





Exploring gravitational-wave detection and parameter inference using deep learning methods

João D Álvares¹ , José A Font^{2,3}, Felipe F Freitas⁴,
Oswaldo G Freitas¹ , António P Morais⁴,
Solange Nunes¹ , Antonio Onofre^{1,*}  and
Alejandro Torres-Forné^{2,5}

¹ Centro de Física das Universidades do Minho e do Porto (CF-UM-UP),
Universidade do Minho, 4710-057 Braga, Portugal

² Departamento de Astronomía y Astrofísica, Universitat de València, Dr. Moliner
50, 46100, Burjassot (València), Spain

³ Observatori Astronòmic, Universitat de València, Catedrático José Beltrán 2,
46980, Paterna (València), Spain

⁴ Departamento de Física da Universidade de Aveiro and Centre for Research and
Development in Mathematics and Applications (CIDMA) Campus de Santiago,
3810-183 Aveiro, Portugal

⁵ Max Planck Institute for Gravitational Physics (Albert Einstein Institute), D-14476
Potsdam-Golm, Germany

E-mail: joadinis01@gmail.com, j.antonio.font@uv.es, felipefreitas@ua.pt,
ogf1996@gmail.com, aapmorais@ua.pt, solangesilnunes@gmail.com,
antonio.onofre@cern.ch and alejandro.torres-forne@aei.mpg.de

Received 23 December 2020, revised 19 May 2021

Accepted for publication 24 May 2021

Published 8 July 2021



CrossMark

Abstract

We explore machine learning methods to detect gravitational waves (GW) from binary black hole (BBH) mergers using deep learning (DL) algorithms. The DL networks are trained with gravitational waveforms obtained from BBH mergers with component masses randomly sampled in the range from 5 to 100 solar masses and luminosity distances from 100 Mpc to, at least, 2000 Mpc. The GW signal waveforms are injected in public data from the O2 run of the Advanced LIGO and Advanced Virgo detectors, in time windows that do not coincide with those of known detected signals. We demonstrate that DL algorithms, trained with GW signal waveforms at distances of 2000 Mpc, still show high accuracy when detecting closer signals, within the ranges considered in our analysis. Moreover, by combining the results of the three-detector network in a unique RGB image, the single detector performance is improved by as much as

*Author to whom any correspondence should be addressed.

70%. Furthermore, we train a regression network to perform parameter inference on BBH spectrogram data and apply this network to the events from the GWTC-1 and GWTC-2 catalogs. Without significant optimization of our algorithms we obtain results that are mostly consistent with published results by the LIGO–Virgo Collaboration. In particular, our predictions for the chirp mass are compatible (up to 3σ) with the official values for 90% of events. From these results we conclude that the combination of computer vision techniques and deep-learning methods put forward in this work is a worthy addition to the GW astronomer’s toolbox.

Keywords: gravitational waves, deep learning, machine learning, black hole, LIGO, Virgo, parameter inference

(Some figures may appear in colour only in the online journal)

1. Introduction

The detection of gravitational waves (GW) from binary black hole (BBH) mergers [1, 2] during the first data-taking run (O1) of Advanced LIGO [3] was a remarkable milestone that opened up a new window for observing the cosmos. The European detector Advanced Virgo [4] joined the efforts during the second observing run (O2) which helped improve the sky localization of the sources. Notably, O2 included the first observation of GW from a binary neutron star (BNS) merger, GW170817. This event was accompanied by electromagnetic radiation which was observed by dozens of telescopes worldwide and brought forth the field of multimessenger astronomy [5]. During O1 and O2 the LIGO Scientific Collaboration and the Virgo Collaboration (LVC) announced the confident detection of eleven GW signals from compact binary coalescences (CBC) [6]. The third science run (O3) ended on March 2020 after completing almost one full year of data-taking. O3 provided a record number of detections, publicly released as low-latency alerts through the GW catalog event database⁶. Recently, the LVC has released their second GW transient catalog comprising the 39 CBC detections accomplished in the first six months of O3 [7].

The detection of GW signals from CBC relies on accurate waveform templates against which to perform match-filtered searches. Pipelines to perform these searches include pyCBC [8], GstLAL [9] and MBTA [10]. Faithful templates can be built either by solving the gravitational field equations with numerical relativity techniques or by using approximations to the two-body problem in general relativity. Current gravitational waveform models (or approximants) combine analytical and numerical approaches and they are able to describe the entire inspiral–merger–ringdown signal for a large variety of possible configurations of the parameter space (see e.g. [11–14] and references therein). Once a CBC source is detected, the estimation of its characteristic physical parameters such as component masses, individual spins or distance, is based on Bayesian inference, using software such as LALInference [15] and Bilby [16]. However, Bayesian inference can be computationally expensive as it may take of the order of days to obtain sufficient number of posterior samples for BBH [17]. The situation aggravates as the number of detections increases, as it is expected in the forthcoming observational campaigns of the GW detector network. As an example, the predicted detection count of BBH mergers in one-calendar-year observing run of the network during O4 is 79^{+89}_{-44} [18]. To over-

⁶ gracedb.ligo.org.

come this difficulty deep learning (DL) algorithms constitute an attractive choice to speed-up parameter estimation [17, 19–23].

Machine learning (ML) and DL are bringing about a revolution in data analysis across a variety of fields and GW astronomy is not alien to that trend. In particular, the use of deep neural networks (DNN) [24] for classification and/or prediction tasks has become the standard on data analysis applications, ranging from industrial applications [25–27] and medical diagnosis [28, 29] to particle physics [30–32] and cosmology [33, 34]. This trend has now organically been extended to GW astronomy, both for signal detection [35–37] and for detector characterization, by reducing the impact of noise artifacts or ‘glitches’ of instrumental and environmental origin [38–46]. Glitches can potentially affect GW detection as they contribute to the background in transient searches, decreasing the statistical significance and increasing the false alarm ratio of actual GW events. For this reason, the detection and classification of glitches has become an important application of DL in GW astronomy. Worth highlighting is the Gravity Spy project [47] which combines DL with citizen science to identify and label families of glitches from the twin Advanced LIGO detectors. Recent approaches to eliminate, or at least mitigate, the effect of glitches are discussed in [48–50]. We also note that recent developments of GW signal denoising algorithms include total-variation methods [51, 52], dictionary learning approaches [53], DL autoencoders and deep filtering [54] or wavenets [55].

Contrary to CBC signals there are other types of GW sources whose detection is not template-based, namely burst sources, continuous-wave sources and stochastic sources. Archetypal examples of the first two are core-collapse supernovae (CCSN) and rotating neutron stars (RNS), respectively. In the case of CCSN the number of waveform templates available is fairly scarce due to the complexity of the very computationally expensive numerical simulations required to model the supernova mechanism, rendering a comprehensive survey of the vast parameter space of the problem impracticable. In the case of RNS, their continuous, monochromatic GW signals are very stable, which makes match-filtering-based techniques unfeasible due to the computational resources the task would involve. Both sources are however excellent candidates for DL methods and specific pipelines have been developed in the two cases. Recent approaches that use DL to improve the chances of detection of CCSN GW signals are discussed in [56–58]. Additionally, as DL methods are designed to efficiently deal with large amounts of data, they could offer a very effective solution to detect and analyze signals from RNS [59–62]. Though still not widely adopted by the GW astronomy community, DL-based techniques are quickly showing themselves to be a promising addition to the GW astronomer’s toolbox.

In this paper we explore the use of computer vision techniques and DL methods to both detect GW from BBH mergers and perform parameter inference using RGB spectrograms that combine open data from the Advanced LIGO and Advanced Virgo three-detector network. Our usage of residual networks (ResNets) in spectrogram data is a novel approach when compared to existing pipelines for detection and parameter estimation. To achieve our goal we train a cross-residual network (\times ResNet) which allows us to extract information about source parameters such as luminosity distance, chirp mass, network antenna power (NAP), and effective spin. As we show below, the application of our ResNet to the BBH detections included in the GWTC-1 and GWTC-2 catalogs yields a remarkable agreement with the LVC results. All our work was developed using the python programming language. The pyCBC [63] and gwpy [64] libraries were used for data generation and treatment, while DNN models were implemented with the fastai [65] library.

This paper is organized as follows: section 2 describes the generation of the GW datasets used for training and testing. Section 3 deals with the general characteristics of our deep neural network, describing its various architectures and the methodology employed for classification

Table 1. Description of the full classification and regression datasets for training and validation with both single and multiple detectors, using an 80/20 training/validation split. The images are generated from the wave-forms calculated by pyCBC. For the classification datasets, the individual masses (m_1, m_2) are sampled with a uniform distribution within the range of 5 to $100M_\odot$. For the regression dataset the parameters of individual masses, distances, inclination and spin are also uniformly sampled. The extra dataset is generated in order to include more examples of small masses and complement the dataset generated with the SEOBNRv4HM_ROM approximant.

Classification		
Parameters	Train size	Validation size
Single detector		
$(m_1, m_2) \sim U(5, 100)M_\odot$,	4000 images	1000 images
$d_L = [100, 300, 1000, 1500, 2000]$ Mpc,	560×560 pixels	560×560 pixels
$\iota = \frac{\pi}{2}$,	8-bit gray scale	8-bit gray scale
Approximant: SEOBNRv4_ROM		
Total images	20 000	5000
Multiple detector	4000 images	1000 images
(Same parameters as above)	560×560 pixels	560×560 pixels
	8-bit RGB	8-bit RGB
Total images	20 000	5000
Regression		
Parameters	Train size	Validation size
Multiple detector		
$(m_1, m_2) \sim U(5, 100)M_\odot$,	12 769 images	3192 images
$d_L \sim U(100, 4000)$ Mpc,	224×224 pixels	224×224 pixels
$\iota \sim U(0, \pi)$,	8-bit RGB	8-bit RGB
spin $\sim U(-1, 1)$,		
Restriction: SNR > 5		
Approximant: SEOBNRv4HM_ROM		
(Same parameters as above)	10 338 images	2584 images
Approximant: IMRPhenomPv2	224×224 pixels	224×224 pixels
	8-bit RGB	8-bit RGB
(Same parameters as above)	10 625 images	2656 images
Approximant: IMRPhenomD	224×224 pixels	224×224 pixels
	8-bit RGB	8-bit RGB
Total images	43 009	14 689
Extra dataset		
Multiple detector		
$(m_1, m_2) \sim U(5, 35)M_\odot$,	15 538 images	
$d_L \sim U(100, 4000)$ Mpc,	224×224 pixels	
$\iota \sim U(0, \pi)$,	8-bit RGB	
spin $\sim U(-1, 1)$,		
Restriction: SNR > 5		
Approximant: SEOBNRv4HM_ROM		

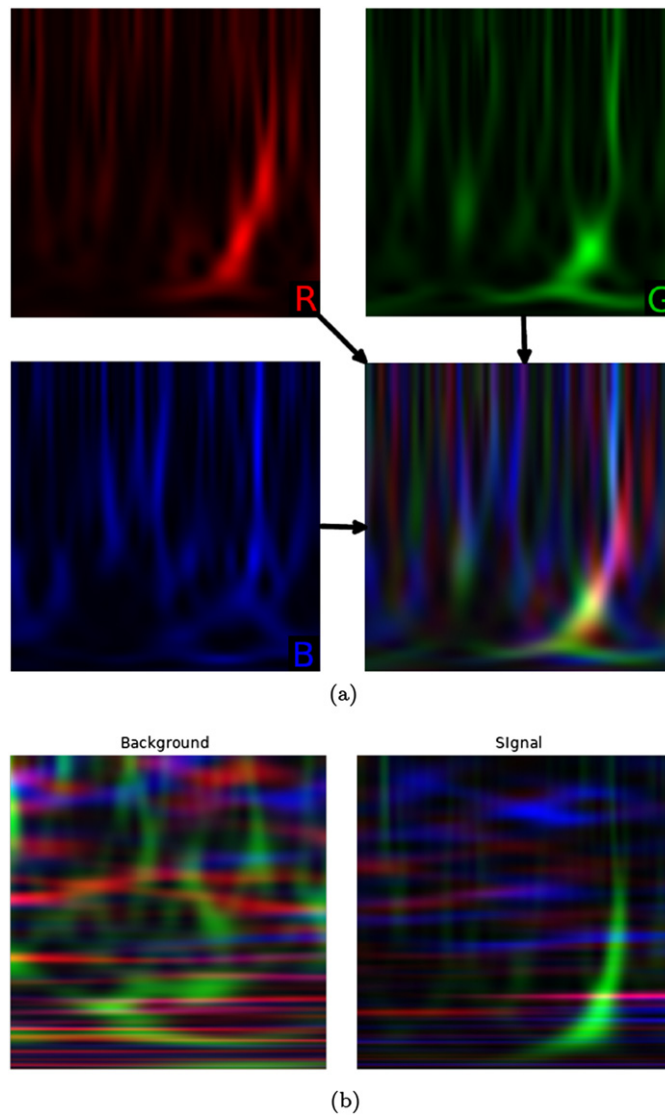


Figure 1. (a) Combining single detector spectrogram data into a single RGB image, to be used combined by the DL network architectures. The Hanford (top-left), Livingston (top-right) and Virgo (bottom-left) spectrogram data, are used as the red, green and blue images, respectively, to build the full RGB image (bottom-right). In all plots the x -axis is time and the y -axis is frequency, while the colour intensity stands for the signal power normalised to the 8-bit integer range (0–255). (b) RGB image from background labeled spectrogram (left) as compared with a spectrogram where a GW waveform was injected into real conditions noise (right).

and regression. The assessment of the network is discussed in section 4 and our main results regarding the use of our network for the analysis (detection and inference) of real GW events is presented in section 5. Finally, our main conclusions are summarized in section 6.

2. GW datasets generation

We begin by describing the generation of the datasets used in our analysis. All CBC waveforms employed in the classification datasets were obtained using pyCBC [63] with the SEOBNRv4_ROM approximant while the regression datasets use SEOBNRv4HM_ROM [66], IMRPhenomPv2 [67] and IMRPhenomD [68, 69]. For the sake of simplicity we start by considering spinless black holes and quasi-circular binaries with no orbital eccentricity. Furthermore, since current GW detector networks are far more sensitive to the plus polarization than to the cross one, we only generate plus-polarized waves. We note that this has the drawback of making impossible to break the degeneracy between luminosity distance and inclination.

2.1. Single detector waveforms for classification

The purpose of this first dataset is to allow our DL models to discern the presence of a GW signal with data collected from a single detector. In particular we employ a 500 s noise segment from the Hanford detector with initial GPS time $t_{\text{GPS}} = 1187\,058\,342$ s. Defining τ to be some time from the start of our noise segment, we randomly select $\tau_0 \in [5, 495]$ s and isolate the window $[\tau_0 - 5 \text{ s}, \tau_0 + 5 \text{ s}]$. This strain selection, which we denote as n , is then whitened through inverse spectrum truncation, using its own amplitude spectral density (ASD). Then, we apply a bandpass filter from 20 Hz to 300 Hz, as well as notch filters at the individual frequencies 60 Hz, 120 Hz and 240 Hz. For the generation of the waveform signal strain h , a random pair of black hole masses $(m_1, m_2) \in U([5, 100]) M_\odot$ is selected for a BBH merger with luminosity distance $d_L = 2000$ Mpc and inclination $\iota = \frac{\pi}{2}$. This waveform is whitened using the ASD of the selected noise strain n and the same filtering process is undertaken. The resulting waveform is injected into the noise window in such a way that the maximum amplitude occurs at τ_0 . Following this, the constant- Q transform is calculated for the $[\tau_0 - 0.16 \text{ s}, \tau_0 + 0.4 \text{ s}]$ interval in the composite signal $S = h + t$, and a spectrogram is produced. A second spectrogram without signal injection is also generated for the same interval. Both spectrograms are saved as images and appropriately labeled as ‘signal’ and ‘background’. This process is iterated 5000 times to build our dataset. The same procedure is taken for the luminosity distances $d_L = 100, 300, 1000, 1500$ and 2000 Mpc. A summary of the single detector dataset summary is shown in the first row of table 1.

2.2. Multiple detector waveforms for classification

In order to combine the data from all three detectors (Hanford, Livingston and Virgo) we select coincident segments of 500 s from all detectors starting at a certain t_{GPS} time. The process is then identical to that of a single detector case with τ_0 , m_1 and m_2 randomly generated. A $[\tau_0 - 5 \text{ s}, \tau_0 + 5 \text{ s}]$ time window is extracted from the longer segments for the three detectors. The resulting background strain data, n_H , n_L and n_V , for Hanford, Livingston and Virgo interferometers respectively, is treated in the same way as described above. However, when injecting a signal, one must make sure that the specific ASD of each detector is being used. After the generation of the signal waveform and its injection into the background noise segments, we include the antenna power from each detector into our time series S_H , S_L and S_V . At this stage, we can emulate the sky position for the signal by randomly choosing one of three detectors as a reference, and shift the beginning of the other two time series according to their time delay with respect to the reference detector. Once the three spectrograms are produced they are combined into a $560 \times 560 \times 3$ array in such a way that each of them is represented by a certain colour channel in a RGB image. Specifically, Hanford, Livingston and Virgo datasets are mapped into the red, green and blue channels respectively, as can be seen in figure 1(a).

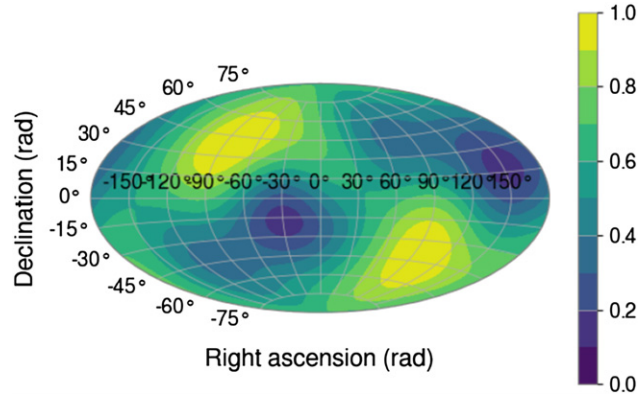


Figure 2. Network antenna power as a function of sky position for $(t \bmod 24 \text{ h}) = 9742$.

Table 2. Convolutional neural networks architectures employed for the classification, multiple (single) detectors, and regression tasks. The first two columns show the setup of the networks as explained in section 3, and the last column shows the performance on a chosen metric, tying in to section 4. The custom header for the classification CNN is described in section 3, the custom header for the regression model has the same structure with the main difference being that the final layer has only one unit with a linear activation function.

Classification		
Base architecture	Hyperparameters	Accuracy ²
ResNet-101 +custom header	Input size: $275 \times 275 \times 3(1)$, Batch size: 8 images, Learning rate: $[2 \times 10^{-3}, 2 \times 10^{-1}]$, Weight decay: 1×10^{-5} Loss function: cross entropy loss (CE)	Single detector: 0.72 Multiple detectors: 0.87
Regression		
Base architecture	Hyperparameters	RMSE
\times ResNet-18 +Blur average layer +MC dropout +custom header	Input size: $128 \times 128 \times 3$, Batch size: 64 images, Learning rate: 1×10^{-2} , Weight decay: 1×10^{-3} , Loss function: mean squared error (MSE)	0.021

As in the single detector case, an equivalent background spectrogram without signal injection is produced. Both, background (left) and signal (right) spectrograms, are represented in figure 1(b). Once again, this process is iterated 5000 times for each luminosity distance we consider. The dataset summary for the multiple detector case is summarized in the second row of table 1.

2.3. Mass dependent dataset

The purpose of this dataset is to check how the trained models perform depending on the binary component masses and luminosity distances, both for the single and multiple detector cases. To this end, we consider a number of mass combinations, where for each m_1 , ranging from 5 to $100M_\odot$ in steps of $2M_\odot$, there is a set of values for m_2 , in the range $[m_1, 100]M_\odot$, covered also with steps of $2M_\odot$. Regarding the luminosity distance d_L , we use distances ranging from 100 to 2000 Mpc, in steps of 100 Mpc. The inclination is kept fixed at $\iota = \frac{\pi}{2}$. For each d_L and each of the 1225 (m_1, m_2) mass combinations, a waveform is generated and injected into the detector's noise following the same procedure as described above. However, in this case, only the spectrograms with the injections are saved and labeled with the corresponding distance and mass.

2.4. Regression datasets

We also study how the DL algorithms can be used to extract information about the physical parameters from the generated data. This procedure is typically denoted as regression. For this purpose, a larger dataset was deemed necessary and only the multiple detector case, was considered. To avoid a dependence on a particular approximant, three different datasets for each of the approximants SEOBNRv4HM_ROM [66], IMRPhenomPv2 [67] and IMRPhenomD [68, 69] were built. It is also relevant to note that, due to an apparent degradation of the regression close to the upper range of the sampled distances, we decided to consider distances up to 4 Gpc, although we set our range of operation to go up to a maximum distance of 2.5 Gpc. Here, we do not build separate datasets for particular values of the distance but instead let d_L be randomly generated within this range. The component masses m_1 and m_2 are again randomly sampled in the $[5, 100]M_\odot$ interval while the inclination takes a random value in the $[0, \pi]$ interval. We also sample the sky position by taking into account the antenna pattern of each detector. An example of the NAP is shown in figure 2. In our regression dataset black holes are assumed to have a dimensionless spin in the range $[-1, 1]$ and those are aligned with the orbital angular momentum, allowing to compute the effective inspiral spin, χ_{eff} ,

$$\chi_{\text{eff}} = \frac{m_1\chi_1^{\parallel} + m_2\chi_2^{\parallel}}{m_1 + m_2}, \quad (1)$$

where χ_i^{\parallel} is the component of the i th spin along the orbital angular momentum. Since we assume from the beginning that a given input to the regression model will necessarily contain a GW signal of some sort, we need not to worry about generating the background-only cases. Furthermore we impose a threshold for the signal-to-noise ratio (SNR) so that we only allow cases where $\text{SNR} > 5$. The sizes of the SEOBNRv4HM_ROM, IMRPhenomPv2 and IMRPhenomD datasets are, respectively, 15 961, 12 922, and 13 281. An extra dataset focused on lower masses ($m_1, m_2 \in [5, 35]M_\odot$) was also generated using the SEOBNRv4HM_ROM approximant, containing 15 538 events. This was combined with the original SEOBNRv4HM_ROM dataset for a lower-mass weighted dataset with a total of 31 499 items. All this information is summarized in table 1.

3. Deep neural network: architectures and methodologies

As mentioned in section 2 we encode the information of the waveforms produced by BBH mergers into a spectrogram. Here we describe how to apply our DL algorithms to identify GW information from spectrogram data. This is done using the classification and regression

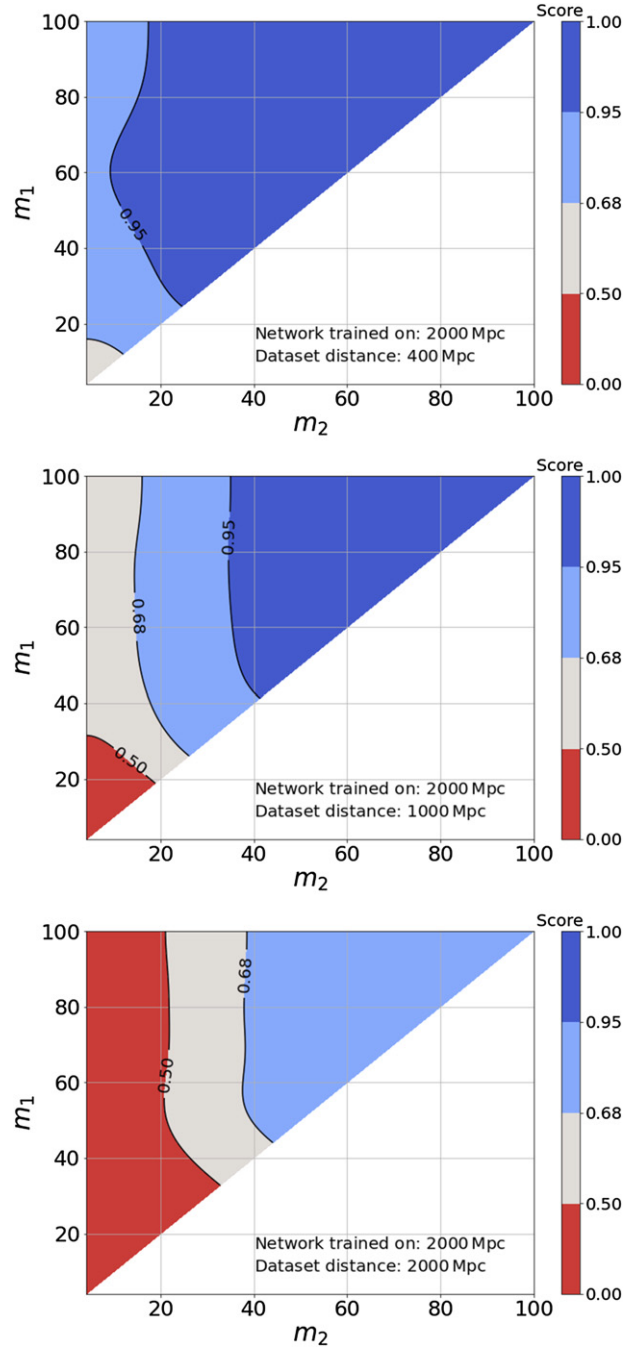


Figure 3. Simulated signal scores using one single detector for different luminosity distances and evaluated with DL networks trained with GW waveforms from BBH mergers at a luminosity distance of 2000 Mpc. Results are shown as a function of the BH masses of the binary system, m_1 and m_2 , for GW signals from sources at 400 Mpc (top), 1000 Mpc (middle) and 2000 Mpc (bottom).

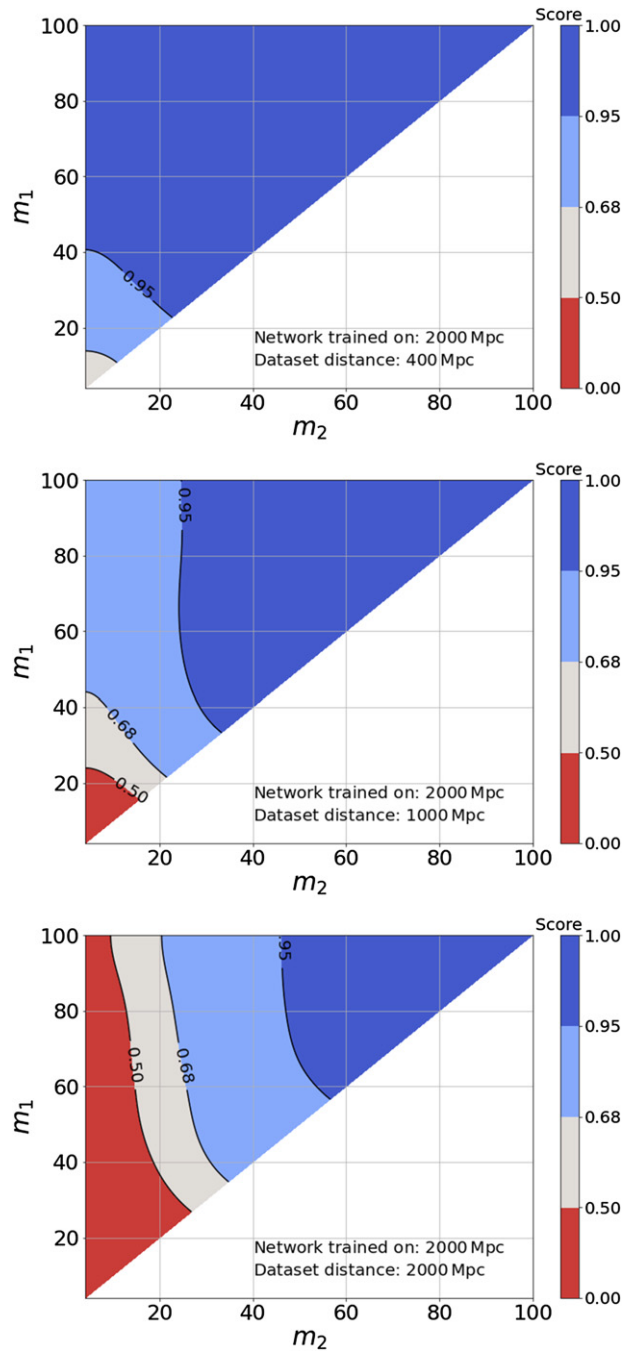


Figure 4. Same as figure 3 but employing multiple detectors to estimate the simulated signal scores displayed.

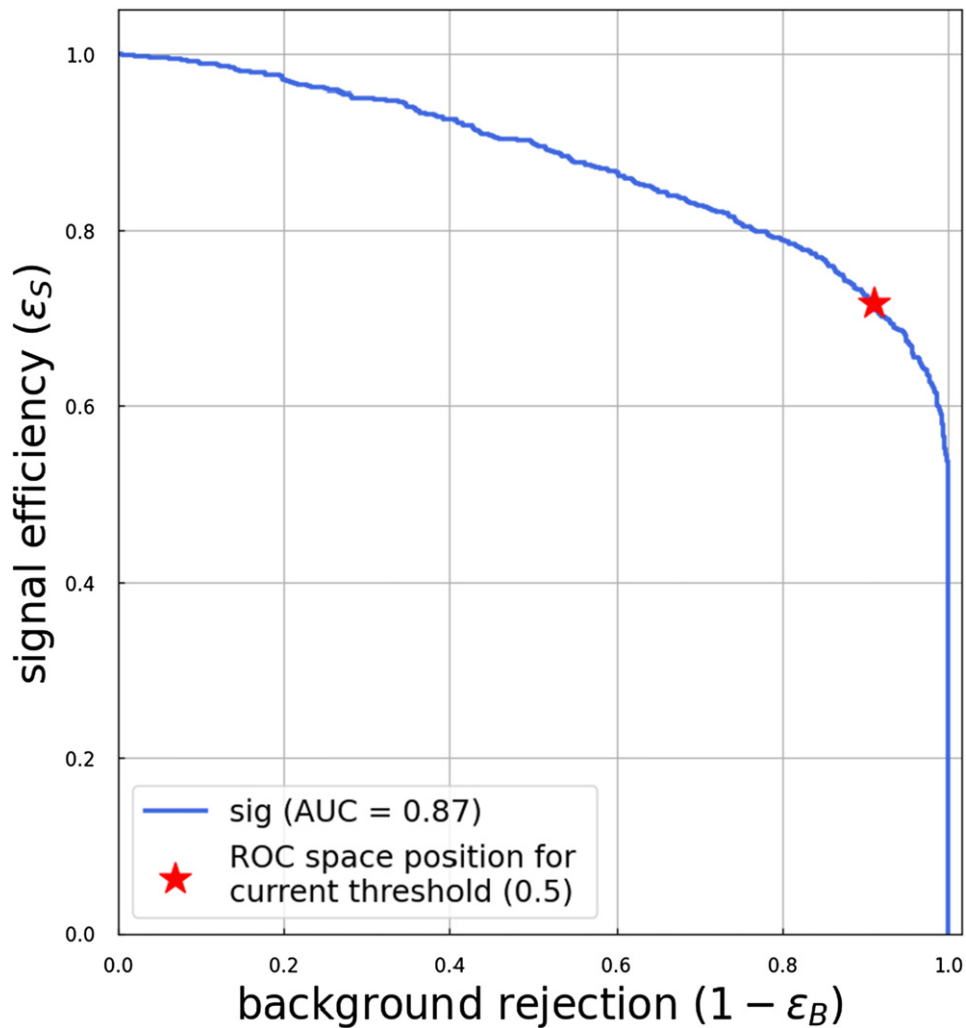


Figure 5. ROC curve for the best-performing classifier. The red star displays the current threshold location on the ROC used for classify the events into signal (score ≥ 0.5) or background (score ≤ 0.5).

networks that we discuss in this section. A summary of the architectures of both networks is reported in table 2.

3.1. Classification network

Our first task is to test whether a DNN can distinguish between possible signal events over a random background. For this, we choose a ResNet [70], which consists of a DNN built as blocks of convolutional layers together with shortcut connections (to skip layers) that make them easier to optimize and overcome the ‘vanishing/exploding gradient’ problem (discussed in [71]). In our analysis, we have tested the accuracy of the ResNet with an increasing number of layers, namely ResNet-18, ResNet-34, ResNet-50, and ResNet-101, using the single-detector and multiple-detector classification data sets described in section 2. The choice of using this

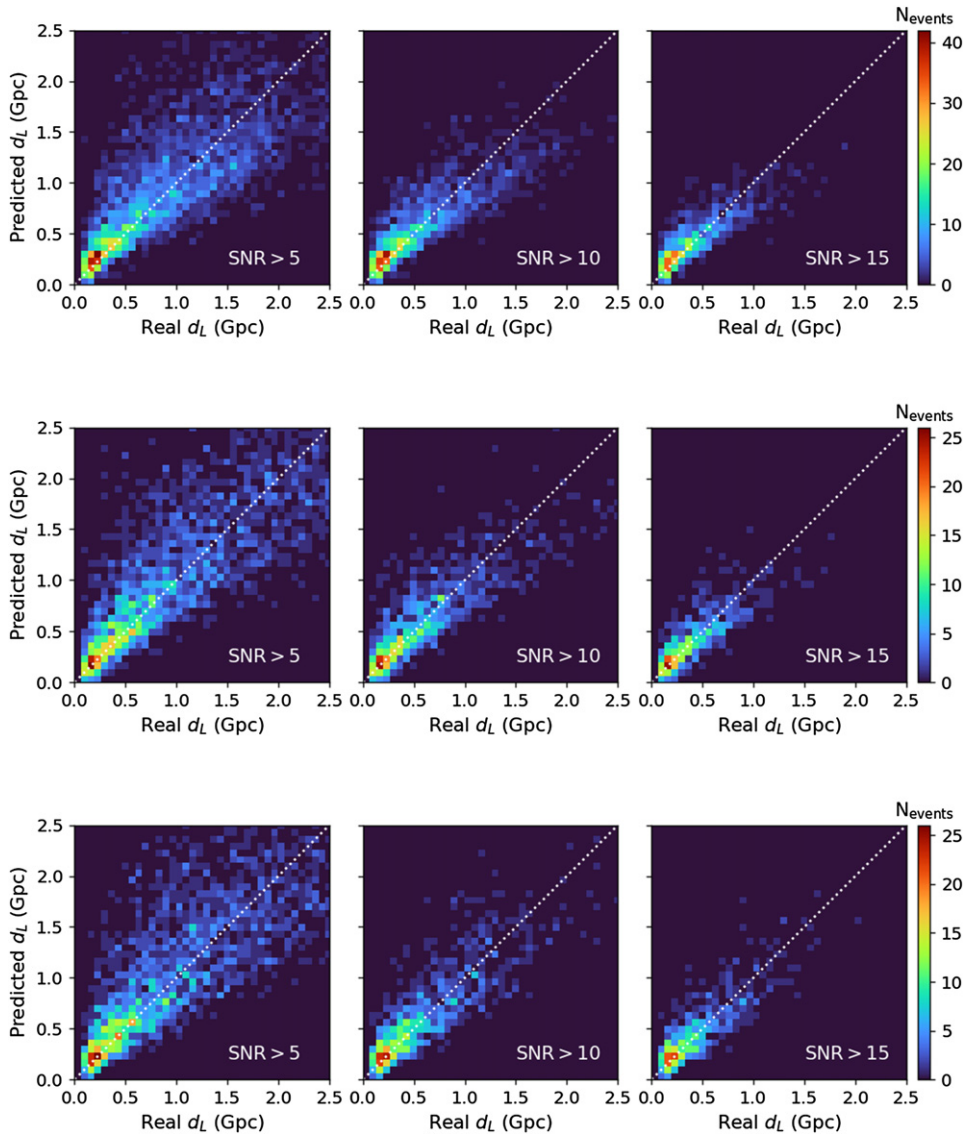


Figure 6. Calibration results for d_L using the different approximant datasets, SEOBNRv4HM_ROM (top), IMRPhenomPv2 (middle) and IMRPhenomD (bottom), and for different SNR thresholds.

architecture comes from the ease of training such networks as well as its well-documented performance in image recognition applications.

For the classification task the highest accuracy was achieved with a ResNet-101 (see table 2), which consists of 101 layers. For our task, we have replaced the last fully connected layers (or the header) of the ResNet-101, responsible for the classification, with the following sequence of layers:

- An adaptive concatenate pooling layer (`AdaptiveConcatPool2d`),
- A flatten layer,

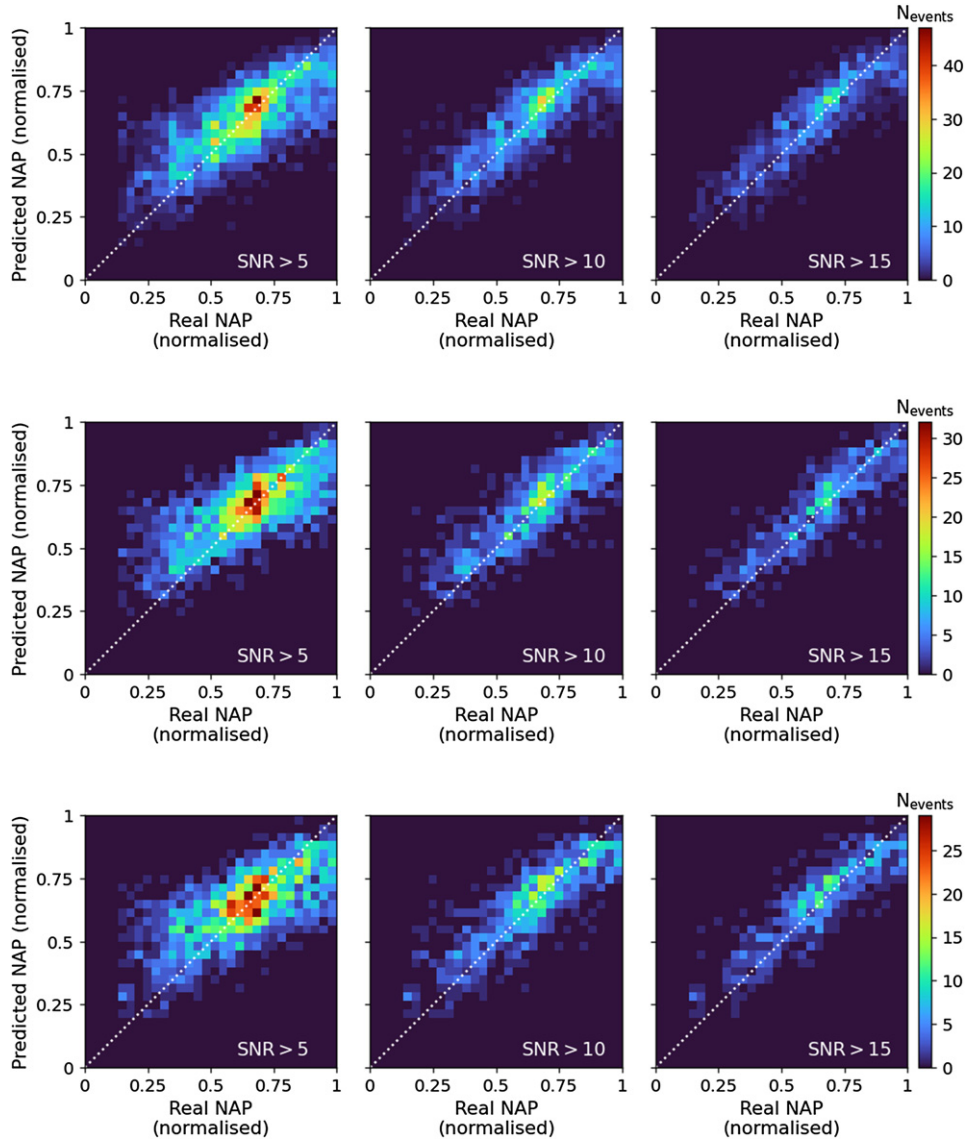


Figure 7. Same as figure 6 but showing the calibration results for the network antenna power.

- A block with batch normalization, dropout, linear, and ReLU layers,
- A dense linear layer with 2 units as outputs, each unit corresponding to *signal* or *background* class and a softmax activation function. This is a necessary step since the fastai ResNet implementation has 1000 output classes by default.

The `AdaptiveConcatPool2d` layer uses adaptive average pooling and adaptive max pooling and concatenates them both. The use of mixed precision (MP) training [72] nearly halves memory requirements and speeds up arithmetic calculations. As the batch size is bounded by the amount of memory available, MP training allows us to use larger batch sizes,

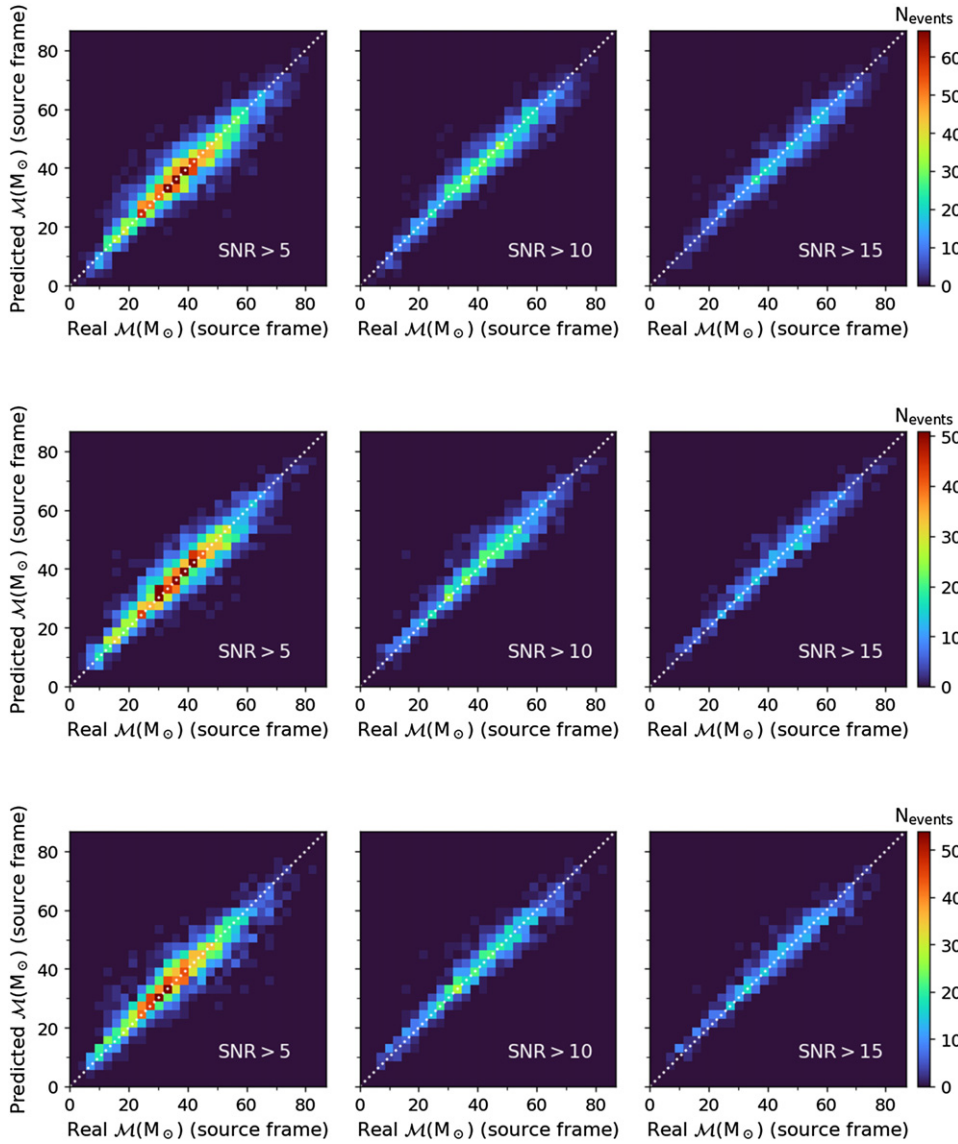


Figure 8. Same as figure 6 but for the chirp mass in the source frame.

which prevents overfitting [73–75]. For each network, we used the maximum possible batch size supported by the hardware.

The learning rate and weight decay are other two key hyperparameters to train DNNs. A good choice of these two parameters can greatly improve the model performance. In our particular case it implies a high accuracy classification and a good background rejection, while drastically reducing the training time. Instead of using a fixed value for the learning rate we opted to use the so called cyclical learning rates (CLR) [74]. To this end one must specify the minimum and maximum learning rate boundaries as well as a step size. While the latter corresponds to the number of iterations used for each step, a cycle consists of two such steps:

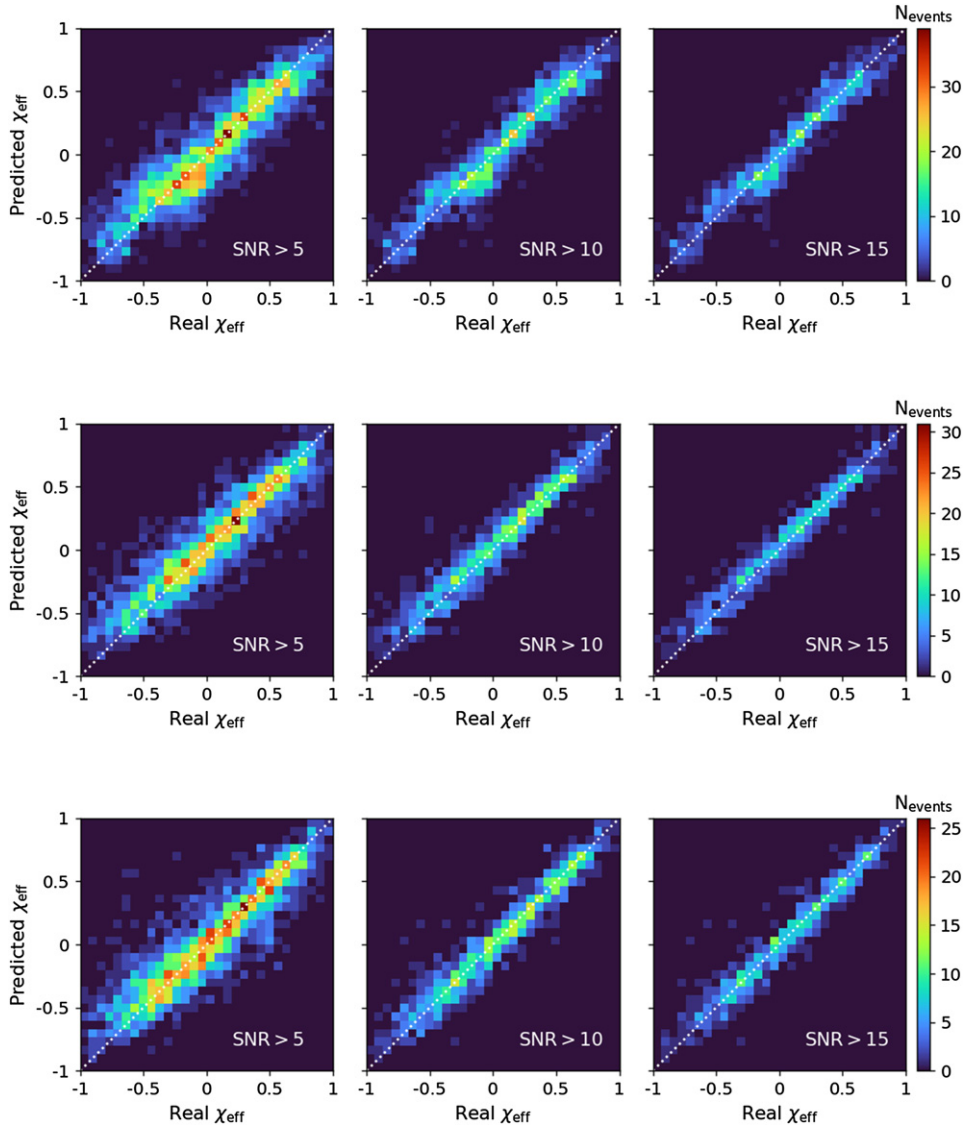


Figure 9. Same as figure 6 but for the effective inspiral spin.

one in which the learning rate increases and the other in which it decreases [74]. Following the guidelines from reference [76], we have performed a scan over a selected range of values for both learning rates and weight decays. According to [76] the best initial values for learning rates are the ones who give the steepest gradient towards the minimum loss value. In our case, we have found it to be 2×10^{-3} for the learning rate and 1×10^{-5} for the weight decay, while for the maximum learning rate value we just multiply the initial value by 10.

3.2. Regression network

We based the regression network architecture on a cross-residual network (\times ResNet; see table 2) [77] and following the guidelines in [78] replaced the average pooling layers with

blur pooling ones. Furthermore, we used dropout layers before pooling, kept activated during test time, a technique known as Monte Carlo (MC) dropout. The effect of MC dropout is equivalent to applying a Bernoulli distribution over the model weights and approximates a Gaussian process [79]. During test time, random realizations of node activations for the same input are translated into a probabilistic distribution in the output values, giving us a simple way of estimating uncertainty. For training, we use once again the CLR, with 1×10^{-2} as the initial value for the learning rate and 1×10^{-3} as the weight decay. It is important to mention that we use the spectrogram images generated from the GW signals to infer continuous values for variables such as the chirp mass \mathcal{M} of the BBH system

$$\mathcal{M} \equiv \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}}, \quad (2)$$

or the luminosity distance of the source d_L . While this approach seems to be rather non-intuitive, CNN's carry inductive biases rooted in translation invariances. Such biases are a direct consequence of the convolutional filters and can be used to extract information from patterns in the spectrogram images and correlated to physical continuous variables.

4. Network assessment

4.1. Classifier

4.1.1. Single detector performance. For the single detector case we find that training with injections at 2000 Mpc retains a 90% classification accuracy⁷ for injections at 100 Mpc, while managing 72% accuracy on the 2000 Mpc validation data. In figure 3 we show contour plots of the scores given to the events in the mass dependence dataset for this last, top-performing model. The signals generated by more massive and closer objects have a higher SNR, therefore easier to classify. Figure 3 illustrates that the signal scores are directly proportional to the component masses and inversely proportional to the luminosity distance. It is worth stressing that the scores increase for GW signals from sources at shorter distances, even when the network is trained with signals at 2 Gpc. This may suggest that, when searching for potential GW signals from BBH mergers with DL networks, the training phase can start by using signals from far away sources.

4.1.2. Multiple detector performance. For the multiple detector case we observe the same trend as in the single detector scenario. However, the performance significantly improves in what concerns the distance of the sources used in the training set. Combining three detectors yields an across-the-board improvement in accuracy of up to 82% figure 4 shows the performance of the best multiple detector model as a function of the binary component masses. Once again, it is noticeable that larger masses and smaller distances result in higher scores. Comparing figures 3 and 4 shows that the multiple detector model yields more confident results, as the region of score > 0.95 (dark blue) is overall more prominent. Figure 5 exhibits the receiver operating characteristic (ROC) curve for our best-performing network, using 2000 Mpc data, showing a value of 0.87 for the area under the curve (AUC)⁸. The x -axis shows the fraction of background-only spectrograms that are successfully rejected by the network, $1 - \varepsilon_B$, while the y -axis represents its efficiency at detecting signals, ε_S . As can be seen in the ROC curve, we

⁷ Accuracy is defined as $\frac{TP+TN}{TP+TN+FP+FN}$, where TP (TN) corresponds to the correctly identified signals (backgrounds) and FP (FN) correspond to incorrectly classified signals (backgrounds).

⁸ An AUC of 1 (0.5) represents a perfect (random) classification.

Table 3. Classifier scores for GWTC-1 marginal detections (left), GWTC-1 confident detections (middle) and GWTC-2 detections (right).

GWTC-1 Confident		GWTC-1 Marginal		GWTC-2			
Event	Score	Event	Score	Event	Score	Event	Score
GW170814	1.00	MC151116	0.73	GW190521	1.00	GW190708_232457	0.98
GW150914	1.00	MC161217	0.72	GW190602_175927	1.00	GW190909_114149	0.97
GW170823	1.00	MC170705	0.51	GW190424_180648	1.00	GW190514_065416	0.96
GW170104	1.00	MC170630	0.49	GW190620_030421	1.00	GW190814	0.95
GW170729	0.99	MC170219	0.45	GW190503_185404	1.00	GW190521_074359	0.95
GW170809	0.97	MC161202	0.40	GW190727_060333	1.00	GW190731_140936	0.92
GW151012	0.96	MC170423	0.35	GW190929_012149	1.00	GW190513_205428	0.92
GW170608	0.92	MC170208	0.33	GW190915_235702	1.00	GW190421_213856	0.87
GW170818	0.88	MC170720	0.30	GW190630_185205	1.00	GW190412	0.81
GW151226	0.87	MC151012A	0.26	GW190519_153544	1.00	GW190728_064510	0.77
—	—	MC151008	0.20	GW190706_222641	1.00	GW190719_215514	0.76
—	—	MC170405	0.14	GW190413_134308	1.00	GW190803_022701	0.66
—	—	MC170616	0.12	GW190701_203306	1.00	GW190930_133541	0.58
—	—	MC170412	0.09	GW190517_055101	1.00	GW190828_065509	0.56
—	—	—	—	GW190408_181802	1.00	GW190924_021846	0.40
—	—	—	—	GW190910_112807	1.00	GW190707_093326	0.35
—	—	—	—	GW190828_063405	0.99	GW190720_000836	0.16
—	—	—	—	GW190413_052954	0.99	—	—
—	—	—	—	GW190512_180714	0.98	—	—
—	—	—	—	GW190527_092055	0.98	—	—

could alter the threshold for classification to be more or less strict, according to the necessities of the problem at hand.

4.2. Regression

4.2.1. Luminosity distance regression. Figure 6 shows the performance of d_L regression for each trained model on its respective validation set. Each event is evaluated 100 times using MC dropout and the mean value of the regressed parameter outputs are stored in the histograms. The white dashed diagonal line shows the ideal behaviour. For the lowest SNR threshold (left columns) the deviation from the ideal behaviour can be quite large. However, as the threshold increases to $\text{SNR} > 10$ and $\text{SNR} > 15$ we are able to more confidently resolve the distances. On the other hand, these thresholds mean that we lose the ability to resolve larger distances for which larger SNRs are also rarer. All three approximants exhibit roughly the same behaviour with the SEOBNRv4HM_ROM dataset showing a slightly tighter distribution.

4.2.2. Network antenna power regression. In figure 7 we show the performance of our model in the regression of the NAP parameter. Again, as for the luminosity distance, all three approximants show roughly the same behaviour. Also note that almost no events with $\text{NAP} < 0.2$ are present, which is to be expected due to the SNR requirements. For the $\text{SNR} > 5$ threshold (left column) there seems to be two separate populations, one that broadly follows the diagonal line and a second one that roughly follows a horizontal line around the 0.6 mark. As we increase the SNR threshold, this second population fades away and we isolate a population of predictions that nicely follows the diagonal line.

4.2.3. Chirp mass regression. Figure 8 shows the behaviour of the predicted source frame chirp masses. It is worth mentioning that the bulk of data points in this figure, as well as in figures 6 and 7, tend to populated the central region of the distributions. The predictions we obtain for the chirp masses nicely follow the actual injected values as the events closely cluster along the diagonal lines in the plots. Of all variables we employ to calibrate our method, the chirp mass is the one that shows the smallest scatter from the ideal results. The distribution for $\text{SNR} > 5$ already displays a fairly low error in the predictions and as we increase the SNR threshold the error further shrinks. Again, the results show almost no dependence on the waveform approximant used.

4.2.4. Effective inspiral spin regression. To end the discussion of the calibration of our model we show in figure 9 the predictions for χ_{eff} compared to the real values. In this case we see that all models also follow closely the ideal diagonal line, with the faithfulness of the distribution width increasing as we raise the SNR threshold.

5. Analysis of real GW detections

Initially, our study comprehended the detections published in the GWTC-1 catalog. When GWTC-2 became publicly available, we extended the analysis to include those detections too. The GW detectors' sensitivity improved in O3 compared to O2, and their detections' SNR differs from the one in our training sets. Therefore, we do not expect optimal results with the current training for the GWTC-2 events.

5.1. Classifier

To analyse the real GW events we produce RGB spectrograms using publicly available data for all GWTC-1 and GWTC-2 BBH events, combining the data from Hanford, Livingston and

Virgo. We leave out the BNS events (GW170817, GW190425 and GW190426) as those cases are not present in our datasets and thus we should not expect the network to recognize them. However, we include GW190814 despite the fact that it involves a $\sim 2.6M_{\odot}$ compact object, since its precise nature remains undetermined [80]. In addition to the confident detections from GWTC-1 and GWTC-2, we also analyze the marginal subthreshold triggers for GWTC-1.

The results of our classifier are presented in table 3. All confident detections reported by the LVC for GWTC-1 are corroborated by our classifier: GW150914, GW170104, GW170814 and GW170823 are all given the score of 1.00, the highest value possible. From the remaining events, the lowest score is 0.87. When we analyse the marginal detections from GWTC-1 we obtain, as expected, much lower scores across the board. Under the standard threshold for detection, which assumes a score of 0.50 or higher, only three events are classified as signal. These are MC151116, MC161217 and MC170705, with scores of 0.73, 0.72 and 0.51 respectively. The first two in particular may deserve a more careful analysis in the future, but this stays outside the scope of this paper. Keeping in mind that the classifier is not optimized for the O3 a run, which has a significantly lower noise floor, we look also at the new BBH events of the GWTC-2 catalog. Despite the lack of any optimization, we find that 34 out of 37 events are given a score above our threshold for detection, and from these, 27 (31) events are given a score above 0.90 (0.70). The highest possible score is obtained for a subset of 16 events.

As an aside remark, it is interesting to note that we have also obtained high scores for signals proposed in alternative GW catalogs [81, 82]. As an example, our method yields a score of 0.75 for GW151216, proposed in [82].

5.2. Parameter inference on GWTC-1

The regression on the BBH events in the LVC catalogs was performed twice. First, with three different networks trained on the three regression datasets, we used MC dropout to pass the spectrogram corresponding to each event to each network, 500 times. We then calculated the mean and standard deviation for the networks' predictions. We realized that with these datasets, low \mathcal{M} events were underrepresented. Therefore, we generated a new dataset using the SEOBNRv4HM_ROM approximant, which was concatenated with the corresponding original dataset, for a total of 31 499 items. A new network was trained on this dataset with a 70/30 train/validation split and new calibration plots were produced. We have checked that with this correction we maintain similar patterns for the calibration plots as presented in subsection 4.2. The spectrograms of the catalogued events were then passed to this new network, 1500 times for each event.

We show the inference results with and without the low-mass datasets contribution to highlight the importance of covering the relevant parameter space. The results in both cases coincide for the high mass events, reinforcing the consistency of our method. Figure 10 shows the comparison between our predictions and the published values by the LVC. As we show below, the consistency of the results obtained with our DL methods with the published data is worth highlighting.

5.2.1. Chirp mass. In the leftmost panel of figure 10 we show the combined results for our three approximants for the chirp mass \mathcal{M} of the GWTC-1 confident BBH detections. The red error bars enclose the MC dropout 3σ range. The results in the top panel are obtained without including the low-mass distribution while those in the bottom panel do include it. For the former we find that the published 90% confidence intervals lie outside our predicted 3σ range for 6 events (without systematic uncertainties taken into account). The most significant discrepancies occur for the GW151226 and GW170608 events where the network seems reluctant to predict low-mass values. We expect this to be related to the under-representation of low chirp

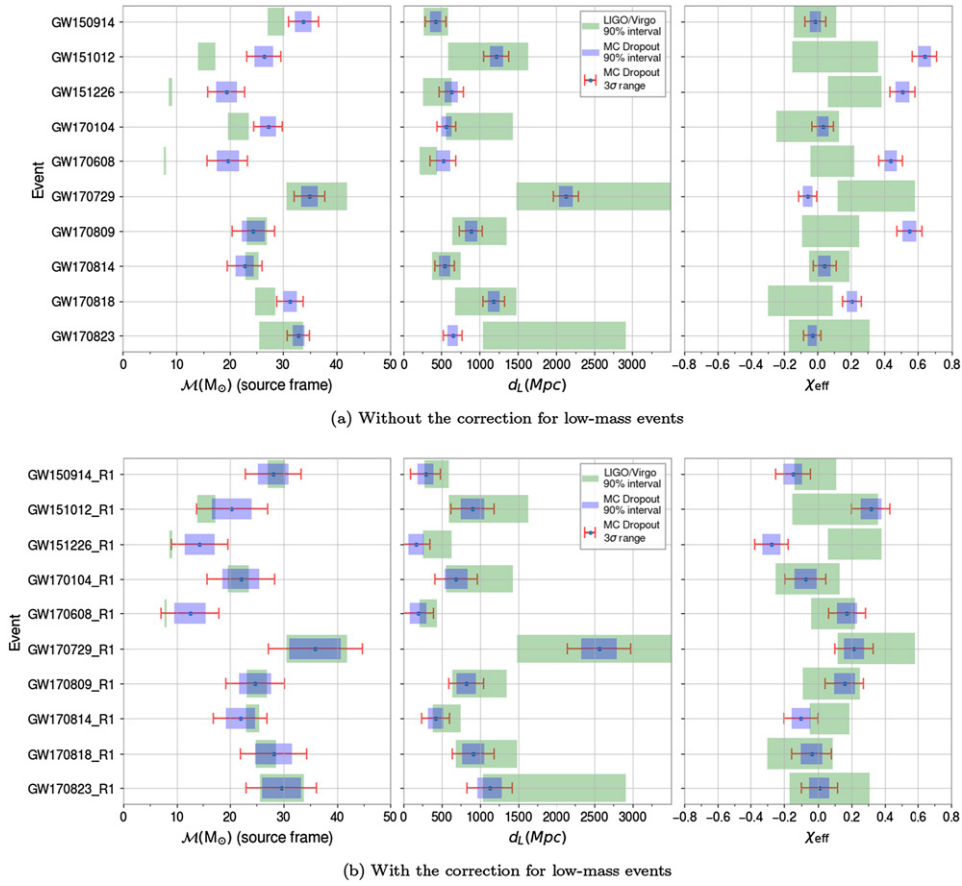


Figure 10. Comparing the parameters of events reported in the gravitation-wave transient catalog (GWTC-1) with the parameters estimated by our algorithm. The parameters include the chirp mass (left), the luminosity distance (middle), and the network for the chirp mass (left), and effective inspiral spin (right). The top panel shows our estimations without a correction (explained in section 5.2) for low-mass events. The bottom panel shows the estimation with the correction for low-mass events, which improved the results.

mass events in the training set. In fact, when the network is trained with the addition of the low-mass dataset, shown in the bottom panel of figure 10, we do indeed observe a considerable improvement. Most of the predictions are now compatible with published data with an uncertainty up to three standard deviations. The one exception is GW151226, where the lower bound on the chirp mass is slightly overestimated by the network when compared to Advanced LIGO results.

5.2.2. Luminosity distance. The top, middle panel of figure 10 displays the combined results for the three approximants without the low-mass distribution, for d_L of the GWTC-1 confident BBH detections. For the luminosity distance most of our predictions, except for GW170823, are compatible with the LIGO/Virgo predictions up to a network uncertainty of 3σ . However, when the network is trained with the addition of the low-mass dataset, as displayed in the



Figure 11. Best-performing DL network’s prediction for the chirp mass (left), luminosity distance (middle) and effective inspiral spin (left) for the BBH events from GWTC-2.

bottom, middle panel of figure 10, all of our predictions become compatible with the LVC values.

5.2.3. Effective inspiral spin. When inferring the effective inspiral spin χ_{eff} using the combined results for the three approximants and without including the low-mass distribution, we find six events with a significant disagreement with published results (in the same sense as discussed previously for the chirp mass) as can be seen in the rightmost, top panel of figure 10. As shown in the corresponding bottom panel of the same figure, when the network is trained with the addition of the low-mass dataset, our results improve. All events, except for GW151226, show compatibility between the LVC 90% confidence interval and the MC dropout 3σ uncertainty.

5.3. Parameter inference on GWTC-2

Finally, we discuss our inference results for the BBH events of GWTC-2. Those are plotted in figure 11 for our best-performing network only, that is, the one trained with the low-mass corrected dataset. As before, we show the parameter inference on the chirp mass, luminosity distance and effective inspiral spin. Without any further optimization of our network, we find that regarding the chirp mass (left panel), the published 90% confidence intervals are compatible with the MC dropout 3σ uncertainty for roughly 90% of the events. While the behaviour for the χ_{eff} predictions tends to also show a reasonable agreement (right panel), the predictions for d_L (middle panel) show a bias towards lower values. We attribute this effect to the usage of O2 noise for the training of the network, as alluded to at the beginning of this section. While this conjecture will be analyzed in future work, the fact that these methods are able to infer some physical parameters from O3 data without any further calibration reinforces the validity of our approach.

6. Conclusions

In this work we have introduced DL methods to study GW from BBH mergers, using spectrograms created from Advanced LIGO and Advanced Virgo open data. By combining data from each of the detectors in the Advanced LIGO and Advanced Virgo network using color channels of RGB images we have shown that the classification procedure improves when compared to a single detector case. For black holes of varying mass and zero spin, we have trained a ResNet classifier on 2000 Mpc data, obtaining a precision of 0.9 and an accuracy of 0.82. This ResNet has been applied on LIGO/Virgo detections and we have corroborated all confident results with high scores, while an analysis of marginal triggers from the O1 and O2 runs has identified 3 cases (MC151116, MC161217 and MC170705) as GW signals, rejecting all others. For GWTC-2 events, despite the lack of any optimization, we have found that 34 out of 37 events pass the threshold for detection.

For the case of black holes of varying mass, spin and sky position, with varying distance, we have trained a \times ResNet to perform parameter estimations on GW spectrogram data from BBH mergers. Using MC dropout we have obtained a natural estimation of the uncertainty of our predictions. We have shown that, at a fledgling level of development, it is possible to successfully perform parameter inference on the distance, chirp mass, antenna power (functioning as a proxy for sky position) and the effective inspiral spin χ_{eff} . The success at resolving this last parameter, especially at high SNR values, shows that our method is sensitive to contributions to the post-Newtonian expansion of the binary system GW radiation up to order 1.5, as this is the first order where a spin-orbit coupling term is observed. Applying this network to spectrogram data from GWTC-1 BBH events, we have found a remarkable agreement with the results published by the LVC in the case of d_L estimations. Most of our chirp mass and effective spin estimations are also compatible with the published 90% confidence intervals up to an MC dropout uncertainty of 3σ , with the exception of GW151226. For GWTC-2 events, again without any optimization, we have found that the published 90% confidence intervals for the chirp mass are compatible with our prediction up to 3σ , for 33 out of 37 BBH events. The behaviour for the χ_{eff} predictions tends to show a reasonable agreement, similar to that of the chirp mass. The predictions for d_L tend to be underestimated, which we suspect is related with the training of the network with injection on O2 noise, which has different characteristics when compared with the O3 run.

While there exist techniques for GW detection and parameter estimation [83], our approach provides the basis for a new pipeline that easily integrates classification and regression tasks. It

is important to stress that we have not carried out a thorough search of network architectures. Going forward, optimizing the architecture, as well as exploring other implementations of Bayesian neural networks, could provide further improvements on our results. Other physical effects of BBH mergers, such as orbital plane precession or eccentricity, may also be explored. Lastly, higher resolution spectrograms, as well as higher colour depth, could in theory be used to increase the sensitivity to smaller effects as well as the predictive power of our tool.

Acknowledgments

We thank Nicolás Sanchis-Gual for very fruitful discussions that allowed to setup the research team involved in this investigation. This work was supported by the Spanish Agencia Estatal de Investigación (PGC2018-095984-B-I00), by the Generalitat Valenciana (PROM-ETEO/2019/071), by the European Union's Horizon 2020 research and innovation (RISE) programme (H2020-MSCA-RISE-2017 Grant No. FunFiCO-777740) and by the Portuguese Foundation for Science and Technology (FCT), project CERN/FIS-PAR/0029/2019. APM and FFF are supported by the FCT project PTDC/FIS-PAR/31000/2017 and by the Center for Research and Development in Mathematics and Applications (CIDMA) through FCT, references UIDB/04106/2020 and UIDP/04106/2020. APM is also supported by the projects CERN/FIS-PAR/0027/2019, CERN/FISPAR/0002/2017 and by national funds (OE), through FCT, IP, in the scope of the framework contract foreseen in the numbers 4, 5 and 6 of the article 23, of the Decree-Law 57/2016, of August 29, changed by Law 57/2017, of July 19.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: <https://arxiv.org/abs/2011.10425>.

ORCID iDs

João D Álvares  <https://orcid.org/0000-0001-5501-9014>
Osvaldo G Freitas  <https://orcid.org/0000-0002-2898-1256>
Solange Nunes  <https://orcid.org/0000-0002-1506-3368>
Antonio Onofre  <https://orcid.org/0000-0003-3471-2703>

References

- [1] Abbott B P *et al* 2016 *Phys. Rev. Lett.* **116** 061102
- [2] Abbott B P *et al* 2016 *Phys. Rev. Lett.* **116** 241103
- [3] LIGO Scientific Collaboration *et al* 2015 *Class. Quantum Grav.* **32** 074001
- [4] Acernese F *et al* (Virgo Collaboration) 2015 *Class. Quantum Grav.* **32** 024001
- [5] Abbott B P *et al* 2017 *Astrophys. J. Lett.* **848** L12
- [6] Abbott B P, Abbott R, Abbott T D, Abraham S, Acernese F, Ackley K, Adams C and Adhikari R e a (LIGO Scientific Collaboration and Virgo Collaboration) 2019 *Phys. Rev. X* **9** 031040
- [7] Abbott R, Abbott T D, Abraham S, Acernese F, Ackley K, Adams C and Adhikari R e a (LIGO Scientific Collaboration and Virgo Collaboration) 2020 arXiv:2010.14527
- [8] Usman S A *et al* 2016 *Class. Quantum Grav.* **33** 215004
- [9] Messick C *et al* 2017 *Phys. Rev. D* **95** 042001 Publisher: American Physical Society
- [10] Adams T, Buskulic D, Germain V, Guidi G M, Marion F, Montani M, Mours B, Piergiorganni F and Wang G 2016 *Class. Quantum Grav.* **33** 175012 Publisher: IOP Publishing

- [11] Blackman J, Field S E, Scheel M A, Galley C R, Ott C D, Boyle M, Kidder L E, Pfeiffer H P and Szilágyi B 2017 *Phys. Rev. D* **96** 024058
- [12] McWilliams S T 2019 *Phys. Rev. Lett.* **122** 191102
- [13] Nagar A et al 2018 *Phys. Rev. D* **98** 104052
- [14] García-Quirós C, Colleoni M, Husa S, Estellés H, Pratten G, Ramos-Buades A, Mateu-Lucena M and Jaume R 2020 *Phys. Rev. D* **102** 064002
- [15] Veitch J et al 2015 *Phys. Rev. D* **91** 042003
- [16] Ashton G et al 2019 *Astrophys. J. Suppl.* **241** 27
- [17] Green S R, Simpson C and Gair J 2020 arXiv:2002.07656
- [18] Abbott B P, Abbott R, Abbott T D, Abraham S and Acernese F (L S C Kagra Collaboration and VIRGO Collaboration) 2020 *Living Rev. Relativ.* **23** 3
- [19] Gabbard H, Williams M, Hayes F and Messenger C 2018 *Phys. Rev. Lett.* **120** 141103
- [20] Gabbard H, Messenger C, Heng I S, Tonolini F and Murray-Smith R 2019 arXiv:1909.06296
- [21] Wang H, Wu S, Cao Z, Liu X and Zhu J-Y 2020 *Phys. Rev. D* **101** 104003
- [22] Chua A J K and Vallisneri M 2020 *Phys. Rev. Lett.* **124** 041102
- [23] Shen H, Huerta E A, Zhao Z, Jennings E and Sharma H 2019 Deterministic and Bayesian neural networks for low-latency gravitational wave parameter estimation of binary black hole mergers arXiv:1903.01998
- [24] Schmidhuber J 2015 *Neural Netw.* **61** 85
- [25] Rummel P 1989 *Machine Vision for Inspection and Measurement* ed H Freeman (New York: Academic) pp 203–21
- [26] Jain R, Rao A R, Kayaalp A and Cole C 1989 *Machine Vision for Inspection and Measurement* ed H Freeman (New York: Academic) pp 283–314
- [27] Dom B 1989 *Machine Vision for Inspection and Measurement* ed H Freeman (New York: Academic) pp 257–82
- [28] Mammone R 1989 *Machine Vision for Inspection and Measurement* ed H Freeman (New York: Academic) pp 185–201
- [29] Elkins A, Freitas F F and Sanz V 2020 *J. Med. Artif. Intell.* **3** 8
- [30] Alves A and Freitas F F 2019 arXiv:1912.12532
- [31] Freitas F F, Khosa C K and Sanz V 2019 *Phys. Rev. D* **100** 035040
- [32] Csáki C, De Freitas F F, Huang L, Ma T, Perelstein M and Shu J 2019 *J. High Energy Phys.* **JHEP05(2019)132**
- [33] Defferrard M, Perraudin N, Kacprzak T and Sgier R 2019 arXiv:1904.05146
- [34] Hong S et al 2020 *Mon. Not. R. Astron. Soc.* **493** 5972
- [35] Gebhard T D, Kilbertus N, Harry I and Schölkopf B 2019 *Phys. Rev. D* **100** 063015
- [36] Lin Y-C and Wu J-H P 2020 arXiv:2007.04176
- [37] Sadeh I 2020 *Astrophys. J.* **894** L25
- [38] Biswas R et al 2013 *Phys. Rev. D* **88** 062003
- [39] Powell J, Trifirò D, Cuoco E, Heng I S and Cavaglià M 2015 *Class. Quantum Grav.* **32** 215012
- [40] Powell J, Torres-Forné A, Lynch R, Trifirò D, Cuoco E, Cavaglià M, Heng I S and Font J A 2017 *Class. Quantum Grav.* **34** 034002
- [41] Razzano M and Cuoco E 2018 *Class. Quantum Grav.* **35** 095016
- [42] Cavaglia M, Staats K and Gill T 2018 *Commun. Comput. Phys.* **25** 963
- [43] George D, Shen H and Huerta E A 2018 *Phys. Rev. D* **97** 101501
- [44] Llorens-Monteaugudo M, Torres-Forné A, Font J A and Marquina A 2019 *Class. Quantum Grav.* **36** 075005
- [45] Coughlin S et al 2019 *Phys. Rev. D* **99** 082002
- [46] Colgan R E, Corley K R, Lau Y, Bartos I, Wright J N, Márka Z and Márka S 2020 *Phys. Rev. D* **101** 102003
- [47] Zevin M et al 2017 *Class. Quantum Grav.* **34** 064003
- [48] Driggers J C, Vitale S, Lundgren A P, Evans M, Kawabe K and Dwyer E a (The LIGO Scientific Collaboration Instrument Science Authors) 2019 *Phys. Rev. D* **99** 042001
- [49] Vajente G, Huang Y, Isi M, Driggers J C, Kissel J S, Szczepańczyk M J and Vitale S 2020 *Phys. Rev. D* **101** 042003
- [50] Torres-Forné A, Cuoco E, Font J A and Marquina A 2020 *Phys. Rev. D* **102** 023011
- [51] Torres A, Marquina A, Font J A and Ibáñez J M 2014 *Phys. Rev. D* **90** 084029
- [52] Torres-Forné A, Cuoco E, Marquina A, Font J A and Ibáñez J M 2018 *Phys. Rev. D* **98** 084013
- [53] Torres-Forné A, Marquina A, Font J A and Ibáñez J M 2016 *Phys. Rev. D* **94** 124040

- [54] Shen H, George D, Huerta E A and Zhao Z 2019 *ICASSP 2019—2019 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)* pp 3237–41
- [55] Wei W and Huerta E A 2020 *Phys. Lett. B* **800** 135081
- [56] Astone P, Cerdá-Durán P, Di Palma I, Drago M, Muciaccia F, Palomba C and Ricci F 2018 *Phys. Rev. D* **98** 122002
- [57] Chan M L, Heng I S and Messenger C 2019 arXiv:1912.13517
- [58] Cavaglià M, Gaudio S, Hansen T, Staats K, Szczepańczyk M and Zanolin M 2020 *Mach. Learn.: Sci. Technol.* **1** 015005
- [59] Miller A L et al 2019 *Phys. Rev. D* **100** 062005
- [60] Beheshtipour B and Papa M A 2020 *Phys. Rev. D* **101** 064009
- [61] Morawski F, Bejger M and Ciecieląg P 2020 *Mach. Learn.: Sci. Technol.* **1** 025016
- [62] Bayley J, Messenger C and Woan G 2020 arXiv:2007.08207
- [63] Nitz A et al 2020 *gwastro/pycbc: Pycbc release v1.16.11*
- [64] Macleod D, Urban A L, Coughlin S, Massinger T, Pitkin M, Altin P, Areeda J, Quintero E and Leinweber K 2019 GWpy: Python package for studying data from gravitational-wave detectors arXiv:1912.016
- [65] Howard J et al 2018 fastai <https://github.com/fastai/fastai>
- [66] Cotesta R, Marsat S and Pürrer M 2020 *Phys. Rev. D* **101** 124040
- [67] Hannam M, Schmidt P, Bohé A, Haegel L, Husa S, Ohme F, Pratten G and Pürrer M 2014 *Phys. Rev. Lett.* **113** 151101
- [68] Husa S, Khan S, Hannam M, Pürrer M, Ohme F, Forteza X J and Bohé A 2016 *Phys. Rev. D* **93** 044006
- [69] Bohé A et al 2017 *Phys. Rev. D* **95** 044028
- [70] He K, Zhang X, Ren S and Sun J 2015 arXiv:1512.03385
- [71] Bengio Y, Simard P and Frasconi P 1994 *IEEE Trans. Neural Netw.* **5** 157
- [72] Micikevicius P et al 2017 arXiv:1710.03740
- [73] Smith S L, Kindermans P and Le Q V 2017 arXiv:1711.00489
- [74] Smith L N 2015 arXiv:1506.01186
- [75] Akiba T, Suzuki S and Fukuda K 2017 arXiv:abs/1711.04325
- [76] Smith L N 2018 arXiv:abs/1803.09820
- [77] He T, Zhang Z, Zhang H, Zhang Z, Xie J and Li M 2018 arXiv:1812.01187
- [78] Zhang R 2019 arXiv:1904.11486
- [79] Gal Y and Ghahramani Z 2015 arXiv:1506.02158
- [80] Abbott R, Abbott T D, Abraham S, Acernese F, Ackley K, Adams C, Adhikari R X and Adya V B (LIGO Scientific Collaboration, and Virgo Collaboration) 2020 *Astrophys. J.* **896** L44
- [81] Nitz A H, Capano C, Nielsen A B, Reyes S, White R, Brown D A and Krishnan B 2019 *Astrophys. J.* **872** 195
- [82] Zackay B, Venumadhav T, Dai L, Roulet J and Zaldarriaga M 2019 *Phys. Rev. D* **100** 023007
- [83] Abbott B P et al 2020 *Class. Quantum Grav.* **37** 055002