Using supervised learning algorithms as a follow-up method in the search of gravitational waves from core-collapse supernovae

Javier M. Antelis,[1, 2, a] Marco Cavaglia,[3] Travis Hansen,[1] Manuel D. Morales,[4]
Claudia Moreno,[1, 4] Soma Mukherjee,[5] Marek J. Szczepańczyk,[6] and Michele Zanolin[1]

[1] *Embry-Riddle Aeronautical University, Prescott, AZ 86301, USA*
[2] *Tecnologico de Monterrey, Escuela de Ingeniería y Ciencias, Monterrey, N.L., 64849, México*
[3] *Institute of Multi-messenger Astrophysics and Cosmology,*
Missouri University of Science and Technology, Rolla, MO 65409, USA
[4] *Universidad de Guadalajara, Guadalajara, Jal., 44430, México*
[5] *The University of Texas Rio Grande Valley, Brownsville, TX 78520, USA*
[6] *University of Florida, Gainesville, FL 32611, USA*

We present a follow-up method based on supervised machine learning (ML) to improve the performance in the search of gravitational wave (GW) burts from core-collapse supernovae (CCSNe) using the coherent WaveBurst (cWB) pipeline. The ML model discriminates noise from signal events using as features a set of reconstruction parameters provided by cWB. Detected noise events are discarded yielding to a reduction of the false alarm rate (FAR) and of the false alarm probability (FAP) thus enhancing of the statistical significance. We tested the proposed method using strain data from the first half of the third observing run of advanced LIGO, and CCSNe GW signals extracted from 3D simulations. The ML model is learned using a dataset of noise and signal events, and then it is used to identify and discard noise events in cWB analyses. Noise and signal reduction levels were examined in single detector networks (L1 and H1) and two detector network (L1H1). The FAR was reduced by a factor of ~ 10 to ~ 100, there was an enhancement in the statistical significance of ~ 1 to $\sim 2\sigma$, while there was no impact in detection efficiencies.

I. INTRODUCTION

The search and characterization of gravitational wave (GW) bursts with the network of laser interferometers LIGO [1], VIRGO [2] and KAGRA [3] need to address issues such as discrimination between GW events and noise artifacts, reconstruction of the GW waveforms, and localization of the source in the sky. In the case of GWs generated by binary black holes (BBH) and binary neutron stars (BNS), the existing algorithms benefits from having highly deterministic signal models, and thus, searches are based on match-filtering detector strain data with available template signals (see for example [4, 5]).

Core-collapse supernovae (CCSNe) are also a primary detection target in the upcoming LIGO and VIRGO observing runs. CCSNe are of special interest because the electromagnetic radiation and emission of neutrinos along with GWs will provide new hints to understand their formation mechanism and dynamic, and also will lead to novel insights in multimessenger astronomy. The morphology of GWs from CCSNe is predominantly stochastic with some deterministic components like the growing trend of the central frequency of the fundamental g-f mode or the characteristic waveform during the core bounce emission in rapidly rotating progenitors. Furthermore, the production of templates from three-dimensional (3D) state-of-the-art numerical simulations is computational extremely expensive, including the fact that multiple run simulations for same progenitor might

produce slightly different morphologies, even if this limitation does not apply to the core bounce component of rapidly rotating progenitors that could be described by deterministic template banks [6]. These limitations are the main reasons why unmodeled or weakly modeled searches are so far the only approaches considered for CCSNe GWs [7, 8].

Coherent WaveBurst (cWB) is a standard method for the search and characterization of GWs using data collected by the LIGO and VIRGO detectors without imposing assumptions about the signal morphology [9]. To detect unmodelled GW transients, cWB identifies coincident excess-power between the strain data in the network of detectors using wavelet-based analyses [10]. Such coincidence between multiple detectors allows to identify and reject noise events since they are present in only one of the detectors. The method also reconstructs the waveform of the detected GW and estimates some signal parameters [11–13]. cWB has been an essential computational tool in the detection of many GW transients from binary systems discovered so far [14–18], and has been a core method in all-sky and targeted searches of burst GW signals in the previous LIGO and VIRGO observing runs [7, 19–22]. Furthermore, cWB will play a critical role in the detection of CCSNe GWs in the upcoming fourth and fifth observing runs (O4 and O5) with the LIGO, VIRGO, and KAGRA (LVK) network [8].

The search of GWs emitted by CCSNe possesses other difficulties in addition to the inherent uncertainties in the waveform models. The production of candidate events is mostly driven by excess energy instances that can be related to stationary or non-stationary noise components. For small signal-to-noise ratio (SNR) the station-

arXiv:2111.07219v1 [gr-qc] 14 Nov 2021

ary ones produce most of the events, while non-stationary ones (usually termed non-Gaussian glitches) dominate the large SNR values. Morphologically both type of noise events tend to carry specific signatures of the physical causes that generated them. The rate of more rare glitches has been largely reduced (see for example ref. [7]) by requiring a temporal coincidence and some morphological similarity between different interferometers (as encoded in the so-called *cc* coefficient of cWB). However, if we have only one interferometer collecting data, temporal coincidence and morphological consistency are not available. But even in the case of a network of detectors, the population of non-Gaussian glitches can have a rate sufficiently large to be a limiting factor in obtaining a large statistical confidence in the detection.

The events formed by cWB at small SNR tend to show a fairly compact time frequency structure that would be independent of the sources, however for increasing amplitudes the morphology can become more varied. The detector characterization group of the LVK network operate routinely to classify the morphology of different glitches [23] and some of them, like the so-called blip glitches can strongly resemble the core bounce waveform for rapidly rotating progenitors [8]. Even if cWB makes use of internal morphological tests with multiple interferometers (especially for extragalactic source locations), all the opportunities to improve either the detection confidence, the detection range or both need to be taken. These issues encourage to identify more morphological metrics to distinguish CCSNe events from noise events, especially for loud glitches that can reduce the statistical confidence of a detection even when they are not temporally overlapped to the detection candidate. Machine learning (ML) methods offers one of these opportunities and this paper is part of systematic the explorations of its potential.

In recent times, ML along with the special class of deep learning (DL) models have attracted a lot of attention to tackle several problems in the realm of GWs [24], for instance, to discriminate between noise and GW signals either from binary systems [25–29] or from CCSNe [30–33], and to identify and remove transient noise events using strain data or auxiliary channels [34–37]. Notably, ML models also have been used to enhance cWB performance, in specific, to distinguish between glitches and GW signals from BBH [38], to construct a statistical veto based on the recognition of noise events to improve the detection efficiencies of GWs from BBH [39], and to achieve higher detection sensitivity of GW signals from CCSNe using signal enhancement [40, 41]. ML models, in specific genetic programming algorithms, had been previously investigated to discriminate CCSNe GW signals from noise transients for the case of single detector searches [42]. This is significant because it has been shown that there are periods of time during the observing runs for which only one detector in the network is in operational conditions and collecting science-quality data [42]. Indeed, ~30% of the collected strain data dur-

ing the first observing runs has been from one detector only [43]. Along this line, this paper investigates the benefits of supervised ML as a follow-up method of cWB that discriminates between noise and signal events in searches of GWs from CCSNe. We extended the previous approach [42] by investigating the noise reduction in two detector network and assessing its effects in the detection range for extragalactic CCSNe; using classification algorithms that rely on a discriminant function that do not require stochastic initialization of parameters and iterative executions thus avoiding to perform several training repetitions or the selection of a sub-optimal solution, as in the case of classifiers based on genetic approaches.

To discriminate between noise and signal triggers, we used as features a set of cWB reconstruction parameters such as duration and central frequency, and the supervised classification models linear discriminant analysis (LDA) and support vector machines (SVM). Detected noise events are discarded, thus reducing the rate of false detections. We carried out systematic studies to assess to quantify the effective improvement in the false alarm rate (FAR) and in the statistical significance with networks of one or two detectors using strain data from the O3a observing run, and with several recent CCSNe GW waveforms from 3D simulations. These studies considered robustness tests where the ML is tuned with a distribution of noise and signal triggers extracted from a stretch of data and then it is applied and tested in a different stretch of data. The manuscript is organized as follows. Section II describes the cWB analyses carried out to generate distributions of noise and signals triggers and the LDA and SVM classification models used to recognize between noise and CCSNe GW signal events. Section III presents the results of the studies devoted, first to assess the noise reduction and signal lost rates produced by the classification models, and second to ascertain the actual improvement in the FAR and the statistical significance. Section IV presents conclusions and future directions.

II. METHODS

A. cWB analysis

In this study we used LIGO strain data from the first half of the third observing run (O3a). cWB analyses were carried out independently in three stretches of open data (i.e., three time windows named TW1, TW2 and TW3) with a duration of three days each. For the case of CCSNe GWs, the search is typically driven by optical observations where the time window ranges from hours to days, while it is a few seconds in the potential case of neutrino flux driven searches. Hence, three days was chosen as a representative window duration of how much we would be able to constrain the GW emission from electromagnetic observations and a relatively rapid discovery of the CCSNe. All cWB analyses were carried out separatelly with one and two detectors network (L1, H1, and

TABLE I. Total time and the background (non-zero lag) time of each network of detectors (L1, H1 and L1H1) for the three stretches of open O3a LIGO data.

Time window	TW1		TW2		TW3	
Initial time	2019-06-17T00:00:01		2019-06-25T00:00:01		2019-08-26T00:00:01	
Final time	2019-06-19T23:59:59		2019-06-27T23:59:59		2019-08-28T23:59:59	
Network	Total time	Background time	Total time	Background time	Total time	Background time
L1	1.73 days	-	1.62 days	-	2.18 days	-
H1	1.77 days	-	1.66 days	-	2.38 days	-
L1H1	1.50 days	4.6 years	1.06 days	3.3 years	1.96 days	6.2 years

L1H1) with the aim of studying the rate of false detection (background analysis), the detectability of CCSNe GWs (sensitivity analysis), and to generate datasets of noise and signal events to train and to test the classification algorithms. Table I shows the total time of the three time windows for each network of detectors used in this study.

1. cWB Background

With more that one detector, the data from one detector is shifted with respect to the other in a time length that has to be longer than the GW travel time between detectors (∼10ms between L1 and H1). This time-shift procedure is repeated multiple times to obtain a total background search time (also called non-zero lag time) long enough to attach a certain statistical significance. All detected events from this analysis are of non-astrophysical origin and therefore correspond to false detections or noise events. Then, the FAR can be estimated simply as the ratio between the number of detected events and the total background time. Table I also shows the total background time of the three time windows considered in this study. Note that with only one detector available in the network the time-shifting is not possible, neither coincident test to remove glitches can be performed. This drastically worsen the FAR in comparison with networks of two or more operational detectors.

In the cWB background analysis, the FAR is produced for different values of the network SNR or ρ, yielding to the FAR versus ρ curves. Note that for the case of a single interferometer, ρ is equivalent to the SNR of the available detector. The statistical significance is computed in terms of the false alarm probability (FAP) as follows [7]:

$$FAP = 1 - e^{T_{on} \times FAR}, \tag{1}$$

where T_{on} is the on-source window where a GW signal is searched for.

Results for the background analysis in the time window TW1 are presented in figure 1. Note how the FAR (figure 1a) is considerably higher with one detector than with two detectors. Likewise, the corresponding values of the FAP (figure 1b, where the ρ threshold was set to 5) for the case of a search in a on-source window of 1s

reveals a statistical significance barely close to 3σ, and a quite higher significance close to 5σ for the two detector network. This illustrates the need of a follow-up method to identify and discard noise events (not only in single detector based searches but also in searches with two detectors) which reduces the FAR and improves the statistical significance of the search.

2. cWB sensitivity

The goal in this analysis is to determine the sensitivity in the search of known CCSNe GWs, and to obtain a distribution of signal events. Here, waveforms are systematically added into detector noise data, rescaled to different amplitudes corresponding to diverse source distances, and at different time delays between detectors corresponding to diverse source sky locations. For the case of single detector network, the injection is carried out only in the available detector and thus no time delay is required. Subsequently, the search of GWs is carried out and the detection efficiency (DE) is simply measured as the fraction of successfully detected GWs. DE is computed for each injected type of GW and for each distance to construct DE versus source distance curves.

All simulation analyses were done at different source distances from 0.1 up to 10kpc. Furthermore, several families of CCSNe GWs were used (see next subsection). Results for the analysis around time window TW1 are presented in figure 1c for the case of the waveform named "he3.5" from [44]. These DE curves are for a FAR of 2.74×10^{-2}, 2.47×10^{-2} and 9.30×10^{-6} Hz for L1, H1, and L1H1, respectively. This result shows how the efficiency reduces as the distance increases, and quite similar performance for all network of detectors. Despite the good and steady sensitivity irrespective of the number of detectors, the low statistical significance imply not only to recognize the noise events with high accuracy, but also to not to affect these detection efficiencies.

3. CCSNe GW waveforms

To carry out simulation analyses, freely-available CC-SNe GWs from recent 3D simulations were selected to be added into noise data. We selected recent families
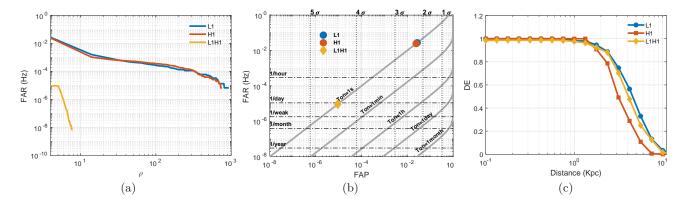
FIG. 1. cWB search results for the cases of single detector network (L1 and H1) and two detectors network (L1H1) using strain data in the time window TW1. (a) False Alarm Rate (FAR) versus the ρ statistic from the background analysis. (b) False alarm rate (FAR) versus False Alarm Probability (FAP) for the case of a on-source window of 1 second. (c) Detection efficiency (DE) with respect to the source distance achieved with the CCSNe GW "he3.5" from *Powell et al. 2019* for a corresponding FAR of 2.74×10^{-2}, 2.47×10^{-2} and 9.30×10^{-6} Hz for L1, H1, and L1H1, respectively.

of CCSNe GW waveforms that were used in the all-sky search of short GW bursts in O3 [45], and that were part of the recent study devoted to investigate the detectability of GWs from CCSNe in the upcoming fourth and fifth observing runs [8]. In specific, we used the following waveform families representing a wide variety of physical phenomena and modeling methods:

- *Scheidegger et al. 2010.* This a large set of GWs obtained in 3D-simulations of magnetohydrodynamic (MHD) driven explosions with diverse rotating progenitors [46]. For this work we considered three representative GWs generated by a progenitor start of $15M_\odot$ with different rotational speeds and with neutrino leakage scheme, R1E1CA_L (slowly rotating), R3E1AC_L (moderate rotating), and R4E1FC_L (rapidly rotating).

- *O'Connor et al. 2018.* This is a family of seven GWs generated by a $20M_\odot$ progenitor star mass [47]. The simulations considered neutrino physics and the resulting GW signatures exhibit strong g-modes and SASI components. The name of the seven waveforms are mesa20, mesa20_LR, mesa20_pert, mesa20_pert_LR, mesa20_v_LR, mesa20_2D, and mesa20_2D_pert.

- *Powell et al. 2019.* This is set of two GWs (named he3.5 and sl8) that were computed from simulations considering low to regular energies in the explosion mechanisms [44]. They are non-rotating models with 3.5 helium core solar masses after the star has lost its outer layers due to binary interactions for model he3.5 and with a a zero age main sequence (ZAMS) of $18M_\odot$ for model sl8. The GW waveforms exhibit the typical frequency rise associated with emission of g-modes.

- *Powell et al. 2020.* This family of three GWs were obtained from simulations including explo-

sion properties of the progenitor star [48]. The name of the waveforms are m39, s18np, y20. The first model is from a rapidly rotating progenitor of $39M_\odot$, while the other two are from non-rotating progenitors of 18 and $20M_\odot$. The model s18np has the same progenitor as the model s18 in Powell et al. 2019, but it does not include perturbations (np).

- *Mezzacappa et al. 2020.* This is a single GW named C15-3D generated by a $15M_\odot$ progenitor start that includes a neutrino-driven convection mechanisms [49]. The GW signature presents low- and high-frequency components and SASI emission.

B. Features

The features used to feed the classification algorithms to be able to discriminate between noise and signal events are reconstruction parameters provided by cWB for each event. We used the same set of parameters proposed in [42] which are representative to describe the duration, frequency and time-frequency characteristics of the reconstructed GW burst transients:

- ρ: cWB detection statistic.

- *Volume*: number of time-frequency pixels composing the event.

- *Duration 1* and *Duration 2*: time length of the event computed from the energy-weighted and from the reconstructed waveform.

- *Frequency 0* and *Frequency 1*: central frequency of the event computed from the energy-weighted and from the reconstructed waveform.

- *Low* and *High*: minimum and maximum frequency of the time-frequency map pixels.

- *Bandwidth 1* and *Bandwidth 2*: bandwidth estimated from the energy-weighted and from the time-frequency map.

- *Norm*: effective number of time-frequency resolutions used for GW reconstruction.

Hence, the vector of features is $\mathbf{x} \in \mathbb{R}^{N_f \times 1}$ where N_f is the number of reconstruction parameters extracted from each event.

C. Classification algorithms

The classifier is a computational model that takes as input a vector of features extracted from a cWB event, and assigns to it one class label indicating "noise" or "signal". Common and robust classification models used in diverse applications are Linear Discriminant Analysis (LDA) and Support Vector Machines (SVM). These classifiers allow to identify a linear or non-linear separation hyper-surface in such a way that the class assigned to a input vector of features depends on which region the vector is located [50]. They consist of a discriminant function defined by $\mathbf{w}^T \cdot f(\mathbf{x}) = 0$, where $\mathbf{x} \in \mathbb{R}^{N_f \times 1}$ is the vector of N_f features, $f(\cdot)$ is a transformation function, and $\mathbf{w} \in \mathbb{R}^{N_f \times 1}$ is a vector of classification weights (i.e., discriminant vector) that have to be calculated from a training dataset $\{\mathbf{x}_i, y_i\}, i = 1, ..., N$, where y_i indicates whether $\mathbf{x}_i$ is a feature vector extracted from a noise or a signal event, and N in the number of instances.

LDA defines the transformation function as $f(\mathbf{x}) = \mathbf{x}$, thus it is only able to construct linear separation surfaces since the discriminant function becomes $\mathbf{w}^T \cdot \mathbf{x} = 0$ [51] . In this classifier, the discriminant vector $\mathbf{w}$ is estimated by seeking the projection that maximizes the difference between the means of the classes while minimizes their variance (leading to a classification model that is optimal when the two classes are Gaussian with equal co-variance). Thus, the hyper-surface is found by solving this optimization problem:

$$\widehat{\mathbf{w}} = \underset{\mathbf{w}}{arg\,max} \frac{\mathbf{w}^T \, \mathbf{S}_B \, \mathbf{w}}{\mathbf{w}^T \, \mathbf{S}_W \, \mathbf{w}}, \qquad (2)$$

where $\mathbf{S}_B$ is the between-class co-variance matrix and $\mathbf{S}_W$ is the within-class co-variance matrix [51].

SVM is able to compute linear or non-linear separation surfaces in such a way that maximizes the separation between the hyper-surface and the nearest data points of each class which are called support vectors [52]. This involves solving the following optimization problem:

$$\widehat{\mathbf{w}} = \underset{\mathbf{w}, \xi}{arg\,min} \frac{1}{2}\|\mathbf{w}\|^2 + C \sum_{i=1}^{n} \xi_i, \qquad (3)$$

subject to the condition $y_i \left(\mathbf{w}^T f(\mathbf{x}_i) \right) \geq 1 - \xi_i, \forall i = 1, ..., N$, where $\xi_i \geq 0$ measure the error in the misclassification of $\mathbf{x}_i$. The separation hyper-plane is linear when

$f(\mathbf{x}) = \mathbf{x}$, leading to the linear SVM (or SVML). On the contrary, non-linear separation boundaries are obtained through a kernel function. The most common kernel is the radial basis function (RBF), which leads to a non-linear SVM with RBF (or SVMR),

In this work, we first used LDA, SVML and SVMR to test the discrimination between noise and signal events, and then, we selected the one with the higher performance in order to assess the actual improvement produced by classification model as a follow-up method in offline cWB searches.

III. RESULTS

A. Training/Testing with data from the same time window

As a first step to investigate the classification between noise and signal triggers we perform both the training and testing with information from the same stretch of strain data, i.e., noise and signal triggers used to train and test the classification model are from the same background and simulation analyses. To do so, a cross-validation process was employed to assess classification performance. The process was implemented as follows: (i) randomly split the dataset of noise and signal triggers into K non-overlapping subsets or folds; (ii) use the data from $K - 1$ subsets for training and the data from the remaining subset for testing (note that data for training and testing are always mutually exclusive); (iii) use the training set to calculate the parameters of the classification model; (iv) fed the classification model with all the triggers from the test set and compute performance metrics by comparing the output labels provided by the classifier with the corresponding true labels; (v) repeat steps (ii) to (iv) until all the K combinations of train and test data are exhausted.

This cross-validation process with $K = 5$ was repeated 10 times to be able to compute distributions of the classification performance metrics. Prior to the training of the classification model, the number of noise and signal triggers in the training set is balanced to avoid overfitting to one of the classes, moreover, the training data was normalized according to $x_i = (x_i - \mu_i)/\sigma_i, (i = 1, 2, \cdots, N_f)$ where μ_i and σ_i are the mean and standard deviation of the i-th feature which are computed exclusively from training data. The normalization is later applied to to each feature vector in the test set.

To assess performance the following metrics were computed: true negative rate (TNR) or specificity (i.e., correct classification percentage of noise events) and false negative rate (FNR) (i.e., percentage of signal triggers incorrectly classified as noise triggers). TNR is expected to be large to reduce the FAR, while FNR is expected to be small to not affect the efficiency. Classification analyses were carried out separately for each variant of:
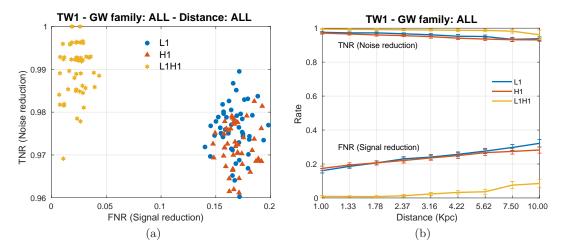
- *Detector network:* L1, H1 and L1H1.

FIG. 2. Classification results from the cross-validation analysis obtained with the dataset of triggers extracted from the time window TW1 in detector networks H1, L1, and L1H1. (a) Distribution of classification rates (TNR and FNR) for the case of signal triggers from all families of GW and from all distances combined. (b) Mean and standard deviation values of classification rates (TNR and FNR) for the case of signal triggers from all families of GW and for each distance individually. All these results were obtained when using the classification algorithm SVMR.

- *Time window of strain data:* TW1, TW2, and TW3 (see table I).

- *Distance:* 1.00, 1.33, 1.78, 2.37, 3.16, 4.22, 5.62, 7.5, 10kpc, and all those distances together.

- *Family of CCSNe GW: Scheidegger et al. 2010, O'connor et al. 2018, Powell et al. 2019, Powell et al. 2020,* and *Mezzacappa et al. 2020*) and from all those families together.

- *Classification models:* LDA, SVML, SVMR.

Figure 2a shows the correct noise classification rate (TNR) and the incorrect signal classification rate (FNR) obtained with the dataset from the time window TW1 separately with detector networks H1, L1, and L1H1. These results are for the case of signal triggers from all distances and from all families of CCSNe GW, which represents the most difficult situation for a classification model because they combine signal triggers with different SNR and with different GW signatures. The percentage of noise triggers that are correctly classified is on average 97.6 ± 0.6, 97.2 ± 0.5 and $99.4 \pm 0.6\%$ for detector networks L1, H1 and L1H1, respectively. This indicates very high specificity (nearly 100%) with very low variability (less than 0.6%) irrespective of the number of detectors in the network. On the contrary, signal triggers are lost, on average, 16.6 ± 1.2, 17.2 ± 1.4 and $2.2 \pm 0.9\%$ for detector networks L1, H1 and L1H1, respectively. These results shows a very high noise reduction irrespective of the detector network and low signal lost especially for the detector network with two interferometers.

To examine the effect of the distance, figure 2b shows the classification results obtained with signal triggers from each distance separately. As the distance increases, correct noise classification (TNR) decreases while signal misclassification (FNR) increases and the proportion of lost signals increases faster than the decrease in correct noise classification. This is due to the fact that as the distance increases the SNR of the GW signals decreases, and therefore it becomes more difficult to discriminate between noise and signal triggers because they tend to have similar characteristics.

The results presented in figure 2 a and b include signal triggers from all GW families considered herein, which represent the case with the more variability in the characteristics of signal triggers. Similar results and observations were drawn with analysis carried out with datasets of noise and signal triggers obtained from time windows TW2 and TW3. Altogether, these results shows that a better classification between noise and signal triggers with the network of two detectors L1H1, and the lower performance with the network of one detector H1.

Figure 3 shows the classification results from the cross-validation analysis for the three time windows, each family of CCSNe GW, and for the three classification algorithms. First, with regard to the three stretches of strain data (Figure 3a), high noise detection and low signal misclassification was obtained in the three cases, and there is a similar distribution of the performance metrics. Second, there is some variable classification performance considering the family of CCSNe GW (Figure 3b). In particular, using signal triggers from *O'Connor et al. 2018* yielded to a lower noise detection and higher signal lost, while triggers from *Scheidegger et al. 2010, Powell et al. 2019* and *Powell et al. 2020* resulted in the best performance with the higher TNR and lower FNR. Importantly, note that for the case of using signal triggers from all families combined the performance metrics are distributed within all families. Finally, regarding the comparison between classification algorithms, all classifiers achieved similar and high rates of noise reduction,
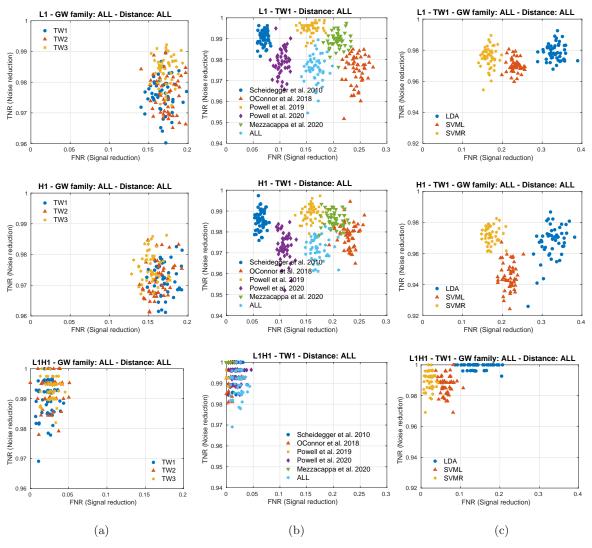
FIG. 3. Distribution of classification metrics (TNR and FNR) obtained with (a) the three different CCSNe time windows, (b) each family of CCSNe GW, and (c) the three classification algorithms. These results are for the case of the single detector network L1 (upper panel), H1 (central panel) and L1H1 (lower panel).

however, the rate of signal lost varied across them, in particular, SVMR presented the lower signal lost. In consequence, the non-linear SVMR classifier was selected to be used is the subsequent analyses. The same behavior of the noise reduction and signal lost is observed in all detector networks (L1, H1 and L1H1), but the network of two detectors always presents the best performance

A summary of the cross-validation results with the percentage of noise and signal reduction for each network of detectors (L1, H1 and L1H1), for each family of CCSNe GW, for different distances, and for the first time window of data is presented in table II. These results shows that classification performance degrades as the distance increases, shows no large differences across families of CCSNe GW and is higher for the two detector network.

B. Training/Testing with data from different time windows

The single interferometer scenario highlight the problem of only having a small amount of live time to tune the parameters of the ML model. This second study aimed to assess the actual improvement (i.e., noise reduction) and the potential drawback (i.e., signal reduction) given by the incorporation of the ML model as a follow-up method in cWB offline searches of GWs from CCSNe. Here, the classification model is learned with a dataset obtained from a given cWB analysis, and then, it is used in a different cWB analysis. This allows to effectively quantify the improvement in FAR and FAP, while also measuring the impact in the DE. In addition, to assess the robustness to unknown GWs, the ML model is tuned using all but one of families of GW signals, and then, it is tested with the remaining family.

TABLE II. Percentage of noise and signal reduction obtained by the three detector networks for each familiy of CCSNe GW waveforms and for several distances of signal triggers. These results are dataset of noise and signal triggers extracted from the time window TW1. For the two detector network there are missing results because it was not possible to train and test the classification model since no signal triggers were obtained in the simulation analysis.

	1kpc		2.37kpc		4.22kpc		7.5kpc		10kpc		All	
	Noise	Signal	Noise	Signal	Noise	Signal	Noise	Signal	Noise	Signal	Noise	Signal
L1												
Scheidegger at al. 2010	99.6	5.4	99.6	9.5	99.4	13.4	99.0	16.5	98.7	18.5	99.0	7.0
O'connor at al. 2018	97.0	20.9	93.4	35.7	91.9	39.0	90.9	39.3	91.5	39.0	97.5	24.3
Powell at al. 2019	99.4	16.2	98.4	18.5	96.7	22.1	93.7	29.6	92.5	34.5	99.5	15.8
Powell at al. 2020	98.8	8.2	97.5	12.8	96.7	16.5	95.8	20.3	95.5	22.4	97.9	10.4
Mezzacappa at al. 2020	99.2	18.1	97.2	23.7	94.0	35.8	93.5	39.5	93.0	39.1	98.9	20.8
All	97.6	16.1	96.7	23.0	95.3	25.7	93.4	29.7	93.8	32.2	97.5	16.6
H1												
Scheidegger at al. 2010	99.0	4.3	98.9	9.1	98.8	12.6	98.5	15.2	98.2	16.5	98.6	6.4
O'connor at al. 2018	96.6	23.3	93.4	32.7	93.8	33.2	93.5	33.0	93.8	33.7	97.8	22.9
Powell at al. 2019	99.4	15.1	97.8	18.7	96.2	24.5	93.5	31.4	92.3	33.0	98.9	15.9
Powell at al. 2020	97.9	7.7	96.3	14.0	94.4	16.6	93.7	19.2	93.7	20.9	97.3	10.9
Mezzacappa at al. 2020	99.2	18.2	96.2	27.3	93.5	33.2	93.8	33.3	93.1	33.5	98.6	20.1
All	97.1	17.4	95.6	22.1	94.1	25.0	93.2	27.4	93.1	28.2	97.2	17.2
L1H1												
Scheidegger at al. 2010	98.6	2.2	99.7	1.0	99.9	0.9	99.7	0.1	99.7	0.2	99.5	1.9
O'connor at al. 2018	98.7	1.3	–	–	–	–	–	–	–	–	99.4	1.1
Powell at al. 2019	100.0	0.3	99.3	0.3	99.0	0.3	95.9	0.8	–	–	99.8	0.1
Powell at al. 2020	99.8	0.1	99.3	1.5	98.5	3.2	98.7	5.0	98.2	7.0	99.4	1.9
Mezzacappa at al. 2020	100.0	0.4	99.3	0.5	–	–	–	–	–	–	100.0	0.4
All	99.5	0.9	99.0	1.3	98.7	3.3	98.2	7.5	96.1	8.5	99.0	2.2

Once the classification model is applied to a offline cWB analysis, the following performance metrics are computed: (i) noise reduction rate (TNR), and signal reduction rate (FNR); (ii) FAR and FAP before and after the application for the ML model; (iii) DE before and after the application for the ML model. Note that FAR, FAP and DE "before" refer to the original search results provided by cWB, while FAR, FAP and DE "after" refer to the results obtained with cWB in combination with the ML model. As in the first study, analyses were carried out separately with each detector network (L1, H1, L1H1), signal triggers from each distance and from all distances combined.

The results presented below are for the specific case of training the ML model with triggers extracted from the cWB analysis in the time window TW1, whereas the model is applied to the cWB analysis carried out in the time window TW2 using exclusively CCSNe GWs from *Powell et al. 2019*. Firstly, figure 4 shows the distribution of noise and signal events before and after the application of the ML model. Noise events are significantly reduced with only a small proportion remaining after classification, indeed, noise reduction is 99.4%, 99.2%, and 100.0%, for L1, H1 and L1H1, respectively. At the same time, most of the signal events remains after the application of the classifier since there is low signal reduction of 15.7%, 14.3%, and 0.3%, for L1, H1 and L1H1,

respectively. This shows that irrespective of the detector network, a high proportion of the noise triggers are removed while a small proportion of signal triggers are lost. In addition, the higher noise reduction and the lower signal lost were obtained with the two detector network.

Secondly, figure 5 shows the FAR versus ρ (or SNR for the case of single interferometer), and the FAP versus FAP before and after the application of the classifier. These are crucial results to determine the actual improvement produced by the incorporation of the classifier as a follow-up method. On the one hand, figure 5a shows a clear FAR reduction, especially in low values of ρ. For the specific operating point of the cWB search (i.e., network $\rho = 5$), the FAR before/after the classification model is $2.97 \times 10^{-2}/1.78 \times 10^{-4}$, $2.53 \times 10^{-2}/1.95 \times 10^{-4}$, and $9.76 \times 10^{-6}/9.58 \times 10^{-9}$ for L1, H1, and L1H1, respectively. This indicates that the FAR is effectively lowered in a factor of up to ~ 170, ~ 130, and ~ 1019. On the other hand, figure 5b shows the false alarm probability (FAP) before and after for on-source windows of 1s (representing a potential neutrino driven search) and of 1day (representing a optically targeted search). For the single detector network, the FAP improvement is about 1.5σ in the case of a on-source window of 1s, while no improvement was achieved in the case of a on-source window of 1day. In contrast, with the two detector network there is FAP improvement is about 1σ. We stress that these re-
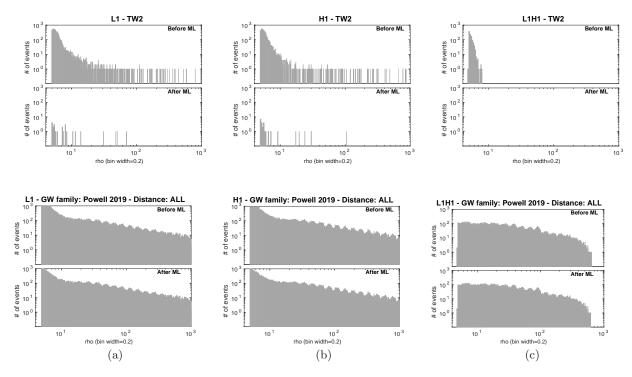
FIG. 4. Distribution of noise and signal triggers before and after the application of the classification model with CCSNe GWs from *Powell et al. 2019*. These results are for the cWB analysis in the time window TW2 for each detector network (a) L1, (b) H1, and (c) L1H1. For each detector, the classification model was tuned with the corresponding dataset of noise and signal triggers extracted from time window TW1 with signal triggers from all distances. Top panel shows the noise triggers while bottom panel shows the signal triggers.

sults represent the most difficult situation to tune and to test the classification model since it involves signal triggers of different distances, and because, the ML model is recognizing unknown GW waveforms that were not employed in the training.

Figure 6 shows the DE versus distance before and after the application of the classification model for the CCSNe GW he3.5 from *Powell et al. 2019* which enables to quantify the actual impact of the classifier. There is a minimum signal reduction for single detector network and almost null effect for two detector network. Horizontal lines allow to identify the distances with 90, 50 and 10% of efficiency. For L1 and H1, there is no effect in the distance with 10% of DE, while there is minimum reduction in the distance for 90 and 50% DE. For L1H1, no effect in the distances is observed. We conclude that for single detector networks there is marginal impact in the detection efficiency whereas there is no effect at all for the two detector network.

For the cWB analysis in time window TW2, table III shows the noise and signal reduction whereas table IV shows FAR improvement factor. As in the previous study, noise detection reduces and signal lost increases for larger distances and varies across the families of CC-SNe GW. Also, the two detector network provides the better performance with the higher noise reduction rates and lower signal reduction rates. Equivalently, FAR improvement factor is higher for lower distances, and there

are cases in which all noise triggers are completely identified and therefore the FAR after the classification is applied reaches its minimum value of the one over the background time.

IV. CONCLUSIONS

This work investigated the use of supervised machine learning (ML) methods to recognize between noise and signal events using a set of reconstruction parameters from cWB outputs. This constitutes a follow-up method devised to recognize and discard noise events while preserving signal events, which is essential to reduce the FAR and to increase the range of detection.

The proposed follow-up ML method to enhance cWB searches was analyzed in two different studies. The first aimed to ascertain the classification accuracy between noise and signal events (i.e., noise and signal reduction rates), while the second aimed to quantify the actual improvement in the statistical significance (i.e., reduction in FAR and FAP for the cases of on-source windows of 1 second and 1 day which represent a potential neutrino flux and optically targeted searches, respectively) and the impact in the detection efficiency. The two studies considered different conditions as the number of detectors in the network (L1, H1, and L1H1); three stretches of open O3a train data; diverse distances of signal events;

TABLE III. Percentage of noise and signal reduction obtained by the three detector networks and for several distances of signal triggers. These results are for the time window TW2 and for each familiy of CCSNe GW waveforms. The classification model is trained from the dataset of noise and signal triggers obtained from the cWB analysis in the time window TW1. For the two detector network there are missing results because it was not possible to train and test the classification model since no signal triggers were obtained in the simulation analysis.

	1kpc		2.37kpc		4.22kpc		7.5kpc		10kpc		All	
	Noise	Signal	Noise	Signal	Noise	Signal	Noise	Signal	Noise	Signal	Noise	Signal
L1												
Scheidegger et al. 2010	99.6	4.9	99.8	8.4	99.8	12.7	99.3	16.4	98.7	17.7	99.4	8.3
O'connor at al. 2018	96.8	19.7	93.3	31.7	92.0	35.7	91.7	35.3	91.0	35.2	97.4	17.8
Powell at al. 2019	99.3	15.3	98.3	18.8	97.1	21.7	93.6	28.1	91.9	33.1	99.4	15.7
Powell at al. 2020	98.8	8.7	97.3	12.5	95.4	15.4	94.5	19.2	93.9	21.2	97.7	11.2
Mezzacappa at al. 2020	99.2	16.9	96.8	21.9	94.2	32.3	93.9	36.3	93.2	36.0	99.1	16.8
H1												
Scheidegger at al. 2010	99.6	4.8	99.4	8.2	99.3	11.9	99.1	15.9	99.1	16.8	99.3	8.0
O'connor at al. 2018	97.9	22.4	96.2	39.2	95.8	40.2	95.5	40.4	95.5	40.2	98.4	17.8
Powell at al. 2019	99.7	15.3	98.7	18.6	97.8	25.4	95.9	37.8	95.3	40.2	99.2	14.3
Powell at al. 2020	98.2	6.9	97.3	12.1	96.3	16.3	96.3	20.4	96.2	22.2	98.1	11.6
Mezzacappa at al. 2020	99.4	16.0	97.6	27.5	95.4	39.4	95.2	40.3	95.5	40.8	99.1	15.9
L1H1												
Scheidegger at al. 2010	99.3	1.0	100.0	1.5	99.9	0.8	99.9	0.0	99.9	0.1	99.9	1.7
O'connor at al. 2018	99.4	0.8	-	-	-	-	-	-	-	-	99.7	0.7
Powell at al. 2019	100.0	0.1	99.9	0.1	99.4	0.6	98.2	0.6	-	-	100.0	0.3
Powell at al. 2020	99.8	0.1	99.2	0.5	98.3	2.8	98.4	6.0	98.1	5.9	99.6	1.9
Mezzacappa at al. 2020	100.0	0.2	99.7	0.3	-	-	-	-	-	-	100.0	0.3

TABLE IV. Improvement factor in the FAR (i.e., FAR before / FAR after) achieved for the cWB analysis in the time window TW2. The classification model is trained from the dataset of noise and signal triggers obtained from the cWB analysis in the time window TW1. Missing values indicate that no classification model was learned since no triggers were available.

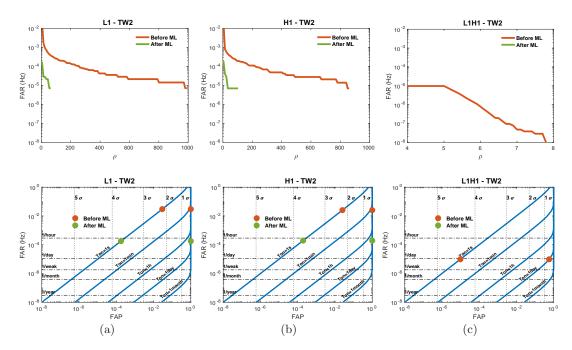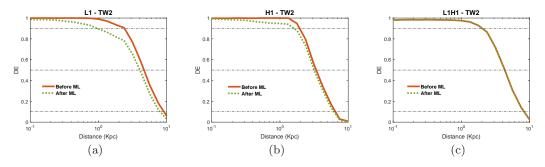|	1kpc	2.37kpc	4.22kpc	7.5kpc	10kpc	All
L1						
Scheidegger at al. 2010	260.7	595.9	417.1	139.0	74.5	173.8
O'connor at al. 2018	30.9	14.8	12.6	12.1	11.1	38.6
Powell at al. 2019	149.0	57.9	34.5	15.7	12.4	166.8
Powell at al. 2020	80.2	37.2	22.0	18.1	16.4	43.9
Mezzacappa at al. 2020	122.7	31.4	17.1	16.4	14.7	115.9
H1						
Scheidegger at al. 2010	242.7	173.4	140.0	107.1	117.5	134.8
O'connor at al. 2018	48.5	26.2	23.6	22.1	22.1	60.7
Powell at al. 2019	364.1	77.5	45.0	24.6	21.2	130.0
Powell at al. 2020	55.2	37.2	27.4	27.0	26.6	53.5
Mezzacappa at al. 2020	158.3	41.4	21.9	20.7	22.1	113.8
L1H1						
Scheidegger at al. 2010	145.6	1019.0	1019.0	1019.0	1019.0	1019.0
O'connor at al. 2018	169.8	-	-	-	-	339.7
Powell at al. 2019	1019.0	1019.0	169.8	56.6	-	1019.0
Powell at al. 2020	509.5	127.4	59.9	63.7	53.6	254.8
Mezzacappa at al. 2020	1019.0	339.7	-	-	-	1019.0

FIG. 5. Results of the cWB background search in the time window TW2 before (red) and after (green) the application of the classification model. Top panel shows the false alarm rate (FAR) versus the signal-to-noise ratio (ρ), and bottom panel shows the FAP versus false alarm rate (FAR) for the two representative cases in the search of GWs from CCSNe (on-source of 1s resembling a neutrino-driven search, and a on-source of 1day representing a optical-targeted search). These results are for each detector network, (a) L1, (b) H1, and (c) L1H1. For each detector network, the classification model was trained with the corresponding dataset of triggers extracted from the time window of TW1 and with signal triggers from all distances.

FIG. 6. Results of the cWB simulation analysis for the time window TW2 before (red) and after (green) the application of the classification model. These results are the detection efficiency (DE) versus distance for each detector network, (a) L1, (b) H1, and (c) L1H1, obtained with the CCSNe GW he3.5 from *Powell et al. 2019*. The FAR before/after the classification model is $2.97 \times 10^{-2}/1.78 \times 10^{-4}$, $2.53 \times 10^{-2}/1.95 \times 10^{-4}$, and $9.76 \times 10^{-6}/9.58 \times 10^{-9}$ for L1, H1, and L1H1, respectively. Again, for each detector network the classification model was trained with the corresponding dataset of triggers extracted from the time window DC1 and with signal triggers from all distances.

various CCSNe GW families with diverse characteristics; and three classification algorithms. Overall, the results of these studies and variety of conditions showed high noise reduction rates greater than 90% and low signal misclassification rates lower than 10%.

Importantly, in our analyses we quantified the impact bring by the ML model in cWB offline searches. The recognition and discharging of noise triggers from cWB outputs reduces the FAR in a factor of ~ 10 for one detector (or even more for some families of GWs), and in a factor of > 100 for two detectors in the network. Notably, these FAR reductions are equivalent to an im-

provement in the statistical significance of $\sim 1.5\sigma$ with one detector and even more with two detectors for the case of a on-source windows of 1 second. For the case of optically targeted searches with on-source windows of 1 day, on the other hand, the FAR reduction yields to a $> 1\sigma$ improvement with a two detector network.

The variety of explored conditions allowed to study several important aspects. First, comparing the classification performance when tuning the ML model with a single or with various families of CCSNe GW showed no large differences in performance. This shows the robustness of the ML approach with respect to different types of

CCSNe GW signatures, and points out that the morphology of noise and signal triggers are effectively distinguishable irrespective of the GW signals used to train and to test the classifier.vThis is noteworthy since in practice the ML model must be tuned with a pool of synthetic GWs obtained for example through numerical simulations (as is done in this work), but its usage in a real situation implies correctly identifying signal events generated by actual GWs that will not exactly match GWs used to train the model. Secondly, there are high noise reduction rates with low signal reduction rates irrespective of the number of detectors in the network, though the classification performance is better with two detectors. This is simply because with two detectors the algorithm is tuned with a uniform distribution of noise triggers with low and moderate SNR since high SNR noise triggers do not survive coherence tests between detectors. Also, regarding the different classification models, the results indicated a consistent superior performance with the support vector machine with radial basis function as kernel (SVMR). This model provides non-linear separation surfaces allowing to account for non-linearities in the feature space. Despite other methods can also be used for this problem, we recommend the use of SVMR since it has been proved to be stable and robust.

With regard to the training of the classification model, it is important to indicate that the training data in all analyses was always balanced, hence, there is an equal probability for each class (noise and signal). However, during the training procedure of the classifiers there is the possibility to assign more probability to one class. This would allow to increase the noise reduction at the expense of the signal reduction, or vice-verse. This is an important aspect to be able to trade off between statistical significance and efficiency. It is also possible to train the classification model using only a subset with loudest noise events, or with a subset of signal events with certain characteristics (duration, bandwidth, etc.). This can be used to bias the follow-up ML model to prioritize specific aspects during the search by recognizing better one specific type of noise or signal event.

To sum up, this work presented a follow-up machine learning method for cWB based searches of GWs from CCSNe which can be use with one or multiple detectors. The method identifies and discharges non-astrophysical noise transients allowing to reduce the false alarm rate, to improve the statistical significance and to increase the detection range of the searches. The model is simply tuned with a set of noise and signal triggers and then can be easily incorporated into the search pipeline.

V. ACKNOWLEDGEMENTS

[1] J. Aasi et al. (LIGO Scientific), Class. Quant. Grav. 32, 074001 (2015), arXiv:1411.4547 [gr-qc].
[2] F. Acernese et al. (Virgo), Class. Quant. Grav. 32, 024001 (2015), arXiv:1408.3978 [gr-qc].
[3] Y. Aso, Y. Michimura, K. Somiya, M. Ando, O. Miyakawa, T. Sekiguchi, D. Tatsumi, and H. Yamamoto (KAGRA), Phys. Rev. D 88, 043007 (2013), arXiv:1306.6747 [gr-qc].
[4] B. Abbott et al. (LIGO Scientific, Virgo), Phys. Rev. X 9, 031040 (2019), arXiv:1811.12907 [astro-ph.HE].
[5] R. Abbott et al., (2020), arXiv:2010.14527 [gr-qc].
[6] S. Richers, C. D. Ott, E. Abdikamalov, E. O'Connor, and C. Sullivan, Phys. Rev. D 95, 063019 (2017).
[7] B. Abbott et al. (LIGO Scientific, Virgo), Phys. Rev. D 101, 084002 (2020), arXiv:1908.03584 [astro-ph.HE].
[8] M. J. Szczepańczyk, J. M. Antelis, M. Benjamin, M. Cavaglià, D. Gondek-Rosińska, T. Hansen, S. Klimenko, M. D. Morales, C. Moreno, S. Mukher-

jee, G. Nurbek, J. Powell, N. Singh, S. Sitmukhambetov, P. Szewczyk, O. Valdez, G. Vedovato, J. Westhouse, M. Zanolin, and Y. Zheng, Phys. Rev. D 104, 102002 (2021).
[9] M. Drago, S. Klimenko, C. Lazzaro, E. Milotti, G. Mitselmakher, V. Necula, B. O'Brian, G. A. Prodi, F. Salemi, M. Szczepanczyk, S. Tiwari, V. Tiwari, G. V, G. Vedovato, and I. Yakushin, SoftwareX 14, 100678 (2021).
[10] V. Necula, S. Klimenko, and G. Mitselmakher, Journal of Physics: Conference Series 363, 012032 (2012).
[11] S. Klimenko, S. Mohanty, M. Rakhmanov, and G. Mitselmakher, Phys. Rev. D 72, 1 (2005), arXiv:0508068 [gr-qc].
[12] S. Klimenko, I. Yakushin, A. Mercer, and G. Mitselmakher, Class. Quantum Gravity 25, 1 (2008), arXiv:0802.3232.
[13] S. Klimenko, G. Vedovato, M. Drago, F. Salemi, V. Tiwari, G. A. Prodi, C. Lazzaro, K. Ackley, S. Tiwari, C. F. Da Silva, and G. Mitselmakher, Phys. Rev. D 93, 1 (2016).
[14] B. Abbott et al. (LIGO Scientific, Virgo), Phys. Rev. Lett. 116, 061102 (2016),

arXiv:1602.03837 [gr-qc].

[15] B. Abbott *et al.* (LIGO Scientific, Virgo), Phys. Rev. D **93**, 122004 (2016), [Addendum: Phys.Rev.D 94, 069903 (2016)], arXiv:1602.03843 [gr-qc].

[16] R. Abbott *et al.* (LIGO Scientific, Virgo), Phys. Rev. Lett. **125**, 101102 (2020), arXiv:2009.01075 [gr-qc].

[17] M. Szczepanczyk *et al.*, (2020), arXiv:2009.11336 [astro-ph.HE].

[18] R. Abbott *et al.* (LIGO Scientific, Virgo), Astrophys. J. **896**, L44 (2020), arXiv:2006.12611 [astro-ph.HE].

[19] B. P. Abbott *et al.* (LIGO Scientific, VIRGO), Phys. Rev. D **93**, 042005 (2016), arXiv:1511.04398 [gr-qc].

[20] B. P. Abbott *et al.* (LIGO Scientific, Virgo), Class. Quant. Grav. **35**, 065009 (2018), arXiv:1711.06843 [gr-qc].

[21] B. P. Abbott *et al.* (LIGO Scientific, Virgo), Phys. Rev. D **95**, 042003 (2017), arXiv:1611.02972 [gr-qc].

[22] B. Abbott *et al.* (LIGO Scientific, Virgo), Phys. Rev. D **100**, 024017 (2019), arXiv:1905.03457 [gr-qc].

[23] D. Davis, J. S. Areeda, B. K. Berger, R. Bruntz, A. Effler, R. C. Essick, R. P. Fisher, P. Godwin, E. Goetz, and A. F. Helmling-Cornell, (2021), 10.1088/1361-6382/abfd85.

[24] E. Cuoco, J. Powell, M. Cavaglià, K. Ackley, M. Bejger, C. Chatterjee, M. Coughlin, S. Coughlin, P. Easter, R. Essick, H. Gabbard, T. Gebhard, S. Ghosh, L. Haegel, A. Iess, D. Keitel, Z. Márka, S. Márka, F. Morawski, T. Nguyen, R. Ormiston, M. Pürrer, M. Razzano, K. Staats, G. Vajente, and D. Williams, Machine Learning: Science and Technology **2**, 011002 (2020).

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

[26] D. George and E. A. Huerta, Phys. Rev. D **97**, 044039 (2018).

[27] M. D. Morales, J. M. Antelis, C. Moreno, and A. I. Nesterov, Sensors **21**, 3175 (2021).

[28] A. Torres-Forné, A. Marquina, J. A. Font, and J. M. Ibáñez, Phys. Rev. D **94**, 124040 (2016).

[29] H. Wang, S. Wu, Z. Cao, X. Liu, and J.-Y. Zhu, Phys. Rev. D **101**, 104003 (2020).

[30] P. Astone, P. Cerda-Duran, I. D. Palma, M. Drago, F. Muciaccia, C. Palomba, and F. Ricci, Phys. Rev. D **98**, 122002 (2018).

[31] M. L. Chan, I. S. Heng, and C. Messenger, Phys. Rev. D **102**, 043022 (2020).

[32] A. Iess, E. Cuoco, F. Morawski, and J. Powell, Machine Learning: Science and Technology **1**, 025014 (2020).

[33] M. López, I. Di Palma, M. Drago, P. Cerdá-Durán, and F. Ricci, Phys. Rev. D **103**, 063011 (2021).

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

[35] R. Biswas, L. Blackburn, J. Cao, R. Essick, K. A. Hodge, E. Katsavounidis, K. Kim, Y.-M. Kim, E.-O. Le Bigot, C.-H. Lee, J. J. Oh, S. H. Oh, E. J. Son, Y. Tao, R. Vaulin, and X. Wang, Phys. Rev. D **88**, 062003 (2013).

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

[37] G. Vajente, Y. Huang, M. Isi, J. C. Driggers, J. S. Kissel, M. J. Szczepańczyk, and S. Vitale, Phys. Rev. D **101**, 042003 (2020).

[38] V. Gayathri, D. Lopez, P. R. S., I. S. Heng, A. Pai, and C. Messenger, Phys. Rev. D **102**, 104023 (2020).

[39] T. Mishra, B. O'Brien, V. Gayathri, M. Szczepanczyk, S. Bhaumik, I. Bartos, and S. Klimenko, (2021).

[40] S. Mukherjee, L. Salazar, J. Mittelstaedt, and O. Valdez, Phys. Rev. D. **96**, 104033 (2017).

[41] S. Mukherjee, G. Nurbek, and O. Valdez, Phys. Rev. D **103**, 103008 (2021).

[42] M. Cavaglià, S. Gaudio, T. Hansen, K. Staats, M. Szczepańczyk, and M. Zanolin, Machine Learning: Science and Technology **1**, 015005 (2020).

[43] M. Vallisneri, J. Kanner, R. Williams, A. Weinstein, and B. Stephens, Journal of Physics: Conference Series **610**, 012021 (2015).

[44] J. Powell and B. Müller, Mon. Not. R. Astron. Soc. **487**, 1178 (2019), arXiv:1812.05738.

[45] R. Abbott *et al.*, (2021), arXiv:2107.03701 [gr-qc].

[46] S. Scheidegger, R. Käppeli, S. C. Whitehouse, T. Fischer, and M. Liebendörfer, Astron. Astrophys. **514** (2010), 10.1051/0004-6361/200913220.

[47] E. P. O'Connor and S. M. Couch, Astrophys. J. **865**, 81 (2018), arXiv:1807.07579.

[48] J. Powell and B. Müller, **12**, 1 (2020), arXiv:2002.10115.

[49] A. Mezzacappa *et al.*, Phys. Rev. D **102**, 023027 (2020), arXiv:2007.15099 [astro-ph.HE].

[50] R. Duda, P. Hart, and D. Stork, *Pattern Classification* (Wiley, 2001).

[51] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)* (Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006).

[52] C. Cortes and V. Vapnik, in *Machine Learning* (1995) pp. 273–297.