Machine Learning Applications in Gravitational Wave Astronomy

Nikolaos Stergioulas

Abstract Gravitational wave astronomy has emerged as a new branch of observational astronomy, since the first detection of gravitational waves in 2015. The current number of O(100) detections is expected to grow by several orders of magnitude over the next two decades. As a result, current computationally expensive detection algorithms will become impractical. A solution to this problem, which has been explored in the last years, is the application of machine-learning techniques to accelerate the detection and parameter estimation of gravitational wave sources. In this chapter, several different applications are summarized, including the application of artificial neural networks and autoenconders in accelerating the computation of surrogate models, deep residual networks in achieving rapid detections with high sensitivity, as well as artificial neural networks for accelerating the construction of neutron star models in an alternative theory of gravity.

1 Introduction

Since 2015, when the first gravitational waves (GWs) from a binary black hole (BBH) system were detected [1], GW detections have become increasingly common, moving closer to the point of being a regular occurrence. After the third observing run (O3), the most recent catalog (GWTC-3, [2]) from the Advanced LIGO [3], Advanced Virgo [4] KAGRA [5, 6] collaboration contained 90 GW events, almost all of which were BBH mergers. The 4th observing run (O4) is currently underway and a larger number of BBH detections are expected [7]. The addition of a fifth interferometer, LIGO-India [8], is expected to significantly enhance both the sensitivity and the sky localization of the network. Moreover, third-generation ground-based detectors such as the Einstein Telescope [9, 10] and Cosmic Explorer [11, 12] are currently

Nikolaos Stergioulas

Department of Physics, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece e-mail: niksterg@auth.gr

being developed and are anticipated to greatly expand our understanding of the astrophysical processes in the Universe [13, 14, 15].

The advances in GW astronomy described above were made possible by collaborative efforts in multiple areas. Accurate descriptions of the entire coalescence, including the full inspiral, merger, and ringdown, can be obtained in different ways, with IMRPhenomXPHM [16] and SEOBNRv5PHM [17] being two examples of waveform models. Recent implementations of these models take into account the spin-induced precession of the binary orbit and contributions from both the dominant and subdominant multipole moments of the emitted gravitational radiation. However, the increased complexity of the waveforms increases their computational cost.

Astronomical observations have enabled a number of attempts to determine the Equation of State (EoS) of Neutron Stars (NSs). These include the NICER mass and radius measurements [18, 19, 20], the measurement of tidal deformability through gravitational waves [21, 22, 23, 24], as well as joint constraints, e.g., [25, 26, 27, 28]. In particular, the detection of the binary NS merger GW170817 [29, 30] has prompted further research in this area.

In recent years, there has been an increase in the utilization of machine learning approaches for the analysis of gravitational wave data (see [31, 32, 33] for reviews). This chapter provides a summary of different machine learning applications to gravitational-wave astronomy presented in [34, 35, 36, 37].

2 ANN-Accelerated Surrogate Models

Surrogate modeling has been provided to reduce the considerable computational cost of evaluating waveform models [38, 39], which can significantly speed up EOB waveforms (e.g. [38, 40, 41, 42, 43]) while still providing high accuracy within its valid parameter range. The SEOBNRv4 model has a three-dimensional parameter space λ ; the mass ratio q between the two black holes and their spins χ_1 and χ_2 , assuming that they are aligned with the orbital angular momentum. A surrogate model for this waveform family was presented in [44]. Several machine learning techniques can be used to interpolate or fit the projection coefficients of a reduced basis representation of time-domain waveforms, and the most suitable method depends on the desired accuracy and dimensionality. For low-dimensional parameter spaces, interpolation is a viable option. However, as the dimensionality increases, interpolation becomes difficult due to the large number of data points usually needed. Artificial Neural Networks (ANNs) are proposed as a solution to estimate these coefficients since this approach allows for efficient execution on either a CPU or GPU.

In [34], it was observed that the residual errors after training an ANN to evaluate the coefficients of the surrogate model for the SEOBNRv4 model had a pattern with respect to the input parameters. It was then demonstrated that a second neural network could be trained to model these errors, leading to an improved method, in which the maximum mismatch between SEOBNRv4 waveforms and waveforms

generated by the new surrogate model was more than one order of magnitude smaller than the baseline method. Here, we will provide a summary of the steps taken to create the surrogate model and the residual ANN network to accelerate the evaluation of its coefficients, as described in [34].

2.1 Constructing a surrogate model

We express the complex gravitational wave strain as $h(t;\lambda) = h_+(t;\lambda) - ih_\times(t;\lambda)$, where h_+ and h_\times are the two independent polarizations [45], t is the time, and λ is a vector of intrinsic parameters. The SEOBNRv4 model [46] has a three-dimensional parameter space, with each waveform characterized by the mass ratio q (the ratio of the masses of the two black holes) and the dimensionless spins χ_1, χ_2 of the two black holes. Surrogate modeling is a process of approximating given signals using a reduced model, denoted $h_s(t;\lambda)$, such that the approximation given by the surrogate model, $h_s(t;\lambda)$, accurately reconstructs the actual waveform $h(t;\lambda)$ within a preset threshold of error. When considering only the dominant, quadrupole (l=m=2) mode [45], the target becomes $h_s(t;\lambda) \approx h_{2,2}(t;\lambda)$ where l,m are the spherical harmonics. To begin the surrogate modeling process, a training set of N waveforms $\{h_i(t;\lambda_i)\}_{i=1}^N$ is created, where $\lambda_i = (q,\chi_1,\chi_2)_i$. The mass ratio is limited to a predetermined interval, such as $1 \le q \le 8$, within which the surrogate model is designed to be accurate. The two spins can have values in the range $-0.99 \le \chi_{1,2} \le 0.99$.

A Reduced Order Method (ROM) basis is constructed from a training set using a greedy algorithm [38]. This is an iterative process that selects n < N waveforms (and their corresponding $\{\lambda_j\}_{j=1}^n$ values, the greedy points) that, after orthonormalization, form the reduced basis $\{e_j\}_{j=1}^n$. Each λ_i waveform in the training set is then expressed as a linear combination

$$h(t; \lambda_i) \approx \sum_{j=1}^n c_j(\lambda_i) e_j(t),$$
 (1)

within a given error tolerance, where $\{c_j(\lambda_i)\}_{j=1}^n = \langle h(t; \lambda_i), e_j(t) \rangle$ are the orthogonal projection coefficients.

Next, a new *Empirical Interpolation Method* (EIM) basis $B_k(t)$ is obtained such that a waveform $h(t; \lambda_j)$ can be expressed as a linear combination of the basis, i.e. $h(t; \lambda_j) = \sum_{k=1}^n \alpha_k(\lambda_j) B_k(t)$. The coefficients $\alpha_k(\lambda_j)$ are equal to the waveform at particular times, $\{T_k\}_{k=1}^n$, known as the empirical time nodes, i.e. $\alpha_k(\lambda_j) = h(T_k; \lambda_j)$. For any other waveform $h(t; \lambda_i)$ in the training set, the coefficients of the EIM representation are $\alpha_k(\lambda_i) = h(T_k; \lambda_i)$. This does not require the basis $B_k(t)$, so the coefficients can be computed much faster than the projection coefficients in the ROM basis (which require the projection of the whole waveform).

In the end, a surrogate model is created by interpolating the coefficient matrix $\alpha_k(\lambda_i)$ of the training set to find the coefficients $\hat{\alpha}_k(\lambda)$ for any λ , such that

$$h(t;\lambda) \approx \sum_{k=1}^{m} \hat{\alpha}_k(\lambda) B_k(t).$$
 (2)

The complexity of this process increases with the number of parameters in λ . *Neural networks can be used to speed up this part of the process, as demonstrated in* [44].

In practice, the complex waveform can be expressed in terms of its *amplitude A* and *phase* ϕ , defined through

$$h_{+}(t;\lambda) - h_{\times}(t;\lambda) = A(t;\lambda)e^{-i\phi(t;\lambda)},\tag{3}$$

which leads to a more compact EIM basis. To construct the ROM and EIM bases, a training set of $N=2\times 10^5$ waveforms was randomly sampled in the parameter space of $1\le q\le 8, -0.99\le \chi_{1,2}\le 0.99$. The waveforms were aligned in amplitude and initial phase, the phase was unwrapped, and the time series was truncated to a common starting time of -20000M, with a total mass of $M=60M_{\odot}$. This ensured that all waveforms began with a minimum frequency no larger than 15 Hz, and 100M of post-peak ringdown data was kept. The ROM and EIM bases were created using RomPy [38, 47]. To evaluate the accuracy of the reconstructed waveforms (after the training is completed), a *validation set* of 3×10^4 SEOBNRv4 waveforms (not included in the training set) was used.

For two waveforms with parameters λ_1 and λ_2 , the inner product can be defined [48]

$$\langle h(\cdot; \lambda_1), h(\cdot; \lambda_2) \rangle = 4\Re \int_{f_{min}}^{f_{max}} \frac{\tilde{h}(f; \lambda_1) \tilde{h}^*(f; \lambda_2)}{S_n(f)} df, \tag{4}$$

where $\tilde{h}(f; \lambda)$ is the Fourier transform of $h(t; \lambda)$, $S_n(f)$ denotes the noise power spectral density (PSD) of the GW detector and the star notation stands for the complex conjugate. The inner product can be employed to normalize the Fourier transform of a waveform in the following manner:

$$\hat{h}(f;\lambda) = \frac{\tilde{h}(f;\lambda)}{\langle h(\cdot;\lambda), h(\cdot;\lambda) \rangle},\tag{5}$$

Then, the *overlap* between two waveforms is defined as the inner product between normalised waveforms $\hat{h}(\cdot; \lambda_1)$, $\hat{h}(\cdot; \lambda_2)$, maximised over a relative time (t_0) and phase (ϕ_0) shift between the two waveforms:

$$O(\hat{h}(\cdot; \lambda_1), \hat{h}(\cdot; \lambda_2)) = \max_{t_0, \phi_0} \langle h(\cdot; \lambda_1), h(\cdot; \lambda_2) \rangle, \tag{6}$$

and, finally the mismatch is given by

$$\mathcal{M}(\hat{h}(\cdot;\lambda_1),\hat{h}(\cdot;\lambda_2)) = 1 - O(\hat{h}(\cdot;\lambda_1),\hat{h}(\cdot;\lambda_2)). \tag{7}$$

The performance of the surrogate model can be evaluated by comparing the waveforms generated by the SEOBNRv4 model with the predictions of the surrogate, using the mismatch defined above.

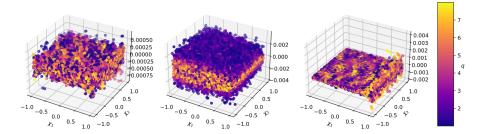


Fig. 1 The residual error for three chosen EIM coefficients for the amplitude is dependent on the input parameters $\lambda = \{\chi_1, \chi_2, q\}$ (the dependence on q is illustrated with a colormap). The example in the left panel shows an unstructured distribution of residuals. In contrast, the example in the center panel reveals a strong dependence on the mass ratio q, while the example on the right displays a large residual error at the highest value of χ_1 . Figure from [34].

2.2 Accelerating the surrogate model using ANNs

To construct the surrogate model, an ANN was employed to interpolate the coefficients $\alpha_k(\lambda_i)$ of the training set to find the coefficients $\hat{\alpha}_k(\lambda)$ for an arbitrary λ . The improved model was compared with a baseline model that followed the architecture of [44]. The ANN had four hidden layers with 320 neurons in each. The batch size was 10^3 and the training lasted for 10^3 epochs. The Adam optimizer [49] with a learning rate of 10^{-3} and the ReLU activation function [50] were used for the amplitude network. For the phase network, the Adamax [49] optimizer with a learning rate of 10^{-2} and the softplus activation function [51] were employed. Preprocessing involved using $\log(q)$ as input instead of q, which was then scaled using the StandardScaler from Scikit-Learn [52]. At the output, the coefficients were used raw for the amplitude network and were scaled using Scikit-Learn's MinMaxScaler for the phase network.

The ANN prediction of the EIM coefficients of the training set waveforms will be referred to as $\hat{y}_i \equiv \{\hat{\alpha}_k(\lambda_i)\}_{k=1}^n$. During training, the standard mean square error

$$MSE = \frac{1}{N} \sum_{i=1}^{N} \|\hat{\mathbf{y}}_i - \mathbf{y}_i\|_2^2$$
 (8)

was measured and minimized, where the $\|\cdot\|_2$ notation represents the Euclidean norm of a vector. The MSEs were in the range $\sim 10^{-8} - 10^{-7}$.

A second ANN was created to predict the residual errors after establishing the baseline ANN surrogate model. This was done due to the presence of structure in the residuals for some EIM coefficients, as seen in Fig. 1. The final predictions are the sum of the outputs of the two models. For all $\{\lambda_i\}_{i=1}^N$ in the training set, one can obtain the corresponding predictions $\{\hat{y}(\lambda_i)\}_{i=1}^N$ and calculate the *residual*

$$\mathbf{e}_i \equiv \mathbf{y}(\lambda_i) - \hat{\mathbf{y}}(\lambda_i),\tag{9}$$

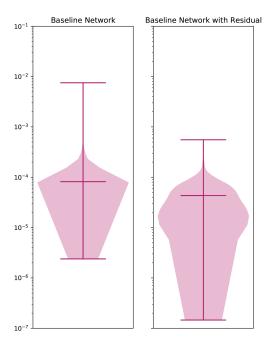


Fig. 2 A comparison of the mismatches (for the validation set) between the baseline network and the case when a second network that models the residual error is added is shown in the violin plots. The median is marked by the middle horizontal line, while the minimum and maximum values are indicated by the extent of the lines. The envelope of each panel is proportional to the density of points, and it is clear that a significant reduction of the mismatch is achieved when the network for the residual error is added. Figure from [34].

where, as already defined, y is the ground truth. The second network was created with the *same input and architecture as the first network*, but this time it was trained on the residuals e_i (which were first scaled using the "MinMaxScaler" from scikit-learn [52]) to make predictions for the residual $\hat{e}(\lambda)$ at any λ . When the prediction \hat{e} for the residual is added to the prediction \hat{y} of the first network, an improved prediction is obtained.

$$\tilde{\mathbf{y}} \equiv \hat{\mathbf{y}} + \hat{\mathbf{e}}.\tag{10}$$

Fig. 2 illustrates the difference in mismatches (for the validation set) between the baseline network and the case where a second network is added that models the residual error, as a violin plot. The median is marked by the middle horizontal line, whereas the minimum and maximum values are shown by the extent of the lines. The envelope of each panel is proportional to the density of points. The results in Fig. 2 demonstrate that adding a second network to learn the residual errors is beneficial for constructing surrogate models for gravitational waves from BBH inspiral. This strategy is likely to be advantageous for other types of GW template banks, such as binary neutron star inspiral waveforms.

3 Efficient Surrogate Models using Autoencoders

Autoencoders (AEs) are a type of unsupervised neural network that is trained to reproduce its input by first transforming it into a lower-dimensional representation [53]. Generally, an autoencoder consists of an encoding component that maps the input to a compressed representation and a decoding component that reconstructs the input. Encoding and decoding functions can have symmetrical or asymmetrical architectures and usually comprise multiple layers of fully connected layers, convolutional layers, or recurrent modules. AEs have been studied for a variety of tasks, such as clustering [54, 55], classification [56, 57], and image retrieval [58, 59], due to their ability to extract semantically meaningful representations without labels. A typical AE architecture is shown in Figure 3, with the input and output layers having the same number of neurons.

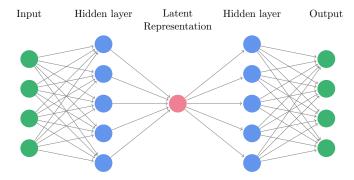


Fig. 3 Single hidden layer architecture of a fully connected Autoencoder. Figure from [35].

In [35] a dataset comprising pairs of mass ratio q_i and corresponding EIM coefficient \mathbf{a}_i (i=1,...,N), created with the EOBNRv2 nonspinning waveform model [60], was used to train an AE, with only the coefficients as input. This unsupervised process revealed a hidden relationship between each mass ratio q_i and the corresponding coefficients, as the mass ratios were unknown to the AE. Specifically, when choosing a two-dimensional intermediate representation, a spiral pattern emerged when visualizing this representation as a function of the mass ratio q, see Figure 4. Below, we summarize the main steps presented in [35] to add a learnable spiral module to the ANN.

Following [38], a dataset of N = 1000 waveforms with mass ratios in the range $1 \le q \le 2$ was generated and a surrogate model was built, with a tolerance of 10^{-10} , resulting in a reduced basis of size n = 11. Next, a simple symmetric encoder-decoder AE architecture was used, with a two-dimensional hidden representation and two hidden fully-connected layers of 128 neurons on either side. The PReLU non-linearity [61] was used in all layers. The model was built using the PyTorch

Deep Learning framework [62]. The EIM coefficients were used as input and output for this network. The AE was trained for 100 epochs with an initial learning rate of 0.001 and a batch size of 32. A multi-step multiplicative schedule was used with a gamma value of 0.9 and a step size of 15. The visual representation of the hidden layer is shown in Figure 4, with the colors indicating the q values for each input coefficient. The spiral manifold in the hidden layer appears to describe a linear relationship between q and the angle θ of the spiral. The mean squared error of the reconstruction is 6.82×10^{-5} .

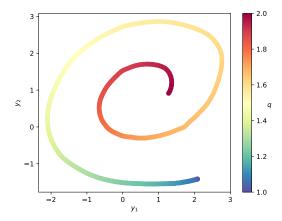


Fig. 4 Hidden representation uncovered by the AE for the empirical interpolation coefficients of a surrogate model of EOBNRv2 waveforms that is valid for $1 \le q \le 2$ (the color bar describes different values of the mass ratio q). When the values at the two neurons of the hidden representation are plotted against each other, a spiral structure emerges, along which the mass ratio appears to vary linearly with angle. Figure from [35].

Based on the spiral pattern that emerged in Figure 4, a neural spiral module was proposed in [35], which first transforms the input q into an angle θ , defined as

$$\theta := w \cdot q + b,\tag{11}$$

and subsequently maps θ onto a spiral structure of the form

$$s_x := (\alpha + \beta \cdot \theta) \cdot \cos \theta,$$

$$s_y := (\alpha + \beta \cdot \theta) \cdot \sin \theta,$$
(12)

where w, b, α and β are parameters. These parameters are learnable, since the output is differentiable with respect to each of them. The spiral is fed to multiple, successive fully-connected layers, each with a nonlinear activation function, before reaching the final linear layer. An example of this architecture with two hidden layers is illustrated

in Figure 5. The inclusion of this module into an ANN accelerates the training process, leading to a significant reduction of the lowest achieved MSE.

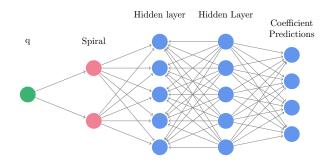


Fig. 5 Fully connected neural network with two hidden layers and the included spiral module. Figure from [35].

The performance of various neural network architectures with fully-connected layers was assessed with and without the spiral module. The metrics used for evaluation were the waveform mismatch, inference speed, and memory requirements, with the maximum batch size that can be processed in a single forward pass on an NVIDIA RTX 2080 Ti GPU. All networks were trained for 2500 epochs with a batch size of 16, using the Adam optimizer [63] and an initial learning rate of 0.001, which was reduced by 0.95 every 150 epochs.

The inclusion of the spiral module significantly improved the mismatch achieved. When only one hidden layer was used, the baseline network with 128 neurons produced waveforms with a very poor mismatch (1.03×10^{-1} median mismatch). However, with the addition of the spiral module, even with only 32 hidden neurons, the median mismatch decreased by about 6 orders of magnitude. The best median and 95th percentile mismatch (9.41×10^{-9} and 3.48×10^{-8}) was achieved by the S-32-64-128-64 network, which was able to generate up to 3.4 million coefficients in a single forward pass on the aforementioned GPU.

Finally, a spiral module was added to a neural network that was trained on a larger dataset of N=56000 waveforms with $1 \le q \le 8$, where the q values were equidistant. A validation and a test set were also created, each with 14000 waveforms, and the q values were randomly chosen in the range of $1 \le q \le 8$. Figure 6 shows the real and imaginary parts of the first ten coefficients of the EIM basis, $\{a_j(q)\}_{j=1}^{10}$. Despite some modulation of amplitude, each coefficient has a sinusoidal dependence with q (except near q=1, where $dq/da_j=0$ for all j).

Several neural networks were trained and tested on the dataset. All networks were trained for 5000 epochs, with a batch size of 32, and the Adam optimizer [63] with an initial learning rate of 0.001, which was reduced by 0.9 every 30 epochs. The training and validation loss per epoch for the 32-64-128-64 network and the corresponding architecture with the addition of the spiral is shown in Figure 7. The

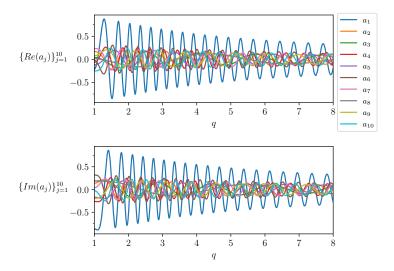


Fig. 6 Real (top) and imaginary parts (bottom) of the empirical interpolation coefficients $a_j(q)$ for a surrogate model of EOBNRv2 waveforms that is valid for $1 \le q \le 8$. Figure from [35].

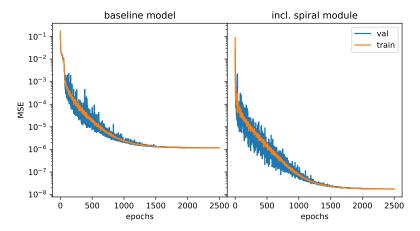


Fig. 7 Training and validation loss per epoch for the 32-64-128-64 network and the corresponding architecture with the addition of the spiral. Figure from [35].

spiral addition resulted in a lower mean squared error, allowing smaller networks to achieve the same accuracy as larger networks, leading to a larger batch size that can be processed in a single forward pass when using a specific GPU card.

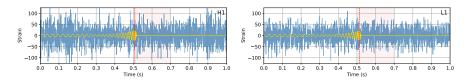


Fig. 8 A 2-channel segment of data from the training set used in [36] is displayed in the left and right panels, respectively, for the Hanford (H1) and Livingston (L1) detectors. The whitened strain of a 1 s segment is shown around the time of coalescence. The coalescence times in the detector frames are within the 0.5-0.7 second range (indicated by the shaded area). The injected waveform is scaled to match the difference between the whitened foreground and background segments. Figure from [36].

4 GW Detection with Deep Residual Networks

The fourth observing run (O4) of gravitational wave detectors, which began in spring of 2023, is expected to result in a significant increase in the number of detections. This increase will be even more pronounced during O5 and the observing runs of the planned third-generation detectors (e.g. Cosmic Explorer [65] and Einstein Telescope [66]). However, the application of traditional matched-filtering techniques to obtain near real-time detection triggers is becoming increasingly costly or even impractical [67], due to both computational efficiency and accuracy. This is especially true for near-threshold systems with random spin directions, which require a much larger parameter space than the aligned-spin case. The situation will become even more difficult if template banks with departures from general relativity (GR) are included. Unmodeled search algorithms, on the other hand, have limited sensitivity, depending on the particular GW source.

Recently, the implementation of machine-learning (ML) methods, such as convolutional neural networks (CNN) or auto-encoders, has been investigated as an attractive solution to the problem of detecting gravitational waves (GWs), see e.g. [68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 64,90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101] and [31, 32, 33] for reviews. However, it has been difficult to evaluate the effectiveness of such efforts in a realistic setting. The first Machine-Learning Gravitational-Wave Mock Data Challenge (MLGWSC-1) was completed [102], providing an objective framework for testing the sensitivity and efficiency of ML algorithms on modeled injections in both Gaussian and real O3a detector noise in comparison to traditional algorithms. In [36], the leading ML algorithm in the case of real O3a noise was presented in more detail and it was shown that with further improvements it surpasses, for the first time, the results obtained with standard configurations of traditional algorithms in this specific setting. This was achieved for a component mass range between $7 - 50M_{\odot}$ (which corresponds to 70% of the announced events in the cumulative GWTC catalog [103]) and a relatively low false-alarm rate (FAR) as small as one per month.

The AresGW algorithm, described in [36], combines several components that increase the sensitive distance. It is based on a 54-layer one-dimensional deep residual

network (ResNet) [104], which is more capable than a simpler CNN. Additionally, the Deep Adaptive Input Normalization (DAIN) [105] was included to address the non-stationary nature of real O3a noise. Furthermore, the dataset was augmented during training to improve the results. The execution speed was increased with the implementation of a framework-specific, module-based whitening layer, which computes the power spectral density (PSD) in batched tensor format. Finally, curriculum learning was used, which allowed the network to learn waveforms with the highest signal-to-noise ratio (SNR) first. The network was created using PyTorch [106] and trained (including validation) on 12 days of data in 31 hours on an A6000 GPU (for 14 epochs). The evaluation of one month of test data on the same hardware took less than 2 hours. The main findings of [36] are summarized below.

4.1 Training and test datasets

The training dataset in [36] spanned a period of 12 days and included real noise from the O3a LIGO run and injections of non-aligned binary black hole waveforms (in accordance with dataset 4 in [102]). Noise was taken from sections of O3a that are available from the Gravitational Wave Open Science Center (GWOSC) [107]. Only segments with a minimum length of 2 hours, where both LIGO detectors had good quality data, and excluding 10 seconds around detections listed in GWTC-2 were included (see [102] for more information). Applying these criteria, the dataset had noise from each of the two aLIGO detectors, Hanford (H1) and Livingston (L1), with a total duration of 11 weeks and a sampling rate of 2048 Hz.

The waveforms injected into the training set were generated using the IMRPhenomXPHM waveform model [108], with a lower-frequency cutoff of 20 Hz. The masses of the individual components, m_1 and m_2 , ranged from 7 to 50 solar masses, resulting in a maximum signal duration of 20 seconds. The signals were uniformly distributed in coalescence phase, polarization, inclination, declination, and right ascension (see [102] for more information). The waveforms were not uniformly distributed in volume, but instead, the chirp distance d_c [102] was sampled (as opposed to the luminosity distance d). This selection increases the number of low-mass systems that can be detected. The spins of the individual components had an isotropically distributed orientation with a magnitude between 0 and 0.99, which means that precession effects were present. All higher-order modes up to (4, -4) available in IMRPhenomXPHM were included. A representative data segment of the training set is shown in Figure 8.

The training data set was made up of the first 12 days of the 11-week dataset, resulting in 740k noise segments that formed the background noise. A set of 38k different waveforms was randomly injected into around 19 different background segments each, resulting in 740k foreground segments that contain injections, creating a balanced training set. Additionally, a validation set was created, based on weeks 4 to 7 of the 11-week dataset and with a different random seed for injections. Lastly, the test data set consisted of noise from weeks 8-11 of the 11-week dataset and injections

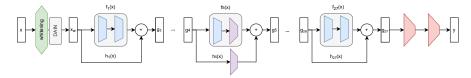


Fig. 9 Description of the residual network architecture in [36]. The input \mathbf{x} is 2×2048 -dimensional (see the text for details). Figure from [36].

with merger times randomly spaced between 24s to 30s. For the test dataset the same random seed and offset as in [102] was used.

The training data was first pre-processed using whitening as in [109, 68] and then normalized using the DAIN algorithm [105, 110]. DAIN is trained by back-propagating the network's gradients to its parameter. Furthermore, DAIN can adjust the normalization scheme applied to the input *during inference*, thus allowing it to handle non-stationary data.

4.2 Deep Residual Networks

Residual neural networks [104] employ skip connections to improve training, allowing gradients to better reach the earliest layers of the neural network architecture, thus solving the problem of vanishing gradients [111]. This, in combination with well-crafted training methods [112], leads to more effective training as the number of layers increases. This allows for the training of much deeper networks than simple CNNs.

The deep residual network developed in [36] was based on 1D convolutions for the binary classification of 1s long (i.e. 2×2048 -dimensional) segments into either positive (containing an injection) or negative (pure noise) segments. The network had a depth of 54 layers, which were grouped into 27 blocks containing two convolutional layers with a varying number of filters. Blocks 5, 8, 11, 14, and 17 were 2-strided, which means that the dimensionality was halved and an additional layer was used in the residual connection. Each convolutional layer was followed by batch normalization and a ReLU activation function. The two final individual convolutional layers reduced the output to a binary outcome (noise plus injected waveform vs. noise only). The Adam optimizer [113] was used for backpropagation and regularized binary cross entropy (a variant of the finite cross-entropy loss function [85]) was used as the objective function. During training, dynamic augmentation was also employed. A graphical representation of the network architecture is shown in Figure 9.

A learning strategy was developed such that the network was initially trained on the strongest injections and only later on weaker injections. This is accomplished by using the optimal signal-to-noise ratio (SNR) of the injected signal, which is given by the equation

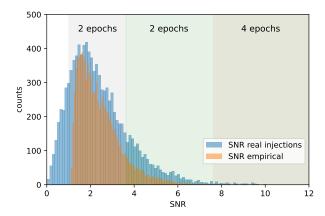


Fig. 10 Comparison of the SNR histogram, calculated with the empirical relation in Eq. (14), to the optimal SNR of 10⁴ randomly chosen injections. The shaded areas in the graph represent the limited SNR values used in the first eight epochs of the *learning strategy*. Figure from [36]

$$SNR = 2\sqrt{\int_0^\infty df \frac{\tilde{h}(f)^2}{S_n(f)}},$$
(13)

where $\tilde{h}(f)$ is the amplitude of the Fourier transform of the injected signal and $S_n(f)$ is the power spectral density of the detector noise. Instead of using the actual optimal SNR, an empirical relation was created that only depends on the chirp mass, the distance, and the inclination angle ι

SNR =
$$\frac{1261 \text{ Mpc}}{D} \left(\frac{\mathcal{M}_c}{M_\odot}\right)^{5/7} [0.7 + 0.3\cos(2\iota)].$$
 (14)

Figure 10 displays a comparison between the optimal SNR (calculated using Eq. (13) and the PSD of the Hanford detector) for a sample of 10⁴ randomly chosen injections and the approximate SNR as determined by Eq. (14). The two distributions are similar; however, the actual SNR is affected by a variety of other factors (such as sky location and spins).

The training process began with signals that were easily recognizable and had a high estimated SNR for the first four epochs. Subsequently, the network was gradually trained on weaker signals, and after the tenth epoch, it learned all the signals in the training set. Figure 10 shows the first eight epochs compared to the SNR distribution. As a result of the chosen learning strategy, initial losses were low. When the network had learned all the signals (after ten epochs), the loss of the training set was equal to that of the validation set.

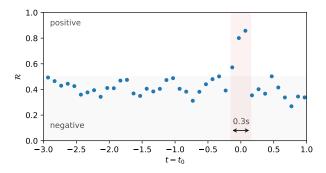


Fig. 11 Representative example of the ranking statistic for overlapping segments (with a stride of 0.1s). Segments with $\mathcal{R} < 0.5$ are classified as negatives. Segments with $\mathcal{R} > 0.5$ that cluster within 0.3s of a known injection at time t_0 are reported as a single true positive. Figure from [36]

4.3 Detection of BBH injections in real noise

The trained network was used to analyze segments of the test dataset, producing a binary output that corresponds to the probability that an injection is present or that the segment only contains noise. The first output was used as a ranking statistic \mathcal{R} (with values between 0 and 1). When $\mathcal{R}>0.5$, a positive result was recorded. Clustering was applied to any positives that were detected within a time interval of 0.3 s, which were reported as a single detection (see Fig. 11 for an example). After deployment, the output was evaluated every 0.1 seconds and compared with the known injection times in the test dataset. If a positive output was within 0.3 seconds of the nominal merger time for a particular injection, it was classified as a true positive and, otherwise, as a false positive.

In order to assess the effectiveness of the search algorithm, one first calculates the false alarm rate as a function of the ranking statistic, that is, the $FAR(\mathcal{R})$ function. Subsequently, the sensitivity of the search is determined as a function of the ranking statistic, which produces a relationship between the sensitivity and FAR (see [102] for definitions and the complete methodology).

Fig. 12 shows the sensitivity distance as a function of FAR for the best model (ResNet54d+SNR), compared to a simpler setup (ResNet54) and two leading algorithms for GW detection, Coherent WaveBurst (cWB) and PyCBC. cWB is a waveform model-agnostic search pipeline for GW signals based on the constrained likelihood method [114, 115, 116], while PyCBC [117] is based on a standard configuration of archival search for compact-binary mergers [118]. The cWB and PyCBC results presented in Figure 12 are taken from [102], where they were obtained on the same test dataset. cWB uses wavelets, which prevents it from achieving ideal fitting factors. It was recently improved with machine-learning techniques [119]. PyCBC implements matched filtering of waveform templates, but in [118, 102] only aligned-spin templates were used (for more general waveforms, the method could become computationally too expensive). Since the injections in the test dataset of

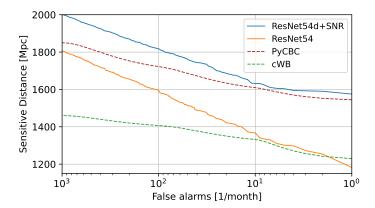


Fig. 12 Comparison of the performance of the best model (ResNet54d+SNR) in [36] with the simpler setup (ResNet54) used in [102] and two leading algorithms for GW detection, Coherent WaveBurst (cWB) and PyCBC (which only used aligned-spin templates, see text). All codes were tested on the same dataset established in [102]. The ResNet54d+SNR model outperformed the other algorithms at all false alarm rates in this setting. Figure from [36].

[36] were based on more general waveforms, this particular PyCBC search could not reach ideal fitting factors, leaving room for other algorithms to outperform it. As seen in Figure 12, the best model presented in [36], which included SNR-based curriculum learning, outperformed the PyCBC results at all FAR and also significantly surpassed the sensitivity of the unmodeled cWB search.

5 ANN Surrogate Models of Neutron Star Mass-Radius Relations in Alternative Theories of Gravity

Several studies have used Artificial Neural Networks (ANNs) to reconstruct the EoS of neutron stars (NSs) based on their observable properties [120, 121, 122, 123, 124]. For instance, [125] investigated the use of ANNs with the autoencoder architecture, while [126, 127] employed ANNs to represent the EoS in a model-independent way, utilizing the unsupervised automatic differentiation framework. Other machine-learning techniques have been applied to explore the NS EoS. For example, [128] used a clustering method to identify patterns in mass-radius curves, and [129] investigated correlations among different EoSs of dense matter using unsupervised Machine Learning (ML) techniques. Additionally, attempts have been made to derive nuclear matter characteristics from NS EoS and observations using deep neural networks; see, e.g., [130, 131].

Neutron stars have been the focus of theoretical investigations in alternative theories of gravity (see [132, 133] for reviews and possible tests, and [134] for the particular theory discussed in this work). Bayesian statistics is commonly used to

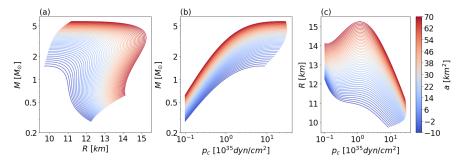


Fig. 13 Data set for EoS BSk20, for modeling the function f_1 . Figure from [37].

infer the NS EoS from observations of their macroscopic properties. This requires a TOV (Tolman-Oppenheimer-Volkoff) solver to be run multiple times to obtain the final posterior distribution of various parameters. If modified theories of gravity are to be included in these studies, a modified TOV solver, such as the one presented in [134], is needed. However, this algorithm is based on an iterative method for solving a differential equation system, which makes Bayesian inference computationally expensive. Therefore, it would be beneficial to find an alternative way to quickly and accurately predict the macroscopic properties of NSs, given some defining characteristics or other macroscopic properties of each equilibrium model.

Motivated by the need for a faster process, we developed an Artificial Neural Network (ANN) regression for two types of functions, $f_1(\text{EoS}; \alpha, p_c) \rightarrow (M, R)$ and $f_2(\text{EoS}; \alpha, M) \rightarrow R$, which was implemented in [37] and used in a Bayesian inference application in [135]. The EoS is a distinct variable, with each type having one ANN model for each EoS. α is the coupling constant of the theory and p_c is the central pressure of the NS. The first type serves as a surrogate model for the numerical iterative method described in [134], which provides the mass and radius of NSs for a specific EoS and a given pair of α and p_c . The aim is to speed up the process while still meeting strict accuracy requirements. The second type cannot be obtained directly using the iterative method since p_c must be an input. Therefore, implementing a root-finding algorithm would be the only solution, resulting in further time delays. On the other hand, training ANNs to predict R based on (EoS; α , M) offers a more straightforward approach to handling type f_2 .

The 4D Horndeski scalar-tensor model, which was studied in [37], was derived from higher-dimensional Einstein-Gauss-Bonnet gravity. The action of this model included several nonlinear terms of a scalar field ϕ , and the mass and radius of the neutron star were affected by the strength of the coupling constant α , which has units of length square. If $\alpha = 0$, then the model is equivalent to Einstein's General Relativity. Further information can be found in [134].

Table 1 Final network architecture in [37].

Layer	Type f_1	Type f_2
Input layer	(α, p_c)	(α, M)
Hidden layer 1	25-tanh	25-tanh
Hidden layer 2	35-relu	35-relu
Hidden layer 3	25-tanh	25-tanh
Output layer	(M,R)	R

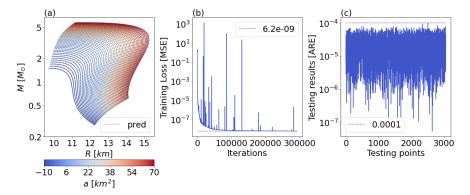


Fig. 14 Training results for f_1 and EoS BSk20. Figure from [37].

6 Training and Testing

The numerical code from [134] requires two inputs to generate the data sets in [37]: the coupling constant α [km²] of the theory and the central pressure p_c [10³⁵ dyn/cm²]. These inputs are used to calculate the mass M [M $_{\odot}$] and radius R [km] of the neutron star. For each type of function, 20 data sets were created, each based on one of the 20 tabulated equations of state.

The type of function $f_1(\text{EoS}; \alpha, p_c)$ maps a given set of 51 values of α and 200 values of p_c to a unique pair of (M, R) values. The α values were evenly distributed on a linear scale, while the p_c values were logarithmically spaced between 0.1 and 1.2 times the central pressure required for a NS to reach its maximum mass for a specific value of the coupling constant. As an example, Figure 13 shows the mapping of (α, p_c) pairs to (M, R) pairs for EoS BSk20. The type of function $f_2(\text{EoS}; \alpha, M) \to R$ had the same size as f_1 , but the range of p_c values was chosen differently. The data comprised 51 values of α ranging from -10 to 70, and 200 values of p_c ranging from 0.1 to p_{max} .

The TensorFlow module Keras¹ was employed with a 70:30 train-test ratio. The Mean Square Error (MSE) was selected as the loss function, while the Absolute Relative Error (ARE) was used as the criterion for evaluating the trained models. The ANN architecture for each type is presented in Table 1. The second-order

¹ https://www.tensorflow.org/api_docs/python/tf/keras

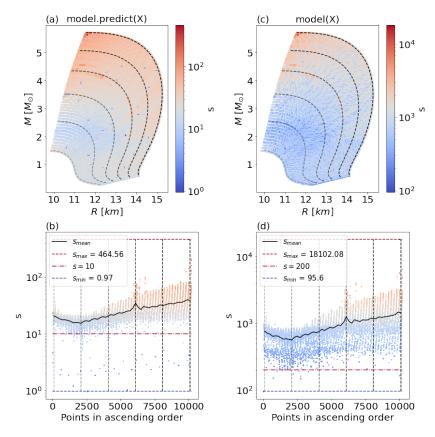


Fig. 15 Speed-up per M-R data point showing how many times faster the trained ANN model is than the numerical code in each case. Figure from [37].

Broyden–Fletcher–Goldfarb–Shanno optimizer algorithm (BFGS), as proposed in [136], was implemented as the preferred optimizer².

Fig. 14a provides an indication of the training results for f_1 , comparing the actual output with the predicted output. Fig. 14b displays the training loss per iteration in terms of the Mean Square Error (MSE) and Fig. 14c presents the Absolute Relative Error (ARE) at a given test point i

$$ARE_{i} = \frac{1}{m} \sum_{j=1}^{m} \left| \frac{Y_{i}^{j} - Y_{i, \text{ true}}^{j}}{Y_{i, \text{ true}}^{j}} \right|$$
 (15)

where m is the number of output neurons, Y_i^j is the network output with index j for test point with index i, and $Y_{i,\text{true}}^j$ is the corresponding real output. The Mean ARE

² BFGS is not included in the list of provided Keras optimizers.

over all n = 3060 points in the test dataset is denoted as MARE. The MARE ranged between 10^{-5} and 10^{-4} for all EOSs, while the maximum ARE was 6×10^{-4} for f_1 and 9×10^{-3} for f_2 , indicating that the absolute relative error never exceeded 1% in the entire domain of the training and test sets.

6.1 Speed-up

The *speed-up s* when comparing the speed of the trained ANN models of f_1 to the speed of the numerical code, is defined as:

$$s = \frac{\Delta t_{\text{ANN}}}{\Delta t_{\text{num}}},\tag{16}$$

where Δt_{ANN} is the run time when using an ANN model and Δt_{num} the run time of the numerical code with the iterative numerical scheme. The output of the models can be calculated in three different ways, leading to a different speed-up:

- 1. model.predict(X), with X being one input value,
- 2. model(X), with X being one input value,
- 3. model.predict(X), with X being an array of input values.

The left panel of Figure 15 shows the speed up for model.predict(X). Panel 15a displays the speed up using different colors, and the dashed lines with increasing transparency represent six different M-R curves for $a=\{-10,6,22,38,54,70\}$ km² respectively. Panel 15b shows the speed-up values for a particular arrangement of the data points, which are sorted in ascending order of α and p_c . Most of the speed-up values were between 10 and 100, with less than 1% outside of this range. Generally, larger input values tend to result in higher speed-ups, with an average speed-up of approximately 25 across all data points.

The right panel of Figure 15 shows the acceleration for the model (X) case, which is set up similarly to the left panel. Most of the speed-up values are between 200 and 18000 (less than 0.2% are outside of this range). The average speed-up, taking into account all data points, was greater than 900. For the model.predict(X) case, the 10200 data points were entered as an array of input values X. In this case, the effective run time Δt_{eff} , can be calculated, which is defined as

$$\Delta t_{eff} = \frac{\Delta t_N}{N},\tag{17}$$

where Δt_N is the *total run time* for the whole array as input and N=10200. The speed-up ranged from ~ 4600 to ~ 565000 , with a mean value of ~ 31000 .

7 Conclusions

This comprehensive review has highlighted several advancements in the field of gravitational wave astronomy through the application of machine learning techniques. The exploration of various neural network architectures, including deep residual networks and autoencoders, has demonstrated the potential for improvements in the detection and analysis of gravitational wave signals. The successful integration of machine learning in this domain not only concerns the optimization of data analysis, but also can open new avenues for understanding complex astrophysical phenomena. Moreover, the application of machine learning in neutron star mass-radius relation studies in alternative theories of gravity could be used to break the degeneracy between the equation of state and some alternative theories of gravity.

In conclusion, the integration of machine learning techniques in gravitational wave astronomy and neutron star studies represents a paradigm shift. Not only does it enhance the accuracy and efficiency of existing methodologies, but it also paves the way for novel discoveries in the realm of compact object astrophysics. The continued collaboration between the gravitational wave and machine learning communities is vital for further advancements.

Acknowledgements I am grateful to my collaborators, Theocharis Apostolatos, Stella Fragkouli, Panagiotis Iosif, Alexandra Koloniari, Ioannis Liodis, Paraskevi Nousi, George Pappas, Nikolaos Passalis, Evangelos Smyrniotis, and Anastasios Tefas, for their contributions that led to the main publications summarized in this review. Many thanks to Elena Cuoco for comments on the manuscript. This work was carried out within the framework of the EU COST action No. CA17137. This research has made use of data or software obtained from the Gravitational Wave Open Science Center (gwosc.org), a service of the LIGO Scientific Collaboration, the Virgo Collaboration, and KAGRA. This material is based upon work supported by NSF's LIGO Laboratory which is a major facility fully funded by the National Science Foundation, as well as the Science and Technology Facilities Council (STFC) of the United Kingdom, the Max-Planck-Society (MPS), and the State of Niedersachsen/Germany for support of the construction of Advanced LIGO and construction and operation of the GEO600 detector. Additional support for Advanced LIGO was provided by the Australian Research Council. Virgo is funded, through the European Gravitational Observatory (EGO), by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale di Fisica Nucleare (INFN) and the Dutch Nikhef, with contributions by institutions from Belgium, Germany, Greece, Hungary, Ireland, Japan, Monaco, Poland, Portugal, Spain. KAGRA is supported by Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan Society for the Promotion of Science (JSPS) in Japan; National Research Foundation (NRF) and Ministry of Science and ICT (MSIT) in Korea; Academia Sinica (AS) and National Science and Technology Council (NSTC) in Taiwan.

References

- B.P. Abbott et al. Observation of Gravitational Waves from a Binary Black Hole Merger. Phys. Rev. Lett., 116(6):061102, 2016.
- R. Abbott et al. GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run. arXiv e-prints, page arXiv:2111.03606,

- November 2021.
- 3. J. Aasi et al. Advanced LIGO. Class. Quant. Grav., 32:074001, 2015.
- F Acernese et al. Advanced Virgo: a second-generation interferometric gravitational wave detector. Classical and Quantum Gravity, 32(2):024001, 2014.
- T Akutsu et al. KAGRA: 2.5 Generation Interferometric Gravitational Wave Detector. Nature Astronomy, 3(1):35–40, Jan 2019.
- T Akutsu et al. Overview of KAGRA: Detector design and construction history. Progress of Theoretical and Experimental Physics, 2021(5), 08 2020. 05A101.
- B. P. Abbott et al. Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA. *Living Reviews in Relativity*, 23(1):3, Sep 2020
- 8. M. Saleem et al. The science case for LIGO-India. Classical and Quantum Gravity, 39(2):025004, January 2022.
- M. Punturo et al. The Einstein Telescope: a third-generation gravitational wave observatory. Class. Quantum Gravity, 27(19):194002, 2010.
- Michele Maggiore et al. Science case for the Einstein telescope. *Journal of Cosmology and Astroparticle Physics*, 2020(3):050, March 2020.
- David Reitze et al. Cosmic Explorer: The U.S. contribution to gravitational-wave astronomy beyond LIGO. In *Bull. Am. Astron. Soc.*, volume 51, page 35, 2019.
- Matthew Evans et al. A Horizon Study for Cosmic Explorer: Science, Observatories, and Community. arXiv e-prints, page arXiv:2109.09882, September 2021.
- 13. B. P. Abbott et al. Exploring the sensitivity of next generation gravitational wave detectors. *Class. Quantum Gravity*, 34(4):044001, 2017.
- 14. David Reitze, Michele Punturo, Peter Couvares, Stavros Katsanevas, Takaaki Kajita, Vicky Kalogera, Harald Lueck, David McClelland, Sheila Rowan, Gary Sanders, B. S. Sathyaprakash, David Shoemaker, and Jo van den Brand. Expanding the Reach of Gravitational Wave Astronomy to the Edge of the Universe: The Gravitational-Wave International Committee Study Reports on Next Generation Ground-based Gravitational-Wave Observatories. arXiv e-prints, page arXiv:2111.06986, November 2021.
- Vicky Kalogera, B. S. Sathyaprakash, Matthew Bailes, Marie-Anne Bizouard, et al. The Next Generation Global Gravitational Wave Observatory: The Science Book. arXiv e-prints, page arXiv:2111.06990, November 2021.
- Geraint Pratten, Cecilio García-Quirós, Marta Colleoni, Antoni Ramos-Buades, Héctor Estellés, Maite Mateu-Lucena, Rafel Jaume, Maria Haney, David Keitel, Jonathan E. Thompson, and Sascha Husa. Computationally efficient models for the dominant and subdominant harmonic modes of precessing binary black holes. *Phys. Rev. D*, 103:104056, May 2021.
- 17. Antoni Ramos-Buades, Alessandra Buonanno, Héctor Estellés, Mohammed Khalil, Deyan P. Mihaylov, Serguei Ossokine, Lorenzo Pompili, and Mahlet Shiferaw. SEOBNRv5PHM: Next generation of accurate and efficient multipolar precessing-spin effective-one-body waveforms for binary black holes. arXiv e-prints, page arXiv:2303.18046, March 2023.
- T. E. Riley, A. L. Watts, S. Bogdanov, P. S. Ray, R. M. Ludlam, S. Guillot, Z. Arzoumanian,
 C. L. Baker, A. V. Bilous, D. Chakrabarty, K. C. Gendreau, A. K. Harding, W. C. G. Ho,
 J. M. Lattimer, S. M. Morsink, and T. E. Strohmayer. A nicer view of PSR j0030+0451:
 Millisecond pulsar parameter estimation. *The Astrophysical Journal*, 887(1):L21, dec 2019.
- M. C. Miller, F. K. Lamb, A. J. Dittmann, S. Bogdanov, Z. Arzoumanian, K. C. Gendreau, S. Guillot, A. K. Harding, W. C. G. Ho, J. M. Lattimer, R. M. Ludlam, S. Mahmoodifar, S. M. Morsink, P. S. Ray, T. E. Strohmayer, K. S. Wood, T. Enoto, R. Foster, T. Okajima, G. Prigozhin, and Y. Soong. Psr j0030+0451 mass and radius from nicer data and implications for the properties of neutron star matter. *The Astrophysical Journal Letters*, 887(1):L24, dec 2019.
- M. C. Miller, F. K. Lamb, A. J. Dittmann, S. Bogdanov, Z. Arzoumanian, K. C. Gendreau, S. Guillot, W. C. G. Ho, J. M. Lattimer, M. Loewenstein, S. M. Morsink, P. S. Ray, M. T. Wolff, C. L. Baker, T. Cazeau, S. Manthripragada, C. B. Markwardt, T. Okajima, S. Pollard, I. Cognard, H. T. Cromartie, E. Fonseca, L. Guillemot, M. Kerr, A. Parthasarathy, T. T.

- Pennucci, S. Ransom, and I. Stairs. The radius of psr j0740+6620 from nicer and xmmnewton data. *The Astrophysical Journal Letters*, 918(2):L28, sep 2021.
- Eric D. Van Oeveren and John L. Friedman. Upper limit set by causality on the tidal deformability of a neutron star. *Physical Review D*, 95(8), apr 2017.
- 22. Tanja Hinderer. Tidal Love numbers of neutron stars. Astrophys. J., 677:1216–1220, 2008.
- Katerina Chatziioannou. Neutron-star tidal deformability and equation-of-state constraints. General Relativity and Gravitation, 52(11):109, 2020.
- 24. Tim Dietrich, Tanja Hinderer, and Anuradha Samajdar. Interpreting binary neutron star mergers: describing the binary neutron star dynamics, modelling gravitational waveforms, and analyzing detections. *General Relativity and Gravitation*, 53(3):27, 2021.
- Bhaskar Biswas. Bayesian model selection of neutron star equations of state using multimessenger observations. The Astrophysical Journal, 926(1):75, feb 2022.
- Tim Dietrich, Michael W. Coughlin, Peter T. H. Pang, Mattia Bulla, Jack Heinzel, Lina Issa, Ingo Tews, and Sarah Antier. Multimessenger constraints on the neutron-star equation of state and the hubble constant. *Science*, 370(6523):1450–1453, 2020.
- Philippe Landry, Reed Essick, and Katerina Chatziioannou. Nonparametric constraints on neutron star matter with existing and upcoming gravitational wave and pulsar observations. *Phys. Rev. D*, 101(12):123007, June 2020.
- G. Raaijmakers, S. K. Greif, K. Hebeler, T. Hinderer, S. Nissanke, A. Schwenk, T. E. Riley, A. L. Watts, J. M. Lattimer, and W. C. G. Ho. Constraints on the Dense Matter Equation of State and Neutron Star Properties from NICER's Mass–Radius Estimate of PSR J0740+6620 and Multimessenger Observations. *Astrophys. J. Lett.*, 918(2):L29, 2021.
- B. P. Abbott et al. GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral. *Phys. Rev. Lett.*, 119(16):161101, 2017.
- B. P. Abbott et al. Properties of the binary neutron star merger GW170817. Phys. Rev. X, 9(1):011001, 2019.
- 31. Elena Cuoco, Jade Powell, Marco Cavaglià, Kendall Ackley, Micha l Bejger, Chayan Chatterjee, Michael Coughlin, Scott Coughlin, Paul Easter, Reed Essick, Hunter Gabbard, Timothy Gebhard, Shaon Ghosh, Leïla Haegel, Alberto Iess, David Keitel, Zsuzsa Márka, Szabolcs Márka, Filip Morawski, Tri Nguyen, Rich Ormiston, Michael Pürrer, Massimiliano Razzano, Kai Staats, Gabriele Vajente, and Daniel Williams. Enhancing gravitational-wave science with machine learning. Machine Learning: Science and Technology, 2(1):011002, dec 2020.
- Vincenzo Benedetto, Francesco Gissi, Gioele Ciaparrone, and Luigi Troiano. Ai in gravitational wave analysis, an overview. Applied Sciences, 13(17), 2023.
- Tianyu Zhao, Ruijun Shi, Yue Zhou, Zhoujian Cao, and Zhixiang Ren. Dawning of a New Era in Gravitational Wave Data Analysis: Unveiling Cosmic Mysteries via Artificial Intelligence – A Systematic Review. arXiv e-prints, page arXiv:2311.15585, November 2023.
- Styliani-Christina Fragkouli, Paraskevi Nousi, Nikolaos Passalis, Panagiotis Iosif, Nikolaos Stergioulas, and Anastasios Tefas. Deep residual error and bag-of-tricks learning for gravitational wave surrogate modeling. *Applied Soft Computing*, 147:110746, 2023.
- Paraskevi Nousi, Styliani-Christina Fragkouli, Nikolaos Passalis, Panagiotis Iosif, Theocharis Apostolatos, George Pappas, Nikolaos Stergioulas, and Anastasios Tefas. Autoencoder-driven Spiral Representation Learning for Gravitational Wave Surrogate Modelling. *Neurocomputing*, 491:67–77, June 2022.
- Paraskevi Nousi, Alexandra E. Koloniari, Nikolaos Passalis, Panagiotis Iosif, Nikolaos Stergioulas, and Anastasios Tefas. Deep residual networks for gravitational wave detection. *Phys. Rev. D*, 108(2):024022, July 2023.
- 37. Ioannis Liodis, Evaggelos Smirniotis, and Nikolaos Stergioulas. A neural-network-based surrogate model for the properties of neutron stars in 4D Einstein-Gauss-Bonnet gravity. *arXiv e-prints*, page arXiv:2309.03991, September 2023.
- 38. Scott E Field, Chad R Galley, Jan S Hesthaven, Jason Kaye, and Manuel Tiglio. Fast prediction and evaluation of gravitational waveforms using surrogate models. *Physical Review X*, 4(3):031006, 2014.
- Manuel Tiglio and Aarón Villanueva. Reduced Order and Surrogate Models for Gravitational Waves. arXiv e-prints, page arXiv:2101.11608, January 2021.

 Michael Pürrer. Frequency domain reduced order model of aligned-spin effective-one-body waveforms with generic mass ratios and spins. *Phys. Rev. D*, 93(6):064041, March 2016.

- Benjamin D. Lackey, Sebastiano Bernuzzi, Chad R. Galley, Jeroen Meidam, and Chris Van Den Broeck. Effective-one-body waveforms for binary neutron stars using surrogate models. *Phys. Rev. D*, 95(10):104036, May 2017.
- 42. Benjamin D. Lackey, Michael Pürrer, Andrea Taracchini, and Sylvain Marsat. Surrogate model for an aligned-spin effective-one-body waveform model of binary neutron star inspirals using Gaussian process regression. *Phys. Rev. D*, 100(2):024002, July 2019.
- Qianyun Yun, Wen-Biao Han, Xingyu Zhong, and Carlos A. Benavides-Gallego. Surrogate model for gravitational waveforms of spin-aligned binary black holes with eccentricities. *Phys. Rev. D*, 103(12):124053, June 2021.
- Sebastian Khan and Rhys Green. Gravitational-wave surrogate models powered by artificial neural networks. *Phys. Rev. D*, 103(6):064015, 2021.
- 45. Michele Maggiore. Gravitational Waves Volume 1: Theory and Experiments. Oxford University Press, 2008.
- 46. Alejandro Bohé, Lijing Shao, Andrea Taracchini, Alessandra Buonanno, Stanislav Babak, Ian W. Harry, Ian Hinder, Serguei Ossokine, Michael Pürrer, Vivien Raymond, Tony Chu, Heather Fong, Prayush Kumar, Harald P. Pfeiffer, Michael Boyle, Daniel A. Hemberger, Lawrence E. Kidder, Geoffrey Lovelace, Mark A. Scheel, and Béla Szilágyi. 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):044028, February 2017.
- 47. Chad R. Galley. RomPy package, 2020. https://bitbucket.org/chadgalley/rompy/.
- 48. Curt Cutler and Eanna E. Flanagan. Gravitational waves from merging compact binaries: How accurately can one extract the binary's parameters from the inspiral waveform? *Phys. Rev. D*, 49(6):2658–2697, Mar 1994.
- 49. Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.
- 50. Vinod Nair and Geoffrey E. Hinton. Rectified linear units improve restricted boltzmann machines. In *ICML*, pages 807–814, 2010.
- Hao Zheng, Zhanlei Yang, Wenju Liu, Jizhong Liang, and Yanpeng Li. Improving deep neural networks using softplus units. 2015 International Joint Conference on Neural Networks (IJCNN), pages 1–4, 2015.
- 52. scikit learn. 6.3. preprocessing data. https://scikit-learn.org/stable/modules/preprocessing.html.
- 53. Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th international conference on Machine learning*, pages 1096–1103, 2008.
- Junyuan Xie, Ross Girshick, and Ali Farhadi. Unsupervised deep embedding for clustering analysis. In *International conference on machine learning*, pages 478–487, 2016.
- 55. Paraskevi Nousi and Anastasios Tefas. Self-supervised autoencoders for clustering and classification. *Evolving Systems*, pages 1–14, 2018.
- Paraskevi Nousi and Anastasios Tefas. Deep learning algorithms for discriminant autoencoding. *Neurocomputing*, 266:325–335, 2017.
- Paraskevi Nousi and Anastasios Tefas. Discriminatively trained autoencoders for fast and accurate face recognition. In *International Conference on Engineering Applications of Neural* Networks, pages 205–215. Springer, 2017.
- Pengcheng Wu, Steven CH Hoi, Hao Xia, Peilin Zhao, Dayong Wang, and Chunyan Miao. Online multimodal deep similarity learning with application to image retrieval. In *Proceedings* of the 21st ACM international conference on Multimedia, pages 153–162, 2013.
- Miguel A Carreira-Perpinán and Ramin Raziperchikolaei. Hashing with binary autoencoders.
 In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 557–566, 2015.
- Yi Pan, Alessandra Buonanno, Michael Boyle, Luisa T. Buchman, Lawrence E. Kidder, Harald P. Pfeiffer, and Mark A. Scheel. Inspiral-merger-ringdown multipolar waveforms

- of nonspinning black-hole binaries using the effective-one-body formalism. *Phys. Rev. D*, 84(12):124052, December 2011.
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.
- Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary De-Vito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. In NIPS-W, 2017.
- Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.
- 64. Weichangfeng Guo, Daniel Williams, Ik Siong Heng, Hunter Gabbard, Yeong-Bok Bae, Gungwon Kang, and Zong-Hong Zhu. Mimicking mergers: mistaking black hole captures as mergers. MNRAS, 516(3):3847–3860, November 2022.
- 65. David Reitze, Rana X. Adhikari, Stefan Ballmer, Barry Barish, Lisa Barsotti, GariLynn Billingsley, Duncan A. Brown, Yanbei Chen, Dennis Coyne, Robert Eisenstein, Matthew Evans, Peter Fritschel, Evan D. Hall, Albert Lazzarini, Geoffrey Lovelace, Jocelyn Read, B. S. Sathyaprakash, David Shoemaker, Joshua Smith, Calum Torrie, Salvatore Vitale, Rainer Weiss, Christopher Wipf, and Michael Zucker. Cosmic Explorer: The U.S. Contribution to Gravitational-Wave Astronomy beyond LIGO. In *Bulletin of the American Astronomical Society*, volume 51, page 35, September 2019.
- 66. Michele Maggiore, Chris Van Den Broeck, Nicola Bartolo, Enis Belgacem, Daniele Bertacca, Marie Anne Bizouard, Marica Branchesi, Sebastien Clesse, Stefano Foffa, Juan García-Bellido, Stefan Grimm, Jan Harms, Tanja Hinderer, Sabino Matarrese, Cristiano Palomba, Marco Peloso, Angelo Ricciardone, and Mairi Sakellariadou. Science case for the Einstein telescope. *Journal of Cosmology and Astroparticle Physics*, 2020(3):050, March 2020.
- 67. Peter Couvares, Ian Bird, Ed Porter, Stefano Bagnasco, Michele Punturo, David Reitze, Stavros Katsanevas, Takaaki Kajita, Vicky Kalogera, Harald Lueck, David McClelland, Sheila Rowan, Gary Sanders, B. S. Sathyaprakash, David Shoemaker, and Jo van den Brand. Gravitational Wave Data Analysis: Computing Challenges in the 3G Era. arXiv e-prints, page arXiv:2111.06987, November 2021.
- Hunter Gabbard, Michael Williams, Fergus Hayes, and Chris Messenger. Matching matched filtering with deep networks for gravitational-wave astronomy. *Phys. Rev. Lett.*, 120:141103, Apr 2018.
- Daniel George and E. A. Huerta. Deep neural networks to enable real-time multimessenger astrophysics. *Phys. Rev. D*, 97:044039, Feb 2018.
- Timothy D. Gebhard, Niki Kilbertus, Ian Harry, and Bernhard Schölkopf. Convolutional neural networks: A magic bullet for gravitational-wave detection? *Phys. Rev. D*, 100(6):063015, September 2019.
- Roberto Corizzo, Michelangelo Ceci, Eftim Zdravevski, and Nathalie Japkowicz. Scalable auto-encoders for gravitational waves detection from time series data. Expert Systems with Applications, 151:113378, 2020.
- Marlin B. Schäfer, Frank Ohme, and Alexander H. Nitz. Detection of gravitational-wave signals from binary neutron star mergers using machine learning. *Phys. Rev. D*, 102:063015, Sep 2020.
- He Wang, Shichao Wu, Zhoujian Cao, Xiaolin Liu, and Jian-Yang Zhu. Gravitational-wave signal recognition of LIGO data by deep learning. *Phys. Rev. D*, 101(10):104003, May 2020.
- 74. Plamen G. Krastev. Real-time detection of gravitational waves from binary neutron stars using artificial neural networks. *Physics Letters B*, 803:135330, April 2020.
- 75. Vasileios Skliris, Michael R. K. Norman, and Patrick J. Sutton. Real-Time Detection of Unmodelled Gravitational-Wave Transients Using Convolutional Neural Networks. *arXiv e-prints*, page arXiv:2009.14611, September 2020.
- Yu-Chiung Lin and Jiun-Huei Proty Wu. Detection of gravitational waves using Bayesian neural networks. *Phys. Rev. D*, 103(6):063034, March 2021.
- E. A. Huerta, Asad Khan, Xiaobo Huang, Minyang Tian, Maksim Levental, Ryan Chard, Wei Wei, Maeve Heflin, Daniel S. Katz, Volodymyr Kindratenko, Dawei Mu, Ben Blaiszik, and

Ian Foster. Accelerated, scalable and reproducible AI-driven gravitational wave detection. *Nature Astronomy*, 5:1062–1068, July 2021.

- Tom Marianer, Dovi Poznanski, and J. Xavier Prochaska. A semisupervised machine learning search for never-seen gravitational-wave sources. *Mon. Not. R. Astr. Soc.*, 500(4):5408–5419, February 2021.
- Wei Wei, Asad Khan, E. A. Huerta, Xiaobo Huang, and Minyang Tian. Deep learning ensemble for real-time gravitational wave detection of spinning binary black hole mergers. *Physics Letters B*, 812:136029, January 2021.
- Shreejit Jadhav, Nikhil Mukund, Bhooshan Gadre, Sanjit Mitra, and Sheelu Abraham. Improving significance of binary black hole mergers in Advanced LIGO data using deep learning: Confirmation of GW151216. *Phys. Rev. D*, 104(6):064051, September 2021.
- 81. Pranshu Chaturvedi, Asad Khan, Minyang Tian, E. A. Huerta, and Huihuo Zheng. Inference-optimized ai and high performance computing for gravitational wave detection at scale. *Frontiers in Artificial Intelligence*, 5, 2022.
- 82. Sunil Choudhary, Anupreeta More, Sudhagar Suyamprakasam, and Sukanta Bose. SiGMa-Net: Deep learning network to distinguish binary black hole signals from short-duration noise transients. *arXiv e-prints*, page arXiv:2202.08671, February 2022.
- 83. Marlin B. Schäfer and Alexander H. Nitz. From one to many: A deep learning coincident gravitational-wave search. *Phys. Rev. D*, 105:043003, Feb 2022.
- 84. Francesco Pio Barone, Daniele Dell'Aquila, and Marco Russo. A Novel Multi-Layer Modular Approach for Real-Time Gravitational-Wave Detection. arXiv e-prints, page arXiv:2206.06004, June 2022.
- Marlin B Schäfer, Ondřej Zelenka, Alexander H Nitz, Frank Ohme, and Bernd Brügmann. Training strategies for deep learning gravitational-wave searches. *Physical Review D*, 105(4):043002, 2022.
- Grégory Baltus, Justin Janquart, Melissa Lopez, Harsh Narola, and Jean-René Cudell. Convolutional neural network for gravitational-wave early alert: Going down in frequency. *Phys. Rev. D*, 106:042002, Aug 2022.
- Michael Andrews, Manfred Paulini, Luke Sellers, Alexey Bobrick, Gianni Martire, and Haydn Vestal. DeepSNR: A deep learning foundation for offline gravitational wave detection. arXiv e-prints, page arXiv:2207.04749, July 2022.
- 88. Chetan Verma, Amit Reza, Gurudatt Gaur, Dilip Krishnaswamy, and Sarah Caudill. Can Convolution Neural Networks Be Used for Detection of Gravitational Waves from Precessing Black Hole Systems? *arXiv e-prints*, page arXiv:2206.12673, June 2022.
- 89. João Aveiro, Felipe F. Freitas, Márcio Ferreira, Antonio Onofre, Constança Providência, Gonçalo Gonçalves, and José A. Font. Identification of binary neutron star mergers in gravitational-wave data using object-detection machine learning models. *Phys. Rev. D*, 106(8):084059, October 2022.
- M. Andrés-Carcasona, A. Menéndez-Vázquez, M. Martínez, and Ll. M. Mir. Searches for mass-asymmetric compact binary coalescence events using neural networks in the LIGO/Virgo third observation period. *Phys. Rev. D*, 107(8):082003, April 2023.
- Jurriaan Langendorff, Alex Kolmus, Justin Janquart, and Chris Van Den Broeck. Normalizing Flows as an Avenue to Studying Overlapping Gravitational Wave Signals. *Phys. Rev. Lett.*, 130(17):171402, April 2023.
- Maximilian Dax, Stephen R. Green, Jonathan Gair, Michael Pürrer, Jonas Wildberger, Jakob H. Macke, Alessandra Buonanno, and Bernhard Schölkopf. Neural Importance Sampling for Rapid and Reliable Gravitational-Wave Inference. *Phys. Rev. Lett.*, 130(17):171403, April 2023.
- Sophie Bini, Gabriele Vedovato, Marco Drago, Francesco Salemi, and Giovanni A. Prodi. An autoencoder neural network integrated into gravitational-wave burst searches to improve the rejection of noise transients. Classical and Quantum Gravity, 40(13):135008, July 2023.
- Minyang Tian, E. A. Huerta, and Huihuo Zheng. Physics-inspired spatiotemporal-graph AI ensemble for gravitational wave detection. arXiv e-prints, page arXiv:2306.15728, June 2023.
- Chinthak Murali and David Lumley. Detecting and denoising gravitational wave signals from binary black holes using deep learning. *Phys. Rev. D*, 108(4):043024, August 2023.

- 96. Philippe Bacon, Agata Trovato, and Micha I Bejger. Denoising gravitational-wave signals from binary black holes with a dilated convolutional autoencoder. *Machine Learning: Science and Technology*, 4(3):035024, September 2023.
- Alistair McLeod, Daniel Jacobs, Chayan Chatterjee, Linqing Wen, and Fiona Panther. Rapid Mass Parameter Estimation of Binary Black Hole Coalescences Using Deep Learning. arXiv e-prints, page arXiv:2201.11126, January 2022.
- Richard Qiu, Plamen G. Krastev, Kiranjyot Gill, and Edo Berger. Deep learning detection and classification of gravitational waves from neutron star-black hole mergers. *Physics Letters B*, 840:137850, May 2023.
- Shang-Jie Jin, Yu-Xin Wang, Tian-Yang Sun, Jing-Fei Zhang, and Xin Zhang. Rapid identification of time-frequency domain gravitational wave signals from binary black holes using deep learning. arXiv e-prints, page arXiv:2305.19003, May 2023.
- 100. Tiago Fernandes, Samuel Vieira, Antonio Onofre, Juan Calderón Bustillo, Alejandro Torres-Forné, and José A. Font. Convolutional neural networks for the classification of glitches in gravitational-wave data streams. Classical and Quantum Gravity, 40(19):195018, October 2023.
- 101. Osvaldo Gramaxo Freitas, Juan Calderón Bustillo, José A. Font, Solange Nunes, Antonio Onofre, and Alejandro Torres-Forné. Comparison of neural network architectures for feature extraction from binary black hole merger waveforms. arXiv e-prints, page arXiv:2307.16668, July 2023.
- 102. Marlin B. Schäfer, Ondřej Zelenka, Alexander H. Nitz, He Wang, Shichao Wu, Zong-Kuan Guo, Zhoujian Cao, Zhixiang Ren, Paraskevi Nousi, Nikolaos Stergioulas, Panagiotis Iosif, Alexandra E. Koloniari, Anastasios Tefas, Nikolaos Passalis, Francesco Salemi, Gabriele Vedovato, Sergey Klimenko, Tanmaya Mishra, Bernd Brügmann, Elena Cuoco, E. A. Huerta, Chris Messenger, and Frank Ohme. MLGWSC-1: The first Machine Learning Gravitational-Wave Search Mock Data Challenge. arXiv e-prints, page arXiv:2209.11146, September 2022.
- 103. R. Abbott et al. GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run. arXiv e-prints, page arXiv:2111.03606, November 2021.
- 104. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern* recognition, pages 770–778, 2016.
- 105. Nikolaos Passalis, Anastasios Tefas, Juho Kanniainen, Moncef Gabbouj, and Alexandros Iosifidis. Deep adaptive input normalization for price forecasting using limit order book data. IEEE Transactions on Neural Networks and Learning Systems, 2019.
- 106. Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In Advances in Neural Information Processing Systems 32, pages 8024–8035. Curran Associates, Inc., 2019.
- R. Abbott et al. Open Data from the Third Observing Run of LIGO, Virgo, KAGRA, and GEO. Astrophys. J. Suppl., 267(2):29, 2023.
- 108. Geraint Pratten, Cecilio García-Quirós, Marta Colleoni, Antoni Ramos-Buades, Héctor Estellés, Maite Mateu-Lucena, Rafel Jaume, Maria Haney, David Keitel, Jonathan E. Thompson, and Sascha Husa. Computationally efficient models for the dominant and subdominant harmonic modes of precessing binary black holes. *Phys. Rev. D*, 103:104056, May 2021.
- 109. Samantha A Usman, Alexander H Nitz, Ian W Harry, Christopher M Biwer, Duncan A Brown, Miriam Cabero, Collin D Capano, Tito Dal Canton, Thomas Dent, Stephen Fairhurst, et al. The pycbc search for gravitational waves from compact binary coalescence. Classical and Quantum Gravity, 33(21):215004, 2016.
- Nikolaos Passalis, Juho Kanniainen, Moncef Gabbouj, Alexandros Iosifidis, and Anastasios Tefas. Forecasting financial time series using robust deep adaptive input normalization. *Journal of Signal Processing Systems*, 93(10):1235–1251, 2021.

111. Boris Hanin. Which neural net architectures give rise to exploding and vanishing gradients? *Advances in neural information processing systems*, 31, 2018.

- 112. Ross Wightman, Hugo Touvron, and Hervé Jégou. Resnet strikes back: An improved training procedure in timm. *arXiv preprint arXiv:2110.00476*, 2021.
- 113. Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014.
- 114. S. Klimenko, G. Vedovato, M. Drago, F. Salemi, V. Tiwari, G. A. Prodi, C. Lazzaro, K. Ackley, S. Tiwari, C. F. Da Silva, and G. Mitselmakher. Method for detection and reconstruction of gravitational wave transients with networks of advanced detectors. *Phys. Rev. D*, 93:042004, Feb 2016.
- 115. Marco Drago, Sergey Klimenko, Claudia Lazzaro, Edoardo Milotti, Gaby Mitselmakher, Valentin Necula, Brendan O'Brian, Giovanni Prodi, Francesco Salemi, Marek Szczepanczyk, Shubhanshu Tiwari, Vaibhav Tiwari, Gayathri V, Gabriele Vedovato, and Igor Yakushin. coherent waveburst, a pipeline for unmodeled gravitational-wave data analysis. *SoftwareX*, 14:100678, 06 2021.
- 116. Sergey Klimenko, Gabriele Vedovato, Valentin Necula, Francesco Salemi, Marco Drago, Rhys Poulton, Eric Chassande-Mottin, Vaibhav Tiwari, Claudia Lazzaro, Brendan O'Brian, Marek Szczepanczyk, Shubhanshu Tiwari, and V. Gayathri. cwb pipeline library: 6.4.1, December 2021.
- 117. Alex Nitz, Ian Harry, Duncan Brown, Christopher M. Biwer, Josh Willis, Tito Dal Canton, Collin Capano, Thomas Dent, Larne Pekowsky, Andrew R. Williamson, Soumi De, Miriam Cabero, Bernd Machenschalk, Duncan Macleod, Prayush Kumar, Francesco Pannarale, Steven Reyes, Gareth S Cabourn Davies, dfinstad, Sumit Kumar, Márton Tápai, Leo Singer, Sebastian Khan, Stephen Fairhurst, Alex Nielsen, Shashwat Singh, Thomas Massinger, Koustav Chandra, Shasvath, and Veronica-villa. gwastro/pycbc: v2.0.5 release of pycbc, July 2022.
- 118. Alexander H. Nitz, Sumit Kumar, Yi-Fan Wang, Shilpa Kastha, Shichao Wu, Marlin Schäfer, Rahul Dhurkunde, and Collin D. Capano. 4-OGC: Catalog of gravitational waves from compact-binary mergers. arXiv e-prints, page arXiv:2112.06878, December 2021.
- 119. T. Mishra, B. O'Brien, M. Szczepańczyk, G. Vedovato, S. Bhaumik, V. Gayathri, G. Prodi, F. Salemi, E. Milotti, I. Bartos, and S. Klimenko. Search for binary black hole mergers in the third observing run of advanced ligo-virgo using coherent waveburst enhanced with machine learning. *Phys. Rev. D*, 105:083018, Apr 2022.
- 120. Yuki Fujimoto, Kenji Fukushima, and Koichi Murase. Mapping neutron star data to the equation of state using the deep neural network. *Phys. Rev. D*, 101:054016, Mar 2020.
- 121. Yuki Fujimoto, Kenji Fukushima, and Koichi Murase. Methodology study of machine learning for the neutron star equation of state. *Phys. Rev. D*, 98:023019, Jul 2018.
- 122. Yuki Fujimoto, Kenji Fukushima, and Koichi Murase. Extensive studies of the neutron star equation of state from the deep learning inference with the observational data augmentation. *Journal of High Energy Physics*, 2021(3), mar 2021.
- 123. Márcio Ferreira and Constança Providência. Unveiling the nuclear matter eos from neutron star properties: a supervised machine learning approach. *Journal of Cosmology and Astroparticle Physics*, 2021(07):011, jul 2021.
- 124. Plamen G. Krastev. Translating neutron star observations to nuclear symmetry energy via deep neural networks. *Galaxies*, 10(1), 2022.
- 125. F. Morawski and M. Bejger. Neural network reconstruction of the dense matter equation of state derived from the parameters of neutron stars. *Astronomy & Astrophysics*, 642(A78):8, oct 2020.
- 126. Shriya Soma, Lingxiao Wang, Shuzhe Shi, Horst Stöcker, and Kai Zhou. Neural network reconstruction of the dense matter equation of state from neutron star observables. *Journal* of Cosmology and Astroparticle Physics, 2022(08):071, aug 2022.
- 127. Shriya Soma, Lingxiao Wang, Shuzhe Shi, Horst Stöcker, and Kai Zhou. Reconstructing the neutron star equation of state from observational data via automatic differentiation. *Physical Review D*, 107(8), apr 2023.

- 128. Ronaldo V. Lobato, Emanuel V. Chimanski, and Carlos A. Bertulani. Cluster structures with machine learning support in neutron star m-r relations. *Journal of Physics: Conference Series*, 2340(1):012014, sep 2022.
- 129. Ronaldo V. Lobato, Emanuel V. Chimanski, and Carlos A. Bertulani. Unsupervised machine learning correlations in eos of neutron stars, 2022.
- 130. Má rcio Ferreira, Valéria Carvalho, and Constança Providência. Extracting nuclear matter properties from the neutron star matter equation of state using deep neural networks. *Physical Review D*, 106(10), nov 2022.
- Plamen G. Krastev. A deep learning approach to extracting nuclear matter properties from neutron star observations, 2023.
- Daniela D. Doneva, Fethi M. Ramazanoğlu, Hector O. Silva, Thomas P. Sotiriou, and Stoytcho S. Yazadjiev. Scalarization. arXiv e-prints, page arXiv:2211.01766, November 2022.
- 133. Emanuele Berti, Enrico Barausse, Vitor Cardoso, Leonardo Gualtieri, Paolo Pani, Ulrich Sperhake, Leo C Stein, Norbert Wex, Kent Yagi, Tessa Baker, C P Burgess, Flá vio S Coelho, Daniela Doneva, Antonio De Felice, Pedro G Ferreira, Paulo C C Freire, James Healy, Carlos Herdeiro, Michael Horbatsch, Burkhard Kleihaus, Antonie Klein, Kostas Kokkotas, Jutta Kunz, Pablo Laguna, Ryan N Lang, Tjonnie G F Li, Tyson Littenberg, Andrew Matas, Saeed Mirshekari, Hirotada Okawa, Eugen Radu, Richard O'Shaughnessy, Bangalore S Sathyaprakash, Chris Van Den Broeck, Hans A Winther, Helvi Witek, Mir Emad Aghili, Justin Alsing, Brett Bolen, Luca Bombelli, Sarah Caudill, Liang Chen, Juan Carlos Degollado, Ryuichi Fujita, Caixia Gao, Davide Gerosa, Saeed Kamali, Hector O Silva, João G Rosa, Laleh Sadeghian, Marco Sampaio, Hajime Sotani, and Miguel Zilhao. Testing general relativity with present and future astrophysical observations. Classical and Quantum Gravity, 32(24):243001, dec 2015.
- C. Charmousis, A. Lehé bel, E. Smyrniotis, and N. Stergioulas. Astrophysical constraints on compact objects in 4d einstein-gauss-bonnet gravity. *Journal of Cosmology and Astroparticle Physics*, 2022(02):033, feb 2022.
- 135. Bhaskar Biswas, Evangelos Smyrniotis, Ioannis Liodis, and Nikolaos Stergioulas. A Bayesian investigation of the neutron star equation-of-state vs. gravity degeneracy. *arXiv e-prints*, page arXiv:2309.05420, September 2023.
- 136. Eric J. Michaud, Ziming Liu, and Max Tegmark. Precision machine learning. *Entropy*, 25(1), 2023