv:1710.05449 [astro-ph.HE].
- [11] B. P. Abbott, R. Abbott, T. D. Abbott, *et al.*, ApJ **848**, L13 (2017), arXiv:1710.05834 [astro-ph.HE].
- [12] D. A. Coulter, R. J. Foley, C. D. Kilpatrick, *et al.*, Science **358**, 1556 (2017), arXiv:1710.05452 [astro-ph.HE].

- [13] S. J. Smartt, T.-W. Chen, A. Jerkstrand, *et al.*, *Nature* **551**, 75 (2017), arXiv:1710.05841 [astro-ph.HE].
- [14] B. P. Abbott, R. Abbott, T. D. Abbott, *et al.*, *ApJ* **848**, L12 (2017), arXiv:1710.05833 [astro-ph.HE].
- [15] A. Bauswein, O. Just, H.-T. Janka, and N. Stergioulas, *ApJ* **850**, L34 (2017), arXiv:1710.06843 [astro-ph.HE].
- [16] B. Margalit and B. D. Metzger, *ApJ* **850**, L19 (2017), arXiv:1710.05938 [astro-ph.HE].
- [17] M. W. Coughlin, T. Dietrich, B. Margalit, *et al.*, *MNRAS* **489**, L91 (2019), arXiv:1812.04803 [astro-ph.HE].
- [18] M. W. Coughlin, T. Dietrich, Z. Doctor, *et al.*, *MNRAS* **480**, 3871 (2018), arXiv:1805.09371 [astro-ph.HE].
- [19] M. W. Coughlin, T. Dietrich, S. Antier, *et al.*, *MNRAS* **492**, 863 (2020), arXiv:1910.11246 [astro-ph.HE].
- [20] E. Annala, T. Gorda, A. Kurkela, *et al.*, *Phys. Rev. Lett.* **120**, 172703 (2018), arXiv:1711.02644 [astro-ph.HE].
- [21] E. R. Most, L. R. Weih, L. Rezzolla, *et al.*, *Phys. Rev. Lett.* **120**, 261103 (2018).
- [22] D. Radice, A. Perego, F. Zappa, *et al.*, *ApJ* **852**, L29 (2018), arXiv:1711.03647 [astro-ph.HE].
- [23] X. Lai, E. Zhou, and R. Xu, *European Physical Journal A* **55**, 60 (2019), arXiv:1811.00193 [astro-ph.HE].
- [24] T. Dietrich, M. W. Coughlin, P. T. H. Pang, *et al.*, *Science* **370**, 1450 (2020), arXiv:2002.11355 [astro-ph.HE].
- [25] S. Huth, P. T. H. Pang, I. Tews, *et al.*, *Nature* **606**, 276 (2022), arXiv:2107.06229 [nucl-th].
- [26] M. W. Coughlin, T. Dietrich, J. Heinzl, *et al.*, *Physical Review Research* **2**, 022006 (2020), arXiv:1908.00889 [astro-ph.HE].
- [27] M. W. Coughlin, S. Antier, T. Dietrich, *et al.*, *Nature Communications* **11**, 4129 (2020), arXiv:2008.07420 [astro-ph.HE].
- [28] B. P. Abbott, R. Abbott, T. D. Abbott, *et al.*, *Nature* **551**, 85 (2017), arXiv:1710.05835.
- [29] K. Hotokezaka, E. Nakar, O. Gottlieb, *et al.*, *Nature Astronomy* **3**, 940 (2019), arXiv:1806.10596 [astro-ph.CO].
- [30] R. Chornock, E. Berger, D. Kasen, *et al.*, *ApJ* **848**, L19 (2017), arXiv:1710.05454 [astro-ph.HE].

- [31] D. A. Coulter, R. J. Foley, C. D. Kilpatrick, *et al.*, Science **358**, 1556 (2017), arXiv:1710.05452 [astro-ph.HE].
- [32] P. S. Cowperthwaite, E. Berger, V. A. Villar, *et al.*, ApJ **848**, L17 (2017), arXiv:1710.05840 [astro-ph.HE].
- [33] S. Rosswog, U. Feindt, O. Korobkin, *et al.*, Classical and Quantum Gravity **34**, 104001 (2017), arXiv:1611.09822 [astro-ph.HE].
- [34] D. Watson, C. J. Hansen, J. Selsing, *et al.*, Nature **574**, 497 (2019), arXiv:1910.10510 [astro-ph.HE].
- [35] M. M. Kasliwal, D. Kasen, R. M. Lau, *et al.*, MNRAS **510**, L7 (2022), arXiv:1812.08708 [astro-ph.HE].
- [36] The LVK Collaboration, arXiv e-prints , arXiv:2508.18080 (2025), arXiv:2508.18080 [gr-qc].
- [37] The LVK Collaboration, arXiv e-prints , arXiv:2508.18081 (2025), arXiv:2508.18081 [gr-qc].
- [38] R. Abbott, T. D. Abbott, F. Acernese, *et al.*, Physical Review X **13**, 041039 (2023), arXiv:2111.03606 [gr-qc].
- [39] The LVK Collaboration, Phys. Rev. D **104**, 122004 (2021), arXiv:2107.03701 [gr-qc].
- [40] C. Messick, K. Blackburn, P. Brady, *et al.*, Phys. Rev. D **95**, 042001 (2017), arXiv:1604.04324 [astro-ph.IM].
- [41] F. Aubin, F. Brighenti, R. Chierici, *et al.*, Classical and Quantum Gravity **38**, 095004 (2021), arXiv:2012.11512 [gr-qc].
- [42] S. Hooper, S. K. Chung, J. Luan, D. Blair, Y. Chen, and L. Wen, Phys. Rev. D **86**, 024012 (2012), arXiv:1108.3186 [gr-qc].
- [43] J. Luan, S. Hooper, L. Wen, *et al.*, Phys. Rev. D **85**, 102002 (2012), arXiv:1108.3174 [gr-qc].
- [44] S. A. Usman, A. H. Nitz, I. W. Harry, *et al.*, Classical and Quantum Gravity **33**, 215004 (2016), arXiv:1508.02357 [gr-qc].
- [45] T. Dal Canton, A. H. Nitz, B. Gadre, *et al.*, ApJ **923**, 254 (2021), arXiv:2008.07494 [astro-ph.HE].
- [46] B. J. Piotrkowski, *Searching for Gravitational Wave Associations with High-Energy Astrophysical Transients*, Ph.D. thesis, University of Wisconsin-Milwaukee, Milwaukee, WI (2022), available at <http://digital.library.wisc.edu/1793/93071>.
- [47] <https://git.ligo.org/emfollow/gwcelery>.
- [48] <https://rtd.igwn.org/projects/gwcelery/en/latest/>.

- [49] <https://gracedb.ligo.org/>.
- [50] S. S. Chaudhary, A. Toivonen, G. Waratkar, *et al.*, Proceedings of the National Academy of Science **121**, e2316474121 (2024), arXiv:2308.04545 [astro-ph.HE].
- [51] <https://emfollow.docs.ligo.org/userguide/>.
- [52] L. P. Singer and L. R. Price, Phys. Rev. D **93**, 024013 (2016), arXiv:1508.03634 [gr-qc].
- [53] J. Creighton and W. Anderson, Gravitational-wave physics and astronomy: An introduction to theory, experiment and data analysis. (John Wiley & Sons, Ltd, 2011) Chap. 7, pp. 269–347, <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9783527636037.ch7>.
- [54] E. Chabanat, P. Bonche, P. Haensel, *et al.*, Nucl. Phys. A **635**, 231 (1998).
- [55] Planck Collaboration, A&A **652**, C4 (2021).
- [56] T. Adams, D. Buskulic, V. Germain, *et al.*, Classical and Quantum Gravity **33**, 175012 (2016), arXiv:1512.02864 [gr-qc].
- [57] C. All  n  , F. Aubin, I. Bentara, *et al.*, Classical and Quantum Gravity **42**, 105009 (2025), arXiv:2501.04598 [gr-qc].
- [58] S. Sachdev, S. Caudill, H. Fong, *et al.*, arXiv e-prints , arXiv:1901.08580 (2019), arXiv:1901.08580 [gr-qc].
- [59] C. Hanna, S. Caudill, C. Messick, *et al.*, Phys. Rev. D **101**, 022003 (2020), arXiv:1901.02227 [gr-qc].
- [60] K. Cannon, S. Caudill, C. Chan, *et al.*, SoftwareX **14**, 100680 (2021), arXiv:2010.05082 [astro-ph.IM].
- [61] S. Sakon, L. Tsukada, H. Fong, J. Kennington, and *et al.*, Phys. Rev. D **109**, 044066 (2024), arXiv:2211.16674 [gr-qc].
- [62] B. Ewing, R. Huxford, D. Singh, L. Tsukada, *et al.*, Phys. Rev. D **109**, 042008 (2024), arXiv:2305.05625 [gr-qc].
- [63] L. Tsukada, P. Joshi, S. Adhicary, *et al.*, Phys. Rev. D **108**, 043004 (2023), arXiv:2305.06286 [astro-ph.IM].
- [64] A. Ray, W. Niu, S. Sakon, B. Ewing, *et al.*, arXiv e-prints , arXiv:2306.07190 (2023), arXiv:2306.07190 [gr-qc].
- [65] P. Joshi, L. Tsukada, C. Hanna, *et al.*, arXiv e-prints , arXiv:2506.06497 (2025), arXiv:2506.06497 [gr-qc].

- [66] P. Joshi, W. Niu, C. Hanna, *et al.*, arXiv e-prints , arXiv:2505.23959 (2025), arXiv:2505.23959 [gr-qc].
- [67] B. Allen, Phys. Rev. D **71**, 062001 (2005), arXiv:gr-qc/0405045 [gr-qc].
- [68] S. A. Usman, A. H. Nitz, I. W. Harry, *et al.*, Classical and Quantum Gravity **33**, 215004 (2016), arXiv:1508.02357 [gr-qc].
- [69] A. H. Nitz, Classical and Quantum Gravity **35**, 035016 (2018), arXiv:1709.08974 [gr-qc].
- [70] T. Dal Canton, A. H. Nitz, B. Gadre, *et al.*, ApJ **923**, 254 (2021), arXiv:2008.07494 [astro-ph.HE].
- [71] G. S. Davies, T. Dent, M. Tápai, *et al.*, Phys. Rev. D **102**, 022004 (2020), arXiv:2002.08291 [astro-ph.HE].
- [72] S. Klimentko, I. Yakushin, A. Mercer, *et al.*, Classical and Quantum Gravity **25**, 114029 (2008), arXiv:0802.3232 [gr-qc].
- [73] S. Klimentko, G. Vedovato, M. Drago, *et al.*, Phys. Rev. D **93**, 042004 (2016), arXiv:1511.05999 [gr-qc].
- [74] M. Drago, S. Klimentko, C. Lazzaro, *et al.*, SoftwareX **14**, 100678 (2021), arXiv:2006.12604 [gr-qc].
- [75] R. Lynch, S. Vitale, R. Essick, E. Katsavounidis, and F. Robinet, Phys. Rev. D **95**, 104046 (2017), arXiv:1511.05955 [gr-qc].
- [76] A. L. Urban, *Monsters in the Dark: High Energy Signatures of Black Hole Formation with Multimessenger Astronomy*, Ph.D. thesis, University of Wisconsin, Milwaukee (2016).
- [77] M.-A. Cho, *Low-Latency Searches for Gravitational Waves and Their Electromagnetic Counterparts with Advanced LIGO and Virgo*, Ph.D. thesis, University of Maryland (2019).
- [78] <https://emfollow.docs.ligo.org/userguide/>.
- [79] S. Ioffe and C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift (2015), arXiv:1502.03167 [cs.LG].
- [80] <https://docs.pytorch.org/docs/stable/generated/torch.nn.CrossEntropyLoss.html>.
- [81] G. E. Hinton, N. Srivastava, A. Krizhevsky, *et al.*, Improving neural networks by preventing co-adaptation of feature detectors (2012), arXiv:1207.0580 [cs.NE].
- [82] X. Glorot and Y. Bengio, in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, Proceedings of Machine Learning Research, Vol. 9, edited by Y. W. Teh and M. Titterton (PMLR, Chia Laguna Resort, Sardinia, Italy, 2010) pp. 249–256.

- [83] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization (2017), arXiv:1412.6980 [cs.LG].
- [84] E. Marx, W. Benoit, A. Gunny, *et al.*, Phys. Rev. D **111**, 042010 (2025), arXiv:2403.18661 [gr-qc].