Physics Letters B 840 (2023) 137850

Contents lists available at ScienceDirect

Physics Letters B

journal homepage: www.elsevier.com/locate/physletb

Deep learning detection and classification of gravitational waves from neutron star-black hole mergers

Richard Qiu [a,b,c], Plamen G. Krastev [d,e,*], Kiranjyot Gill [a,e], Edo Berger [a,e]

[a] Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138-1516, USA
[b] Department of Physics, Harvard University, 17 Oxford Street Cambridge, MA 02138, USA
[c] John A. Paulson School of Engineering and Applied Sciences, Harvard University, 150 Western Ave, Allston, MA 02134, USA
[d] Faculty of Arts and Sciences Research Computing, Harvard University, 52 Oxford Street, Cambridge, MA 02138, USA
[e] The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, USA

ARTICLE INFO

Article history:
Received 30 October 2022
Received in revised form 6 March 2023
Accepted 15 March 2023
Available online 16 March 2023
Editor: B. Balantekin

ABSTRACT

The Laser Interferometer Gravitational-Wave Observatory (LIGO) and Virgo Interferometer Collaborations have now detected all three classes of compact binary mergers: binary black hole (BBH), binary neutron star (BNS), and neutron star-black hole (NSBH). For coalescences involving neutron stars, the simultaneous observation of gravitational and electromagnetic radiation produced by an event, has broader potential to enhance our understanding of these events, and also to probe the equation of state (EOS) of dense matter. However, electromagnetic follow-up to gravitational wave (GW) events requires rapid real-time detection and classification of GW signals, and conventional detection approaches are computationally prohibitive for the anticipated rate of detection of next-generation GW detectors. In this work, we present the first deep learning based results of classification of GW signals from NSBH mergers in *real* LIGO data. We show for the first time that a deep neural network can successfully distinguish all three classes of compact binary mergers and separate them from detector noise. Specifically, we train a convolutional neural network (CNN) on $\sim 500,000$ data samples of real LIGO noise with injected BBH, BNS, and NSBH GW signals, and we show that our network has high sensitivity and accuracy. Most importantly, we successfully recover the two confirmed NSBH events to-date (GW191219 and GW200115) and the two confirmed BNS mergers to-date (GW170817 and GW190425), together with $\sim 90\%$ of all BBH candidate events from the third Gravitational Wave Transient Catalog, GWTC-3. These results are an important step towards low-latency real-time GW detection, enabling multi-messenger astronomy.

1. Introduction

The first gravitational wave (GW) detection on 2015 September 14 [1] by the advanced Laser Interferometer Gravitational-Wave Observatory (LIGO) and Virgo Collaboration [2,3] ushered in a new era of GW Astrophysics. During the first (O1) and second (O2) observing runs the LIGO and Virgo collaborations reported eleven GW signals [4] from compact binary mergers, which included ten binary black-hole (BBH) events and one clear binary neutron star (BNS) merger, GW170817 [5]. The observation of GW170817 in both gravitational and electromagnetic (EM) radiation inaugurated the field of Multi-Messenger Astrophysics (MMA), which uses GWs, EM radiation, cosmic rays, and neutrinos to provide complimentary information about the astrophysical processes and environments of the sources [5,6]. The third observing run (O3) was split into two parts, O3a and O3b, and led to the detection of over 70 new GW events, including one additional BNS merger [7] and for the first time, two NSBH mergers [8,9]. Thus, including the candidates from O3, the recently released third Gravitational-Wave Transient Catalog (GWTC-3) contains over 90 events that include all configurations of compact object mergers [9].

During the upcoming fourth observing run (O4), scheduled to commence in March 2023, and subsequent runs, the detection rates are expected to substantially increase with the greater instrument sensitivity [10,11]. Moreover, with the advent of next-generation GW detectors in the next decade (Einstein Telescope and Cosmic Explorer [12–14]), it is anticipated that millions of events will be detected per year, probing compact object mergers

* Corresponding author at: Faculty of Arts and Sciences Research Computing, Harvard University, 52 Oxford Street, Cambridge, MA 02138, USA.
E-mail addresses: rqiu@college.harvard.edu (R. Qiu), plamenlrastev@fas.harvard.edu (P.G. Krastev), kiranjyot.gill@cfa.harvard.edu (K. Gill), eberger@cfa.harvard.edu (E. Berger).

https://doi.org/10.1016/j.physletb.2023.137850

across the bulk of cosmic history. With the rapid increase in detection rate, the data analysis capabilities and techniques will have to grow and adapt.

At present, the detection of GW signals from compact binary mergers is primarily achieved using conventional matched-filtering methods that rely on large banks of pre-calculated waveform templates [9,15,16]. Each template is a unique combination of a waveform model and source parameters, such as binary component masses and/or spins [17]. To generate a signal-to-noise ratio (SNR) time series, these methods take an inner product between the detector data and each template in the bank. However, the vast parameter space covered by the template bank due to the unknown source parameters, makes these approaches computationally challenging and expensive. As the rate of GW detections increases, unexpected events with unique physical properties will be observed more frequently in the future. Considering additional effects such as eccentricity, precession, and higher order models requires millions of waveform templates to cover the parameter space of all potential signals [18–20], making these methods computationally prohibitive. This is especially the case for GW events involving neutron stars, where the prompt follow-up of electromagnetic (EM) counterparts is crucial. As a result, there is a critical need for more efficient detection and classification algorithms that can overcome the limitations of conventional matched-filtering methods [9,15,16].

In recent times, there has been an upsurge in the application of Deep Learning (DL) approaches [21] in various scientific and technical arenas, to expedite research that would otherwise be computationally demanding and to catalyze scientific discovery [22]. With the aid of GPU computing, these techniques have shown exceptional performance in tasks like image recognition [23] and natural language processing [24]. Furthermore, DL has emerged as a new tool in engineering and scientific applications, supplementing traditional High Performance Computing (HPC), and has led to the evolution of a new field called Scientific Machine Learning [25]. In recent years, there has been a growing interest in applying DL techniques in the field of GW astrophysics (see e.g., Refs. [26–28]). Specifically, the use of Convolutional Neural Network (CNN) algorithms [29] has been pioneered, and has shown promising results in detecting simulated signals from BBH collisions embedded in Gaussian noise with performance that is comparable or even better than that of conventional matched-filtering methods [15,30]. As a result, a growing number of research groups have begun to apply DL algorithms to detect GW BBH events, both in simulated Gaussian noise and realistic LIGO data (see e.g., Refs. [31–40]). These exciting developments demonstrate the potential of DL approaches to transform the field of GW astrophysics and to enhance our understanding of the Universe.

In our previous work we applied, for the first time, a DL approach to detect GW signals from BNS mergers, embedded in both simulated Gaussian noise [41] and real LIGO data [42], and distinguish them from detector noise and BBH signals. In our later work [42] we applied a CNN to successfully recover and classify all eleven GW events from the first public catalog, GWTC-1 [4].

In this article, we extend our detection/classification deep neural network to include the NSBH event category. This allows us to address the detection and classification of GW signals from all compact binary coalescence (CBC) configurations consistently in a unified DL framework. We show for the first time that a neural network can be used to detect GW signals from NSBH mergers embedded in highly non-stationary, non-Gaussian noise. Most importantly, we demonstrate that our DL approach is able to recover all *real* GW events involving neutron stars to date – the two BNS (GW170817 and GW190425) and the two NSBH (GW191219 and GW200115) mergers. These results are an important step towards real-time detection of gravitational waves from BNS and NSBH mergers, enabling prompt follow-up of EM counterparts of these important GW transients and multi-messenger astrophysics.

2. Methods

Following Krastev [41] and Krastev et al. [42], we construct a large data-set of templates of real LIGO noise with injected simulated BBH, BNS, and NSBH waveform signals. Then, we train a CNN to discriminate between these three classes and noise and evaluate its performance on new, unseen injections, as well as on real GW events from GWTC-3 [9].

2.1. Dataset construction

We obtained real LIGO data from the LIGO and Virgo Gravitational Wave Open Science Center [43]. Specifically, we used O2 and O3b data from the LIGO Livingston detector (L1) sampled at 4096 Hz which does not contain known GW events and hardware injections.

To simulate GW CBC signals, we use the LIGO Algorithm Library Suite (LALSuite) [44] to generate BBH, BNS, and NSBH waveforms. In particular, we used the `SEOBNRv4` [17], `TaylorF2` [45], and `SEOBNRv4_ROM_NRTidalv2_NSBH` [46] time domain approximants to generate BBH, BNS, and NSBH waveforms, respectively. For BBH waveforms, we uniformly sample component masses between 2 and 95 $M_{\odot}$ with a maximum mass ratio $m_1/m_2 \leq 10$. For BNS waveforms, we sample component masses uniformly between 1 and 2 $M_{\odot}$. Finally, for NSBH waveforms, we sample NS component masses uniformly between 1 and 2 $M_{\odot}$ and BH component masses uniformly between 2 and 35 $M_{\odot}$.

For all waveforms, we assume both components have zero spin and the binary system has zero eccentricity. For BNS and NSBH waveforms, we use the APR equation of state [47] to obtain the contribution from the tidal deformability parameter Λ of the component neutron star(s); for calculating Λ, see, e.g., Refs. [48] and [49]. We sample waveforms at 4096 Hz for 4 seconds, which we have found is sufficient for achieving strong discrimination between each represented CBC class, and for recovering real GW signals from all CBC configurations. The shorter templates also help to reduce the memory requirements during the neural network training. Both the data and the simulated signals are whitened separately with power spectral density (PSD) computed directly from the raw GW strain data by Welch's method [50]. Whitening of data is an operation of rescaling the noise contribution at each frequency to have equal power [15]. Because whitening is a linear procedure, whitening both parts individually is equivalent to whitening their sum. Subsequently, we position the peak of the waveform uniformly at random between 3.7 and 3.9 s in the template to make the training process robust to moderate time translations in the signal. We scale the injected waveform amplitude to achieve a particular signal-to-noise ratio (SNR), defined as [15]:

$$\rho^2_{\rm opt} = 4 \int_{f_{\rm min}}^{\infty} \frac{\left|\tilde{h}(f)\right|^2}{S_n(f)} df, \tag{1}$$

where $\tilde{h}$ is the frequency domain representation of the GW strain, $S_n(f)$ is the single-sided detector noise PSD, and $f_{\rm min}$ is the frequency of the GW signal at the start of the sample time series. From an astrophysical perspective, rescaling the waveform is equivalent to varying the source distance from the detector.

We generate 480,000 templates for training, 16,000 templates for validation, and 1,600,000 templates for testing, all of which are

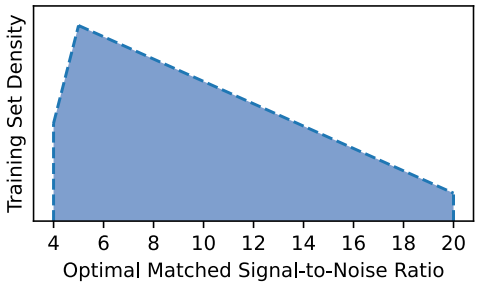

Fig. 1. The distribution from which we sample training signal-to-noise ratios. We emphasize low SNR events in our training set to improve model generalization performance to other low SNR events.

disjoint. Each dataset is approximately 1/4 noise with no event waveform, 1/4 noise + BBH signal, 1/4 noise + BNS signal, and 1/4 noise + NSBH signal. For validation, we sample SNRs uniformly between 5 and 20. For testing, we sample SNRs uniformly between 1 and 20. And for training, to emphasize low SNR detections, we sample SNRs between 4 and 20 using a truncated triangular distribution with lower limit 3, mode 5, and upper limit 27.5 SNR; see Fig. 1. We find empirically that this distribution of SNRs for training improves our sensitivity, accuracy, and ability to recover confirmed events from the LIGO catalog.

2.2. Neural network architecture and training

To choose a CNN architecture, we performed Bayesian hyperparameter optimization over architecture choices and learning rates. The final optimized architecture contained 4 convolutional blocks followed by 2 dense hidden layers and a softmax output. Each convolutional block contains 3 convolutional layers, each followed by a ReLU activation and a batch normalization layer and the entire block ends with a max pooling layer. The filter sizes within a given convolutional block are the same and chosen to be 16, 8, 8, and 8 for each respective convolutional block in the network. Each max pooling layer has filter size 4. The number of filters within a given convolutional block is also constant, and chosen to be 32, 64, 128, and 256. The dense hidden layers have widths 128 and 64. The final softmax output corresponds to the number of predicted classes, 4. The model has 3,441,380 parameters in total. A schematic diagram showing the architecture is provided in Fig. 2.

We built and trained our CNN models using TensorFlow 2.9 [51]. We performed hyperparameter optimization and experiment tracking with Weights and Biases [52]. To train the models, we used the Adam [53] optimizer with AMSgrad [54]. Following hyperparameter optimization, we used $\alpha = 3.986 \times 10^{-3}$ as an initial learning rate, Adam $\beta_1 = 0.1888$, Adam $\beta_2 = 0.9537$, Adam $\epsilon = 1.4975 \times 10^{-3}$, and with a batch size of 256. We used sparse categorical cross-entropy as a loss function and trained with a training budget of 50 epochs. Our final model is taken from the epoch with the lowest validation loss. We also use linear learning rate decay, decaying α by a factor of 0.1 over the first 30 epochs of training. We trained our model using 4 NVidia A100 GPUs with a data-parallel strategy.

3. Results

We evaluate the performance of our model over a large testing dataset of synthetic GW event injections, described in subsection 2.1. We also evaluate the performance of our model on real GW events using the GWTC-3 catalog [9]. This extended catalog includes the two BNS events (GW170817 and GW190425) and the two NSBH events (GW191219 and GW200115). Finally, we compare the performance of the trained DL model with models

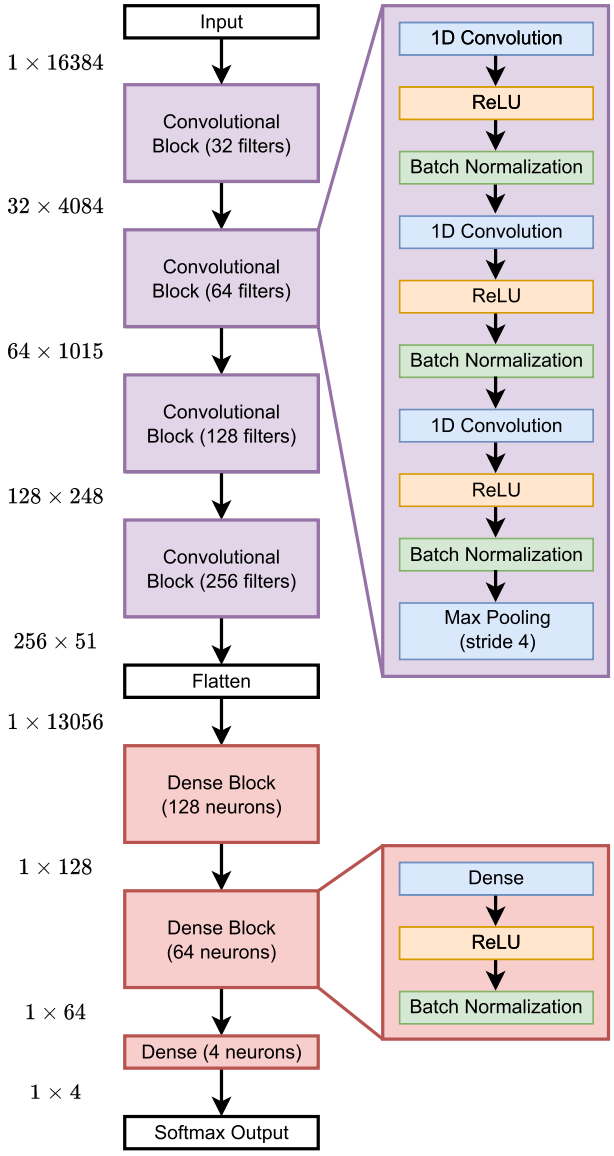

Fig. 2. Schematic diagram of our neural network architecture. All 1D convolution layers within a given convolutional block have the same number of filters. Our model contains 3,441,380 parameters.

from our previous works [41,42] that do not include the NSBH event class.

3.1. Synthetic event detection

Following the same strategy as in our previous works [41,42], we first consider the receiver operating characteristic (ROC) curves for each type of GW signal at a fixed SNR. To determine the ROC curve for a given type of CBC, we calculate the probability of each event in our test dataset being classified as that event type. A ROC curve then displays the proportion of true positives against the proportion of false positives parameterized as a function of the probability threshold to classify a given signal as an event. We calculate the ROC curves with the Python scikit-learn library (https://scikit-learn.org), which constructs empirical ROC curves. An empirical ROC curve shows the relationship between the true alarm probability (TAP) and the false alarm probability (FAP) for various threshold values. Each point on the curve corresponds to a different threshold value. In order to compare different ranking statistics, we can fix the FAP and choose the statistic with the highest

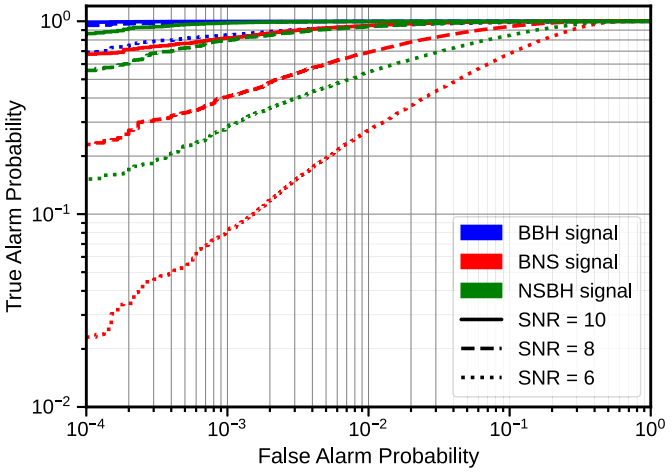

Fig. 3. Receiver operating characteristic curves for all three CBC configurations at fixed signal-to-noise ratios, showing true positive rate and false positive rate, parameterized as a function of detection probability threshold. Note that axes are both shown in logarithmic scale.

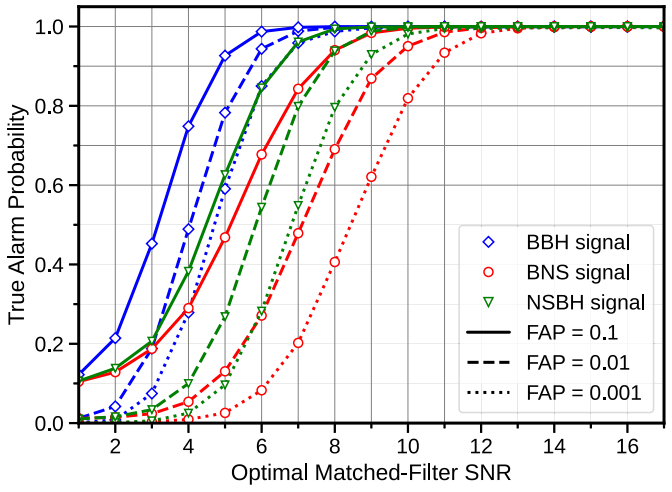

Fig. 4. Sensitivity curves of the demonstrating the true positive rate at fixed signal-to-noise ratios and false positive rates for simulated BBH, BNS, and NSBH signals embedded in real LIGO noise. At $\rho_{\rm opt} \geq 13$, all signals are detected and correctly classified.

TAP (sensitivity) at that FAP. This allows us to assess which statistic performs better overall. We varied the optimal SNR from 1 to 20 in integer steps of 1, and the trained model was applied to inputs with approximately equal fractions of each GW signal class (Noise, BBH Signal, BNS Signal, NSBH Signal).

We show the ROC curves calculated on our synthetic dataset in Fig. 3 for all three event classes at SNRs of 6, 8, and 10. These curves show that our model is most sensitive to BBH events, followed by NSBH events, and least sensitive to BNS events. We also find that our model is more sensitive to both BBH and BNS events than those presented in [41,42]. Note that since the TAP is a function of the FAP, it also reaches a maximal sensitivity for signals with lower optimal SNR at a higher FAP.

We also examined the detection sensitivity of the classifier as a function of the SNR at a fixed FAP, shown in Fig. 4. Similar to ROC curves, we consider the fraction of true positives compared to false positives for all three CBC classes. However, we now plot the portion of true positives as a function of SNR at several representative FAPs (0.1, 0.01, and 0.001). These sensitivity curves represent the ability of the detection CNN to identify and classify GW signals from all three CBC event configurations (BBH, BNS, and NSBH). The lowest FAP used in our analysis translates to a false alarm rate (FAR) of 0.1% or an estimated FAR of $\sim 10^3$ per month.[1] The FAR can be decreased by applying the classifier independently to multiple GW detectors and enforcing coincidence [26,58,59], and also by checking the consistency of the estimated GW source parameters. As before, we observe that our model is the most sensitive to BBH signals, followed by NSBH signals, and finally BNS signals. At a FAP of 0.001, our model saturates the sensitivity curves for BBH events at $\rho_{\rm opt} \geq 8$, for NSBH events at $\rho_{\rm opt} \geq 10$, and for BNS events at $\rho_{\rm opt} \geq 13$. In summary, all curves saturate (at 1) for optimal SNR ≥ 13, which implies that all signals are always detected. Again, we note that, for BBH and BNS signals, our model is more sensitive than the models presented in our previous work [41,42].

3.2. Recovering real GW events

To evaluate the applicability of our model beyond synthetic data, we apply our CNN to real GW strain data containing all events in the GWTC-1 [4], GWTC-2 [55], and GWTC-3 [9] catalogs, which include 82 BBH mergers, 2 BNS mergers, and 2 NSBH mergers. We obtained GW strain data for these events from the LIGO GWOSC through the catalogs provided by PyCBC [60], and preprocessed them following the procedure described in subsection 2.1. For events where multiple detectors were online, we selected the event data from the detector with the highest single-detector SNR.

In Fig. 5, we shown results for all events from the GWTC-1 catalogue, as well as the two BNS mergers, the two confirmed NSBH merger signals, and two potential NSBH candidates. For brevity, in the Figure we omit results for the remaining BBH events from the GWTC-2 and GWTC-3 catalogs. The trained model recovers and classifies successfully all BNS and NSBH events, and 64 of the 72 BBH events (counting GW190814 and GW200210_092254 as BBH) across both catalogs. Most importantly, we show for the first time that a deep neural network can identify *real* GW signals from the NSBH CBC class.

As shown in the Figure, our model successfully identifies all merger events involving neutron stars (BNS and NSBH) to date and distinguish them from BBH events and detector noise with very high confidence. This result is particularly important in the context of the next generation GW detectors where, due to the very high rate of detections, it will be critical to promptly identify events for EM follow-up. Moreover, the results summarized in Fig. 5 illustrate for the first time that GWs from all CBC classes can be detected and classified consistently in a unified DL based framework.

Of the misclassified events, three are from the GWTC-2 catalog (O3a run) and five are from the GWTC-3 catalog (O3b run). These are summarized in Table 1. Of these events, we note that GW190426_152155 is an exceptional event with a large FAR, and eventually demoted to a marginal candidate in a follow-up analysis [56]. GW190814 has an inferred secondary component mass of $\sim 2.6 M_\odot$ and it is possible that the merger is a NSBH rather than a BBH [57]. It is noteworthy that our DL model identifies GW190814 as a NSBH event, despite both components falling within the BBH mass range of $[2-95] M_\odot$. The total mass of the GW190924_021846 BBH merger is merely 13.9 $M_\odot$, making it the BBH coalescence with the lowest mass, and this particular attribute may have led to its erroneous classification as a NSBH merger. Moreover, during the occurrence of GW190924_021846, a moderate glitch was observed, and it was among the signals marked for glitch removal [61], which could have also caused its misclassification.

[1] To estimate the FAR, overlapping time-series segments of duration 0.2 s are used to match the length of the interval in which the signal peak varies. The FAR is calculated from the FAP, which is then converted to false alarms per month.

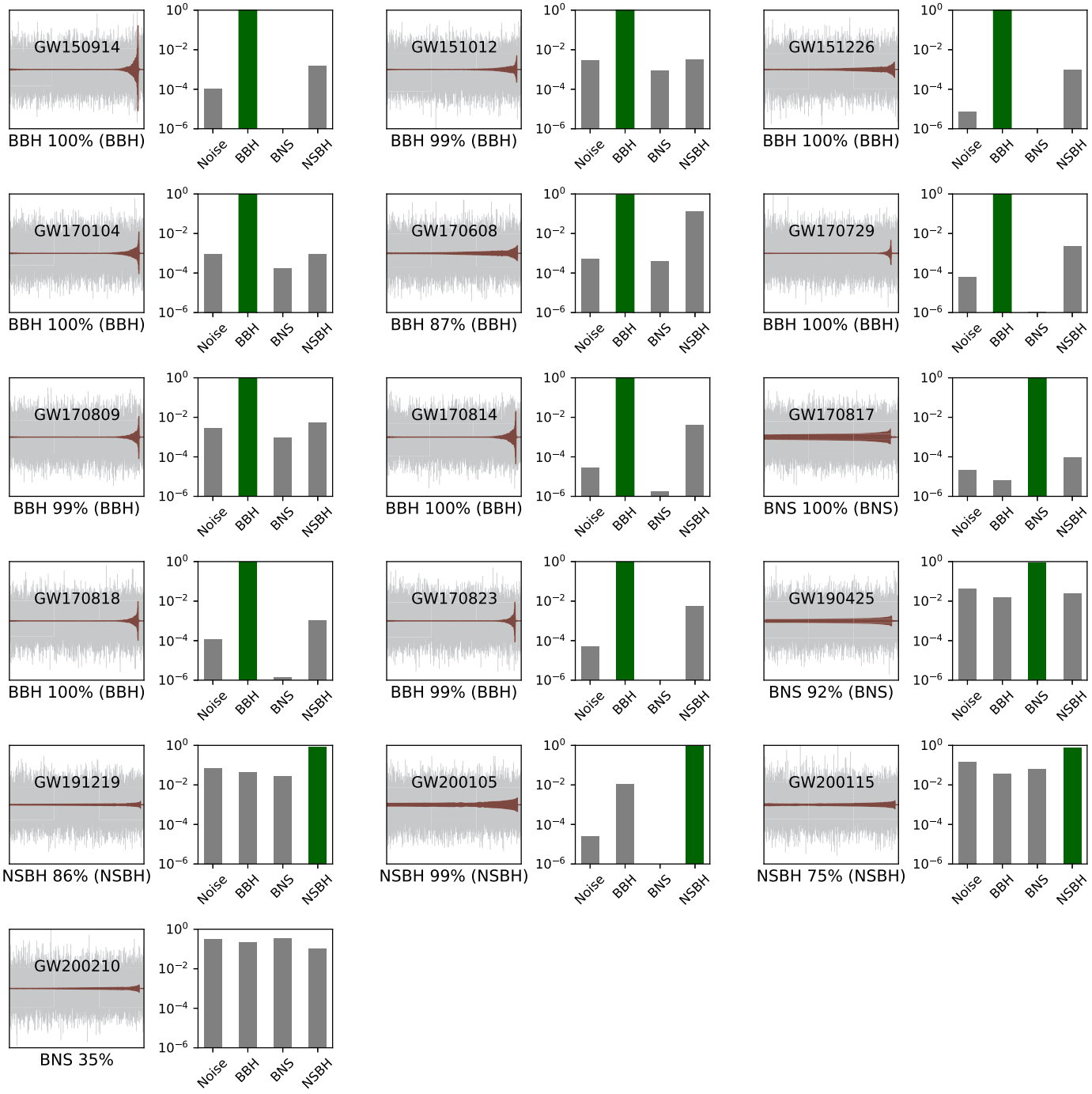

Fig. 5. Detection of GWTC-1 events with our deep neural network. We also include BNS and NSBH (candidate) mergers from GWTC-2 and GWTC-3: GW170817 and GW190425 (BNS event from GWTC-2), GW191219 (NSBH event from GWTC-3), GW200105 (NSBH marginal candidate from GWTC-3), GW200115 (NSBH event from GWTC-3), and GW200210 (BBH or NSBH candidate event from GWTC-3). We show the probability of each event class predicted by our neural network for each GW event. The LIGO/Virgo inferred secondary mass for GW200210 is $2.83^{+0.47}_{-0.42}$, which is out of distribution for both our BNS and NSBH training datasets. Note that these probabilities are shown in logarithmic scale.

Of the misclassified events from O3b, we note that the inferred secondary component mass in the case of GW200210_092254 is $2.83^{+0.47}_{-0.42}$ $M_{\odot}$, which puts the full 90% credible interval outside the neutron star mass distributions of our training datasets, which have neutron star masses up to 2 $M_{\odot}$. As shown in Fig. 5, even though the masses of both binary components of GW200210 fall within the BBH mass range of our training datasets, our DL model cannot accurately classify this system into any specific event class – it only misclassifies it as a BNS event with a confidence level of 35%. Finally, we note that all remaining misclassified events from O3b are BBH merger signals with maximum (across all detection pipelines) single detector SNR < 7.0, and as such are below the maximal sensitivity SNR range of our trained model for BBH signals (see Fig. 4). Thus, it is not surprising that these specific events were misclassified.

3.3. Comparison with other DL models

The DL algorithm presented in this work could be most readily compared with the methods developed in Refs. [15,30,31] as

Table 1
Events of interest or misclassified events in the GWTC-2 and GWTC-3 catalogs. Ref. [55] does not report individual detector SNRs and thus GWTC-2 events (GW190426_152155, GW190814, and GW190924_021846) are missing this data. We note that in follow-up analysis, GW190426_152155 is reclassified as a marginal event [56]. LIGO/Virgo analysis finds that GW190814 is likely a BBH merger though a NSBH merger is also possible [57]; interestingly, our pipeline classifies GW190814 as a NSBH event. All misclassified BBH merger signals in GWTC-3 have single detector SNRs $\rho < 7.0$. Finally, the LIGO/Virgo inferred second component mass of $2.83^{+0.47}_{-0.42}$ $M_{\odot}$ for GW200210_092254 is out-of-distribution for our BNS and NSBH training datasets, which have NS masses up to 2 $M\odot$. Although both binary component masses of this system are within the BBH mass range of our training datasets, the DL model cannot confidently classify it into any specific event class. See text for details.

Event	LIGO/Virgo Inferred Event Type	CNN Inferred Event Type	Maximum Single-Detector SNR	Detector
GW190426_152155	BBH	Noise	—	—
GW190814	BBH/NSBH	NSBH	—	—
GW190924_021846	BBH	NSBH	—	—
GW191103_012549	BBH	Noise	6.9	LIGO Livingston
GW191126_115259	BBH	Noise	6.2	LIGO Livingston
GW200208_222617	BBH	Noise	6.0	LIGO Livingston
GW200210_092254	BBH/NSBH	BNS	8.0	LIGO Livingston
GW200306_093714	BBH	Noise	6.1	LIGO Livingston

they are all employ 1-D CNNs to detect BBH GW signals embedded in both Gaussian and real LIGO noise. In particular, the sensitivity curves for the BBH signals shown in Fig. 4 are very similar to the ones reported in these works. The major difference with the DL algorithms presented in these works is that our algorithm includes also the classification of BNS and NSBH GW signals embedded in real detector noise.

Very recently, Schäfer et al. [40] reported results from the first ML GW search mock data challenge (MLGWSC-1), including results from three algorithms based on the CNN architecture: MFCNN [33], CNN-Conic [15,38], and TPI FSU Jena [39]. The CNN architecture employed by our DL algorithm compares most closely with the algorithms of CNN-Conic and TPI FSU Jena as they both are based on Ref. [15]. The major difference is that these algorithms are specifically trained to detect BBH GW signals, and subsequently applied to data streams from two detectors. On the other hand, our algorithm is designed and trained to detect and classify all CBC event types (BBH, BNS, and NSBH) in a consistent DL framework, but currently still operates on a single-detector data. In order to robustly compare the performance of our DL approach to that of the algorithms from Ref. [40], our DL algorithm needs to be tested on the datasets from MLGWSC-1. However, this is beyond the scope of the current work, and it is left for following articles.

We also compared the performance of the current DL model with models from our previous works [41,42], which did not include the NSBH event class in the training datasets. Previous models that lacked the NSBH classification performed very poorly on test data containing injected NSBH signals where most signals were misclassified as detector noise, or BBH signals. Additionally, the models from our previous works failed on data containing *real* NSBH events (GW191219, GW200105 and GW200115). The NSBH events were misclassified as BBH signals (GW191219 and GW200105) and noise (GW200115) respectively.

In comparison, our current DL model trained on a dataset containing the NSBH GW event type improves dramatically the performance on test data with injected simulated NSBH signals. As seen in Fig. 4, our model identifies 100% of the injected NSBH signals with $SNR \geq 13$ in our test dataset. Moreover, the DL model trained on data containing NSBH templates is able to successfully recover real NSBH events (see Fig. 5). The presented results indicate a significant enhancement in the performance of the current model as compared to the previous models which did not include the NSBH GW event class. These results highlight the significance of the current study, particularly in the context of the precise identification of GW signals arising from CBC events involving neutron stars, where timely observation of the EM counterpart is of utmost importance.

4. Discussion and conclusion

We have demonstrated for the first time that a deep learning algorithm can detect and distinguish GW signals from NSBH mergers in real advanced LIGO data. We have shown that our model achieves high sensitivity on simulated injections of all three CBC classes. Critically, we have also applied the trained neural network to GW events in the GWTC-1, GWTC-2, and GWTC-3 catalogs, and we have shown that our model can recover all high-confidence BNS and NSBH events in these catalogs, as well as 74 of the 82 BBH events; the misclassified BBH signals are all of low SNR. These results are an important step towards a deep learning approach to real-time GW detection from multi-messenger sources, where rapid electromagnetic follow-up is critical.

To further improve fidelity in realistic contexts, upcoming work will need to extend these deep learning algorithms to include realistic glitches and synthesize multiple detector data streams. The results in this work are also fundamental for real-time parameter estimation, where accurate and reliable classification is important. These DL based approaches can be extended to enable rapid parameter estimation, and employ Bayesian networks to quantify model uncertainty.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This research is supported in part by the NSF under grant AST-2108531 and Cooperative Agreement PHY-2019786 (The NSF AI Institute for Artificial Intelligence and Fundamental Interactions http://iafi.org). This work has made use of data, software and/or web tools obtained from the Gravitational Wave Open Science Center (https://www.gw-openscience.org), a service of LIGO Laboratory, the LIGO Scientific Collaboration and the Virgo Collaboration. LIGO is funded by the U.S. National Science Foundation. Virgo is

funded by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale della Fisica Nucleare (INFN) and the Dutch Nikhef, with contributions by Polish and Hungarian institutes. The computations in this paper were run on the FASRC Cannon cluster supported by the FAS Division of Science Research Computing Group at Harvard University.

References

[1] B.P. Abbott, et al., Observation of gravitational waves from a binary black hole merger, Phys. Rev. Lett. 116 (6) (2016) 061102, https://doi.org/10.1103/PhysRevLett.116.061102, arXiv:1602.03837.
[2] J. Aasi, et al., Advanced LIGO, Class. Quantum Gravity 32 (2015) 074001, https://doi.org/10.1088/0264-9381/32/7/074001, arXiv:1411.4547.
[3] F. Acernese, et al., Advanced Virgo: a second-generation interferometric gravitational wave detector, Class. Quantum Gravity 32 (2) (2015) 024001, https://doi.org/10.1088/0264-9381/32/2/024001, arXiv:1408.3978.
[4] B.P. Abbott, et al., GWTC-1: a gravitational-wave transient catalog of compact binary mergers observed by LIGO and Virgo during the first and second observing runs, Phys. Rev. X 9 (3) (2019) 031040, https://doi.org/10.1103/PhysRevX.9.031040, arXiv:1811.12907.
[5] B.P. Abbott, et al., Observation of gravitational waves from a binary neutron star inspiral, Phys. Rev. Lett. 119 (16) (1708) 161101, https://doi.org/10.1103/PhysRevLett.119.161101, arXiv:1710.05832, 2017.
[6] B.P. Abbott, et al., Multi-messenger observations of a binary neutron star merger, Astrophys. J. Lett. 848 (2) (2017) L12, https://doi.org/10.3847/2041-8213/aa91c9, arXiv:1710.05833.
[7] B.P. Abbott, et al., Observation of a compact binary coalescence with total mass $\sim 3.4 M_{\odot}$, Astrophys. J. Lett. 892 (1) (1904) L3, https://doi.org/10.3847/2041-8213/ab75f5, arXiv:2001.01761, 2020.
[8] R. Abbott, et al., Observation of gravitational waves from two neutron star–black hole coalescences, Astrophys. J. Lett. 915 (1) (2021) L5, https://doi.org/10.3847/2041-8213/ac082e, arXiv:2106.15163.
[9] R. Abbott, et al., GWTC-3: compact binary coalescences observed by LIGO and Virgo during the second part of the third observing run, arXiv:2111.03606.
[10] B.P. Abbott, et al., Prospects for observing and localizing gravitational-wave transients with advanced LIGO, advanced Virgo and KAGRA, Living Rev. Relativ. 21 (1) (2018) 3, https://doi.org/10.1007/s41114-020-00026-9, arXiv:1304.0670.
[11] C. Cahillane, G. Mansell, Review of the advanced LIGO gravitational wave observatories leading to observing run four, Galaxies 10 (1) (2022) 36, https://doi.org/10.3390/galaxies10010036, arXiv:2202.00847.
[12] B.P. Abbott, et al., Exploring the sensitivity of next generation gravitational wave detectors, Class. Quantum Gravity 34 (4) (2017) 044001, https://doi.org/10.1088/1361-6382/aa51f4, arXiv:1607.08697.
[13] M. Maggiore, et al., Science case for the Einstein telescope, J. Cosmol. Astropart. Phys. 03 (2020) 050, https://doi.org/10.1088/1475-7516/2020/03/050, arXiv:1912.02622.
[14] M. Evans, et al., A horizon study for cosmic explorer: science, observatories, and community, arXiv:2109.09882.
[15] H. Gabbard, M. Williams, F. Hayes, C. Messenger, Matching matched filtering with deep networks for gravitational-wave astronomy, Phys. Rev. Lett. 120 (14) (2018) 141103, https://doi.org/10.1103/PhysRevLett.120.141103, arXiv:1712.06041.
[16] T. Dal Canton, et al., Implementing a search for aligned-spin neutron star-black hole systems with advanced ground based gravitational wave detectors, Phys. Rev. D 90 (8) (2014) 082004, https://doi.org/10.1103/PhysRevD.90.082004, arXiv:1405.6731.
[17] A. Bohé, et al., Improved effective-one-body model of spinning, nonprecessing binary black holes for the era of gravitational-wave astrophysics with advanced detectors, Phys. Rev. D 95 (4) (2017) 044028, https://doi.org/10.1103/PhysRevD.95.044028, arXiv:1611.03703.
[18] I. Harry, S. Privitera, A. Bohé, A. Buonanno, Searching for gravitational waves from compact binaries with precessing spins, Phys. Rev. D 94 (2) (2016) 024012, https://doi.org/10.1103/PhysRevD.94.024012, arXiv:1603.02444.
[19] I. Harry, J. Calderón Bustillo, A. Nitz, Searching for the full symphony of black hole binary mergers, Phys. Rev. D 97 (2) (2018) 023004, https://doi.org/10.1103/PhysRevD.97.023004, arXiv:1709.09181.
[20] R. Dhurkunde, A.H. Nitz, Sensitivity of spin-aligned searches for neutron star-black hole systems using future detectors, arXiv e-prints arXiv:2207.14645, 2022.
[21] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.
[22] A.M. Deiana, et al., Applications and techniques for fast machine learning in science, Front. Big Data 5 (2022) 787421, https://doi.org/10.3389/fdata.2022.787421, arXiv:2110.13041.
[23] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[24] T. Young, D. Hazarika, S. Poria, E. Cambria, Recent trends in deep learning based natural language processing, IEEE Comput. Intell. Mag. 13 (3) (2018) 55–75.
[25] N. Baker, F. Alexander, T. Bremer, A. Hagberg, Y. Kevrekidis, H. Najm, M. Parashar, A. Patra, J. Sethian, S. Wild, et al., Workshop report on basic research needs for scientific machine learning: Core technologies for artificial intelligence, Tech. Rep., USDOE Office of Science (SC), Washington, DC (United States), 2019.
[26] E.A. Huerta, et al., Accelerated, scalable and reproducible AI-driven gravitational wave detection, Nat. Astron. 5 (10) (2021) 1062–1068, https://doi.org/10.1038/s41550-021-01405-0, arXiv:2012.08545.
[27] E.A. Huerta, Z. Zhao, Advances in machine and deep learning for modeling and real-time detection of multi-messenger sources, arXiv:2105.06479, https://doi.org/10.1007/978-981-15-4702-7_47-1.
[28] E. Cuoco, et al., Enhancing gravitational-wave science with machine learning, Mach. Learn.: Sci. Technol. 2 (1) (2021) 011002, https://doi.org/10.1088/2632-2153/abb93a, arXiv:2005.03745.
[29] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324.
[30] D. George, E.A. Huerta, Deep neural networks to enable real-time multimessenger astrophysics, Phys. Rev. D 97 (4) (2018) 044039, https://doi.org/10.1103/PhysRevD.97.044039, arXiv:1701.00008.
[31] D. George, E.A. Huerta, Deep learning for real-time gravitational wave detection and parameter estimation: results with advanced LIGO data, Phys. Lett. B 778 (2018) 64–70, https://doi.org/10.1016/j.physletb.2017.12.053, arXiv:1711.03121.
[32] T.D. Gebhard, N. Kilbertus, I. Harry, B. Schölkopf, Convolutional neural networks: a magic bullet for gravitational-wave detection?, Phys. Rev. D 100 (6) (2019) 063015, https://doi.org/10.1103/PhysRevD.100.063015, arXiv:1904.08693.
[33] H. Wang, S. Wu, Z. Cao, X. Liu, J.-Y. Zhu, Gravitational-wave signal recognition of LIGO data by deep learning, Phys. Rev. D 101 (10) (2020) 104003, https://doi.org/10.1103/PhysRevD.101.104003, arXiv:1909.13442.
[34] Y.-C. Lin, J.-H.P. Wu, Detection of gravitational waves using Bayesian neural networks, Phys. Rev. D 103 (6) (2021) 063034, https://doi.org/10.1103/PhysRevD.103.063034, arXiv:2007.04176.
[35] M.D. Morales, J.M. Antelis, C. Moreno, A.I. Nesterov, Deep learning for gravitational-wave data analysis: a resampling white-box approach, Sensors 21 (9) (2021) 3174, https://doi.org/10.3390/s21093174, arXiv:2009.04088.
[36] H. Xia, L. Shao, J. Zhao, Z. Cao, Improved deep learning techniques in gravitational-wave data analysis, Phys. Rev. D 103 (2) (2021) 024040, https://doi.org/10.1103/PhysRevD.103.024040, arXiv:2011.04418.
[37] M.B. Schäfer, F. Ohme, A.H. Nitz, Detection of gravitational-wave signals from binary neutron star mergers using machine learning, Phys. Rev. D 102 (6) (2020) 063015, https://doi.org/10.1103/PhysRevD.102.063015, arXiv:2006.01509.
[38] M.B. Schäfer, A.H. Nitz, From one to many: a deep learning coincident gravitational-wave search, Phys. Rev. D 105 (4) (2022) 043003, https://doi.org/10.1103/PhysRevD.105.043003, arXiv:2108.10715.
[39] M.B. Schäfer, O. Zelenka, A.H. Nitz, F. Ohme, B. Brügmann, Training strategies for deep learning gravitational-wave searches, Phys. Rev. D 105 (4) (2022) 043002, https://doi.org/10.1103/PhysRevD.105.043002, arXiv:2106.03741.
[40] M.B. Schäfer, et al., First machine learning gravitational-wave search mock data challenge, Phys. Rev. D 107 (2) (2023) 023021, https://doi.org/10.1103/PhysRevD.107.023021, arXiv:2209.11146.
[41] P.G. Krastev, Real-time detection of gravitational waves from binary neutron stars using artificial neural networks, Phys. Lett. B 803 (2020) 135330, https://doi.org/10.1016/j.physletb.2020.135330, arXiv:1908.03151.
[42] P.G. Krastev, K. Gill, V.A. Villar, E. Berger, Detection and parameter estimation of gravitational waves from binary neutron-star mergers in real LIGO data using deep learning, Phys. Lett. B 815 (2021) 136161, https://doi.org/10.1016/j.physletb.2021.136161, arXiv:2012.13101.
[43] R. Abbott, et al., Open data from the first and second observing runs of advanced LIGO and advanced Virgo, SoftwareX 13 (2021) 100658, https://doi.org/10.1016/j.softx.2021.100658, arXiv:1912.11716.
[44] LIGO Scientific Collaboration, LALSuite: LIGO Scientific Collaboration Algorithm Library Suite, https://doi.org/10.7935/GT1W-FZ16, 2018.
[45] F. Messina, R. Dudi, A. Nagar, S. Bernuzzi, Quasi-5.5PN TaylorF2 approximant for compact binaries: point-mass phasing and impact on the tidal polarizability inference, Phys. Rev. D 99 (12) (2019) 124051, https://doi.org/10.1103/PhysRevD.99.124051, arXiv:1904.09558.
[46] T. Dietrich, et al., Matter imprints in waveform models for neutron star binaries: tidal and self-spin effects, Phys. Rev. D 99 (2) (2019) 024029, https://doi.org/10.1103/PhysRevD.99.024029, arXiv:1804.02235.
[47] A. Akmal, V.R. Pandharipande, D.G. Ravenhall, Equation of state of nucleon matter and neutron star structure, Phys. Rev. C 58 (1998) 1804–1828, https://doi.org/10.1103/PhysRevC.58.1804, https://link.aps.org/doi/10.1103/PhysRevC.58.1804.
[48] T. Hinderer, B.D. Lackey, R.N. Lang, J.S. Read, Tidal deformability of neutron stars with realistic equations of state and their gravitational wave signatures in binary inspiral, Phys. Rev. D 81 (2010) 123016, https://doi.org/10.1103/PhysRevD.81.123016, arXiv:0911.3535.

[49] P.G. Krastev, B.-A. Li, Imprints of the nuclear symmetry energy on the tidal deformability of neutron stars, J. Phys. G 46 (7) (2019) 074001, https://doi.org/10.1088/1361-6471/ab1a7a, arXiv:1801.04620.

[50] P. Welch, The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms, IEEE Trans. Audio Electroacoust. 15 (2) (1967) 70–73.

[51] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng, TensorFlow: large-scale machine learning on heterogeneous systems, software available from tensorflow.org, https://www.tensorflow.org/, 2015.

[52] L. Biewald, Experiment tracking with weights and biases, software available from wandb.com, https://www.wandb.com/, 2020.

[53] D.P. Kingma, J. Ba Adam, A method for stochastic optimization, arXiv preprint arXiv:1412.6980.

[54] S.J. Reddi, S. Kale, S. Kumar, On the convergence of Adam and beyond, in: International Conference on Learning Representations, 2018, https://openreview.net/forum?id=ryQu7f-RZ.

[55] R. Abbott, et al., GWTC-2: compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run, Phys. Rev. X 11 (2021) 021053, https://doi.org/10.1103/PhysRevX.11.021053, arXiv:2010.14527.

[56] R. Abbott, et al., GWTC-2.1: deep extended catalog of compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run, arXiv:2108.01045.

[57] R. Abbott, et al., Gravitational waves from the coalescence of a 23 solar mass black hole with a 2.6 solar mass compact object, Astrophys. J. Lett. 896 (2) (1908) L44, https://doi.org/10.3847/2041-8213/ab960f, arXiv:2006.12611, 2020.

[58] W. Wei, A. Khan, E.A. Huerta, X. Huang, M. Tian, Deep learning ensemble for real-time gravitational wave detection of spinning binary black hole mergers, Phys. Lett. B 812 (2021) 136029, https://doi.org/10.1016/j.physletb.2020.136029, arXiv:2010.15845.

[59] P. Chaturvedi, A. Khan, M. Tian, E.A. Huerta, H. Zheng, Inference-optimized AI and high performance computing for gravitational wave detection at scale, Front. Artif. Intell. 5 (2022) 828672, https://doi.org/10.3389/frai.2022.828672, arXiv:2201.11133.

[60] A. Nitz, I. Harry, D. Brown, C.M. Biwer, J. Willis, T.D. Canton, C. Capano, T. Dent, L. Pekowsky, A.R. Williamson, S. De, M. Cabero, B. Machenschalk, D. Macleod, P. Kumar, F. Pannarale, S. Reyes, G.S.C. Davies dfinstad, S. Kumar, M. Tápai, L. Singer, S. Khan, S. Fairhurst, A. Nielsen, S. Singh, T. Massinger, K. Chandra, Shasvath, Veronica Villa, gwastro/pycbc: v2.0.5 release of pycbc, https://doi.org/10.5281/zenodo.6912865, Jul. 2022.

[61] N.J. Cornish, Rapid and robust parameter inference for binary mergers, Phys. Rev. D 103 (10) (2021) 104057, https://doi.org/10.1103/PhysRevD.103.104057, arXiv:2101.01188.