Gravitational-wave parameter inference using Deep Learning

João D. Álvares*, José A. Font†‡, Felipe F. Freitas§, Osvaldo G. Freitas*, António P. Morais§,
Solange Nunes*, Antonio Onofre*, Alejandro Torres-Forné†¶
*Centro de Física das Universidades do Minho e do Porto (CF-UM-UP), Universidade do Minho, 4710-057 Braga, Portugal
†Departamento de Astronomía y Astrofísica, Universitat de València, Dr. Moliner 50, 46100, Burjassot (València), Spain
‡Observatori Astronòmic, Universitat de València, Catedrático José Beltrán 2, 46980, Paterna (València), Spain
§Departamento de Física da Universidade de Aveiro
and Centre for Research and Development in Mathematics and Applications (CIDMA)
Campus de Santiago, 3810-183 Aveiro, Portugal
¶Max Planck Institute for Gravitational Physics (Albert Einstein Institute), D-14476 Potsdam-Golm, Germany

2021 International Conference on Content-Based Multimedia Indexing (CBMI) | 978-1-6654-4220-6/20/$31.00 ©2021 IEEE | DOI: 10.1109/CBMI50038.2021.9461893

***Abstract*—We explore machine learning methods to detect gravitational waves (GW) from binary black hole (BBH) mergers using deep learning (DL) algorithms. The DL networks are trained with gravitational waveforms obtained from BBH mergers with component masses randomly sampled in the range from 5 to 100 solar masses and luminosity distances from $100\,\mathrm{Mpc}$ to, at least, $2000\,\mathrm{Mpc}$. The GW signal waveforms are injected in public data from the O2 run of the Advanced LIGO and Advanced Virgo detectors, in time windows that do not coincide with those of known detected signals, and the data from each detector in the Advanced LIGO and Advanced Virgo network is combined into a unique RGB image. We show that a classifier network can be trained in order to detect the presence of GW signal with high accuracy. Furthermore, we train a regression network to perform parameter inference on BBH spectrogram data. Without significant optimization of our algorithms we manage to corroborate most of the BBH detections in the GWTC-1 and GWTC-2 catalogs, and obtain parameter inference results that are mostly consistent with published results by the LIGO-Virgo Collaboration in GWTC-1. In particular, our predictions for the chirp mass are compatible (up to 3σ) with the official values for 90% of events.**

***Index Terms*—GW astronomy, convolutional neural networks, spectrogram classification, bayesian neural networks**

I. Introduction

The detection of gravitational waves (GW) from binary black hole (BBH) mergers [1] during the first data-taking run (O1) of Advanced LIGO [2] was a remarkable milestone that opened up a new window for observing the cosmos. The European detector Advanced Virgo [3] joined the efforts during the second observing run (O2) which helped improve the sky localization of the sources. During O1 and O2 the LIGO Scientific Collaboration and the Virgo Collaboration (LVC) announced the confident detection of eleven GW signals from compact binary coalescences (CBC) [4]. The third science run (O3) ended on March 2020 after completing almost one full year of data-taking. Recently, the LVC has released their second GW transient catalog comprising the 39 CBC detections accomplished in the first six months of O3 [5].

The detection of GW signals from CBC relies on accurate waveform templates against which to perform match-filtered searches. Faithful templates can be built either by solving the gravitational field equations with numerical relativity techniques or by using approximations to the two-body problem in general relativity. Current gravitational waveform models (or approximants) combine analytical and numerical approaches and are able to describe the entire inspiral-merger-ringdown signal for a large variety of parameter combinations (see e.g. [6] and references therein). Once a CBC source is detected, the estimation of its characteristic physical parameters such as component masses, individual spins or distance, is based on Bayesian inference [7], [8], which can be computationally expensive as it may take days to obtain a sufficient number of posterior samples for BBH [9].

Machine learning (ML) and DL are bringing about a revolution in data analysis across a variety of fields and GW astronomy is not alien to that trend. In particular, the use of Deep Neural Networks (DNN) for classification and/or prediction tasks has become the standard on data analysis applications, ranging from medical diagnosis [10] to particle physics [11]. This trend has now organically been extended to GW astronomy, both for signal detection [12] and for detector characterization, by reducing the impact of noise artifacts or "glitches" of instrumental and environmental origin [13], [14]. Recent approaches to eliminate, or at least mitigate, the effect of glitches are discussed in [15], [16].

In this paper we explore the use of DL methods to both detect GW from BBH mergers and perform parameter inference using RGB spectrograms that combine open data from the Advanced LIGO and Advanced Virgo three-detector network. To achieve our goal we train a cross-residual network which allows us to extract information about source parameters such as luminosity distance, chirp mass, network antenna power, and effective spin.

II. Deep Neural Network: architectures and methodologies

We encode the information of simulated waveforms of BBH mergers, produced with pycbc using the `SEOBNRv4HM_ROM` approximant [17][18], into 3-channel RGB spectrograms (see

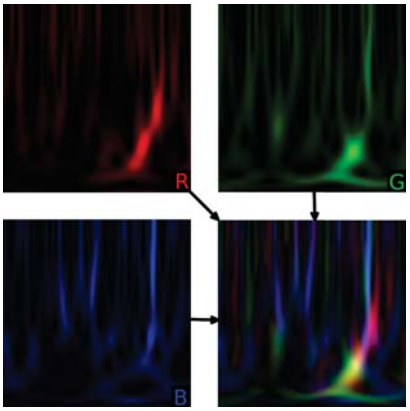

Fig. 1. Combining spectrogram data into a single RGB image, to be used as input to the deep learning networks. In all plots the x-axis is time and the y-axis is frequency, while each channel's colour intensity stands for the signal power normalised to the 8-bit integer range (0-255).

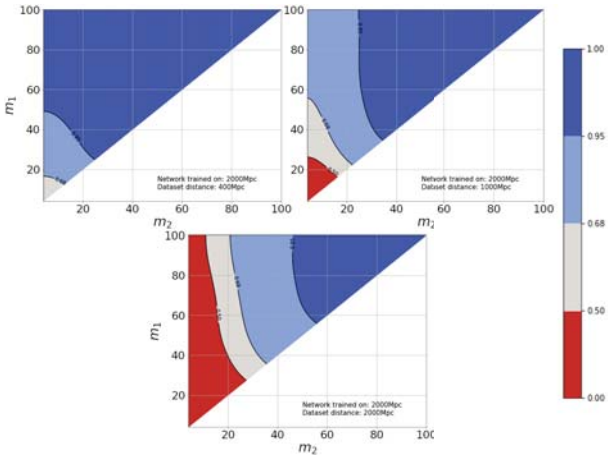

Fig. 2. Simulated signal scores for different luminosity distances and evaluated with a DL network trained with GW waveforms from BBH mergers at a luminosity distance of 2000 Mpc. Results are shown as a function of the BH masses of the binary system, m_1 and m_2, for GW signals from sources at 400 Mpc (top left), 1000 Mpc(top right) and 2000 Mpc(bottom).

Fig. 1), using a 0.2 s window, with frequencies between 20 Hz to 300 Hz. For more information on the composition of the datasets see Tab. I. We now describe how to apply our DL algorithms to identify GW information, using the classification and regression networks discussed in this section. A summary of the architectures of both networks is reported in Table Tab. II.

A. *Classification Network*

Our first task is to test whether a Deep Neural Network (DNN) can distinguish between possible signal events over a random background. For this, we choose a Residual Network (ResNet) [19], which consists of a DNN built as blocks of convolutional layers together with shortcut connections (to skip layers) that make them easier to optimize and overcome the "vanishing/exploding gradient" problem [20].

For the classification task the highest discriminant power was achieved with a ResNet-101 (see Tab. II), which consists of 101 layers, where in between each Conv2D layer we have a series of batch normalizations, average pooling and rectified activations (ReLU). For our task, we have replaced the last fully connected layers of the ResNet-101, responsible for the classification, with the following sequence of layers:

- an adaptive concatenate pooling layer[1],
- a flatten layer,
- a block with batch normalization, dropout, linear, and ReLU layers,
- a dense linear layer with 2 units as outputs, each unit corresponding to *signal* or *background* class and a softmax activation function, outputting a *score* between 0 and 1, such that the sum of both outputs equals 1.

The `AdaptiveConcatPool2d` layer uses adaptive average pooling and adaptive max pooling and concatenates them both.

Learning rate and weight decay are two key hyperparameters to train DNNs. A good choice of these two parameters can greatly improve the model performance. In our particular case it implies a high accuracy classification and a good background rejection, while drastically reducing the training time. Instead of using a fixed value for the learning rate we opted to use Cyclical Learning Rates (CLR) [21]. Following the guidelines from Ref. [22], we have performed a scan over a selected range of values for both learning rates and weight decays. In our case, we have found the optimal values to be 2×10^{-3} for the learning rate and 1×10^{-5} for the weight decay, while for the maximum learning rate value we just multiply the initial value by 10.

[1] `AdaptiveConcatPool2d`

B. *Regression Network*

We based the regression network architecture on a Cross-Residual Network (xResNet; see Table II) [23] and following the guidelines in [24] replaced the average pooling layers with blur pooling ones. Furthermore, we made use of Dropout layers before pooling in order to approximate a Bayesian variational inference process. This has given us a way of estimating the network's uncertainty in the parameter inference at testing time using Monte Carlo (MC) dropout [25]. For training, we use once again the CLR, with 1×10^{-2} as the initial value for the learning rate and 1×10^{-3} as the weight decay. It is important to mention that we use the spectrogram images generated from the GW signals to infer continuous values for variables such as the chirp mass $\mathcal{M} \equiv \frac{(m_1 m_2)^{3/5}}{(m_1+m_2)^{1/5}}$ of the BBH system or the luminosity distance of the source d_L. While this approach seems to be rather non-intuitive, convolutional neural networks carry inductive biases rooted in translation invariance. Such biases are a direct consequence of the convolutional filters and can be used to extract information from patterns in the spectrogram images and correlated to physical continuous variables.

Classification		
Parameters	Train size	Validation size
Multiple detector $(m_1, m_2) \sim U(5, 100)\,\mathrm{M}_\odot$, $d_L = [100, 300, 1000, 1500, 2000]\mathrm{Mpc}$, $\iota = \frac{\pi}{2}$,	4000 images 560 × 560 pixels 8-bit RGB	1000 images 560 × 560 pixels 8-bit RGB
Total images	20000	5000

Regression		
Parameters	Train size	Validation size
Multiple detector $(m_1, m_2) \sim U(5, 100)\ M_\odot$, $d_L \sim U(100, 4000)\mathrm{Mpc}$ $\iota \sim U(0, \pi)$, $spin \sim U(-1, 1)$, restriction: SNR > 5	22049 images 224 × 224 pixels 8-bit RGB	9450 images 224 × 224 pixels 8-bit RGB

TABLE I

DESCRIPTION OF THE CLASSIFICATION AND REGRESSION DATASETS FOR TRAINING AND VALIDATION. THE IMAGES ARE GENERATED FROM WAVEFORMS CALCULATED BY PYCBC, USING THE SEOBNRV4HM_ROM APPROXIMANT. FOR THE CLASSIFICATION DATASETS, THE INDIVIDUAL MASSES (m_1, m_2) ARE SAMPLED WITH AN UNIFORM DISTRIBUTION WITHIN THE RANGE OF 5 TO 100 $M_\odot$. FOR REGRESSION, THE PARAMETERS ARE ALSO UNIFORMLY SAMPLED.

Classification		
Base architecture	Hyperparameters	Metric performance
ResNet-101 + custom header	input size: $275 \times 275 \times 3(1)$, batch size: 8 images, learning rate: $[2 \times 10^{-5}, 2 \times 10^{-3}]$, weight decay: 1×10^{-5}, loss function: Cross Entropy Loss (CE)	AUC: 0.82

Regression		
Base architecture	Hyperparameters	RMSE
xResNet-18 + Blur average layer + MC Dropout + custom header	input size: $128 \times 128 \times 3$, batch size: 64 images, max learning rate: $[1 \times 10^{-4}, 2 \times 10^{-2}]$, weight decay: 1×10^{-3}, loss function: Mean Squared Error (MSE)	RMSE: 0.021

TABLE II

CONVOLUTIONAL NEURAL NETWORKS ARCHITECTURES EMPLOYED FOR THE CLASSIFICATION, MULTIPLE (SINGLE) DETECTORS, AND REGRESSION TASKS. THE CUSTOM HEADER FOR THE CLASSIFICATION DNN IS DESCRIBED IN SEC. II-A, THE CUSTOM HEADER FOR THE REGRESSION MODEL HAS THE SAME STRUCTURE WITH THE MAIN DIFFERENCE THAT THE FINAL LAYER HAS ONLY ONE UNIT WITH A LINEAR ACTIVATION FUNCTION.

III. NETWORK ASSESSMENT

A. Classifier

For the classifier network we found that training for lower amplitude signals (i.e. larger distances) allowed the model to detect both low-amplitude and high-amplitude signals. As such, going forwards we chose to use the network trained on signals injected at 2000 Mpc. Figure 2 shows the performance of the model as a function of the binary component masses. It is noticeable that larger masses and smaller distances (higher amplitudes) result in higher scores. Fig. 3 exhibits the receiver operating characteristic (ROC) curve for our network, using 2000 Mpc data. The x-axis shows the fraction of background-only spectrograms that are successfully rejected by the network, $1 - \varepsilon_B$, while the y-axis represents the fraction of successfully detected signals, ε_S. As can be seen in the ROC curve, we could alter the threshold for classification to be more or less strict, according to the necessities of the problem at hand.

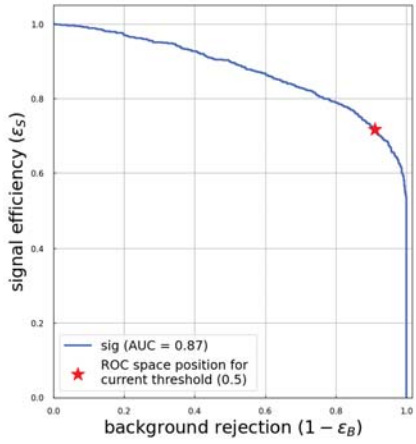

Fig. 3. ROC curve for the best-performing classifier. The red star displays the current threshold location on the ROC used for classify the events into signal (score ≥ 0.5) or background (score ≤ 0.5).

B. Regression

1) Luminosity Distance Regression: Figure 4 shows the performance of d_L regression for our model on the validation set. Each event is evaluated 100 times using MC dropout and the mean value of the regressed parameter outputs are stored in the histograms. The white dashed diagonal line shows the

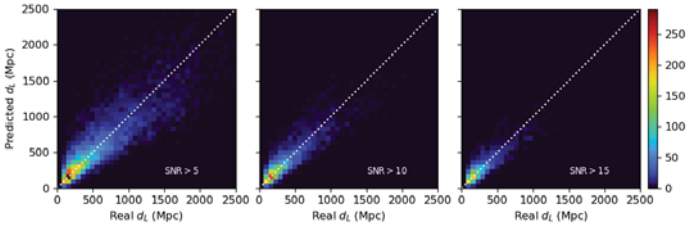

Fig. 4. Calibration for d_L. Columns show different thresholds for the signal to noise ratio: SNR > 5 (left), SNR > 10 (middle) and SNR > 15 (right).

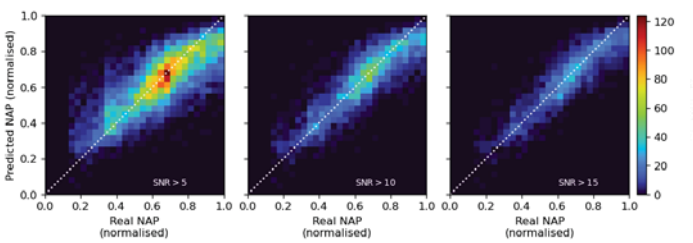

Fig. 5. Calibration for NAP. Columns show different thresholds for the signal to noise ratio: SNR > 5 (left), SNR > 10 (middle) and SNR > 15 (right).

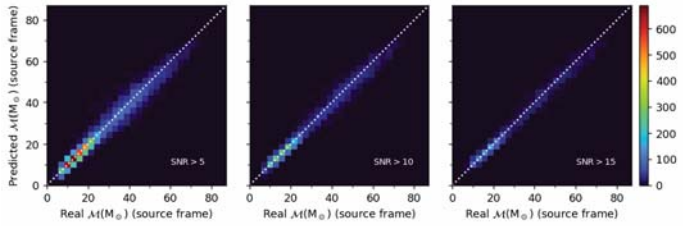

Fig. 6. Calibration for $\mathcal{M}$. Columns show different thresholds for the signal to noise ratio: SNR > 5 (left), SNR > 10 (middle) and SNR > 15 (right).

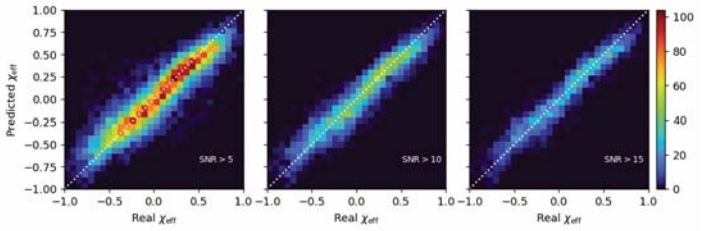

Fig. 7. Calibration for $\chi_{\rm eff}$. Columns show different thresholds for the signal to noise ratio: SNR > 5 (left), SNR > 10 (middle) and SNR > 15 (right).

ideal behaviour. For the lowest SNR threshold (left column) the deviation from the ideal behaviour can be quite large. However, as the threshold increases to SNR>10 and SNR>15 we are able to more confidently resolve the distances. On the other hand, these thresholds mean that we lose the ability to resolve distances for which larger SNRs are also rarer.

2) Network Antenna Power Regression: in Fig. 5 we show the performance of our model in the regression of the network antenna power (NAP) parameter. Note that almost no events with NAP<0.2 are present, which is to be expected due to the SNR requirements during event generation. For the SNR>5 threshold (left column) there seem to be two separate populations, one that broadly follows the diagonal line and a second one that roughly follows a horizontal line around the 0.6 mark (likely due to the model assigning an overall average NAP value to inputs whose relevant features it cannot resolve). As we increase the SNR threshold, this second population fades away and we isolate a population of predictions that nicely follows the diagonal line.

3) Chirp Mass Regression: Figure 6 shows the behaviour of the predicted source frame chirp masses. Our predictions for the chirp masses nicely follow the actual injected values as the events closely cluster along the diagonal lines in the plots. Of all variables we employ to calibrate our method, the chirp mass is the one that shows the smallest average deviation from the ideal results. The distribution for SNR>5 already displays a fairly low error in the predictions and as we increase the SNR threshold the error further shrinks.

4) Effective Inspiral Spin Regression: to end the discussion of the calibration of our model we show in Fig. 7 the predictions for $\chi_{\rm eff}$ compared to the real values. We see that all models follow closely the ideal diagonal line, with the faithfulness of the distribution width increasing as we raise the SNR threshold.

IV. Analysis of real GW detections

We turn now to assess the model's performance on actual GW detections. To this end we explore the BBH detections published in the GW transient catalogs from the LVC, GWTC-1 [4] and GWTC-2 [5], though due to length constraints we will perform parameter inference only on GWTC-1 events.

A. Classifier

To analyse the real GW events we produce RGB spectrograms using publicly available data for all GWTC-1 and GWTC-2 BBH events, combining the data from Hanford (R), Livingston (G) and Virgo (B). We leave out the binary neutron star events (GW170817, GW190425 and GW190426) as those cases involve different time scales and are not present in training. However, we include GW190814 despite the fact that it involves a $\sim 2.6 M_\odot$ compact object, since its precise nature remains undetermined [26]. In addition to the confident detections from GWTC-1 and GWTC-2, we also analyze the marginal subthreshold triggers for GWTC-1.

The results of our classifier are presented in Table III. All confident detections reported by the LVC for GWTC-1 are corroborated by our classifier: GW150914, GW170104, GW170814 and GW170823 are all given the score of 1.00, the highest value possible. From the remaining events, 0.87 was the lowest score obtained. When we analyse the marginal detections from GWTC-1 we obtain, as expected, much lower scores across the board. Under the default threshold for detection, which assumes a score of 0.50 or higher, only three events are classified as signal. These are MC151116, MC161217 and MC170705, with scores of 0.73, 0.72 and 0.51 respectively. The first two in particular may deserve a more careful analysis in the future, but this stays outside the scope of

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

TABLE III
CLASSIFIER SCORES FOR GWTC-1 CONFIDENT DETECTIONS (LEFT), GWTC-1 MARGINAL DETECTIONS (MIDDLE) AND GWTC-2 DETECTIONS (RIGHT).

this paper. Keeping in mind that the classifier is not optimized for the O3a run, which has a significantly lower noise floor, we look also at the new BBH events of the GWTC-2 catalog. Despite the lack of any optimization, we find that 34 out of 37 events are given a score above our threshold for detection, and from these, 27 (31) events are given a score above 0.90 (0.70). The highest possible score is obtained for a subset of 16 events.

B. Parameter Inference on GWTC-1

To perform parameter inference on GWTC-1 events we used MC dropout to pass the spectrogram corresponding to each event to our model 1500 times. Due to the properties of MC dropout the output of this is an array of normally-distributed values, from which we calculate the mean and standard deviation for the values of the chirp mass, luminosity distance and effective inspiral spin. Figure 8 shows the comparison between the results obtained with our model and the values published in the GWTC-1 paper: the green shading covers the LVC published values for the 90% interval for each parameter, the blue shading covers our MC dropout inferred 90% interval, and the red bar delimits our 3σ range. We observe a remarkable compatibility with published data with an uncertainty up to three standard deviations, with the one exception being GW151226, where our model overestimates $\mathcal{M}$ and underestimates $\chi_{\rm eff}$ when compared with the LVC results. The reasons for this may be related to the homoskedastic nature of MC dropout. Furthermore, due to the significant size of the parameter space, a larger dataset may be required. This warrants further investigation.

V. CONCLUSIONS

In this work we introduce Deep Learning (DL) methods to study gravitational waves from BBH mergers, using spectrograms created from Advanced LIGO and Advanced Virgo open data. For black holes of varying mass and zero spin, we have trained a residual network classifier, which has been applied on GWTC-1 and GWTC-2 detections, corroborating 34 out of 37 confident results with high scores. An analysis of marginal triggers has identified 3 cases (MC151116, MC161217 and MC170705) as GW signals, rejecting all others.

We have also trained a cross-residual network to perform parameter estimations on GW spectrogram data from BBH mergers. Using MC dropout we have obtained a natural estimation of the uncertainty of our predictions. We have shown that it is possible to perform parameter inference on the distance, chirp mass, network antenna power and effective inspiral spin. Applying this network to spectrogram data from GWTC-1 BBH events, we have found a remarkable agreement with the results published by the LVC in the case of d_L estimations. Most of our chirp mass and effective spin estimations are also compatible with the published 90% confidence intervals up to an MC dropout uncertainty of 3σ, with the exception of GW151226. Going forward, optimizing the network architecture could provide further improvements on our results. Other physical effects of BBH mergers, such as orbital plane precession or eccentricity, may also be explored. For building a full detection pipeline, a more thorough comparison with existing methods should be done.

ACKNOWLEDGEMENTS

We thank Nicolás Sanchis-Gual for fruitful discussions that allowed the setup of the team involved in this project. This work was supported by the Spanish Agencia Estatal de Investigación (PGC2018-095984-B-I00), by the Generalitat Valenciana (PROMETEO/2019/071), by the EU's Horizon 2020 research and innovation (RISE) programme (H2020-MSCA-RISE-2017 Grant No. FunFiCO-777740) and by the Portuguese Foundation for Science and Technology (FCT), project CERN/FIS-PAR/0029/2019. APM and FFF are supported by the FCT project PTDC/FIS-PAR/31000/2017 and

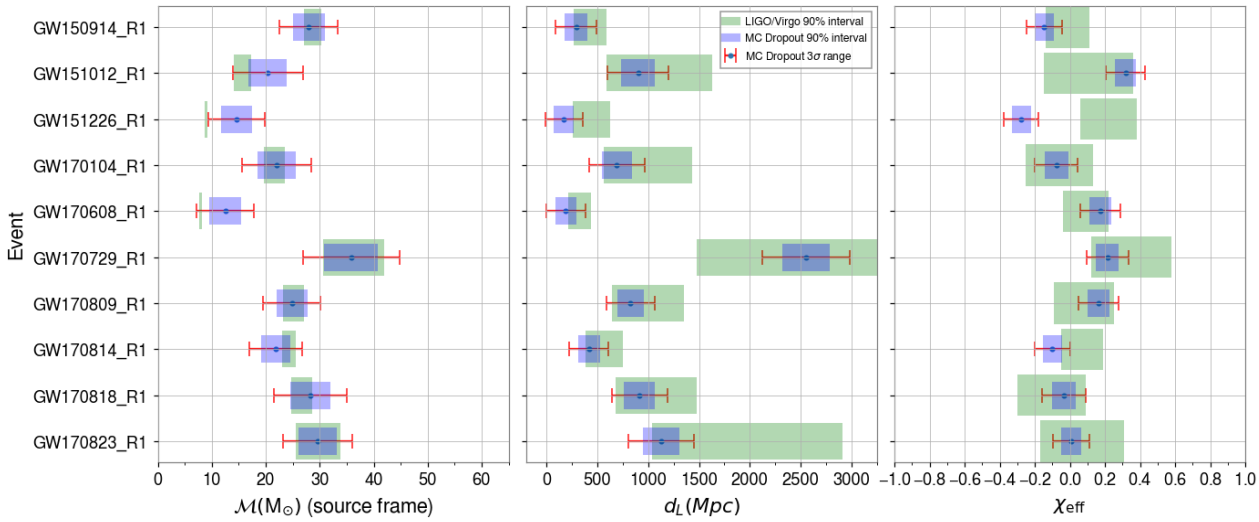

Fig. 8. Predictions of the DL network for the chirp mass (left), luminosity distance (middle) and effective inspiral spin (left).

by CIDMA through FCT, references UIDB/04106/2020 and UIDP/04106/2020. APM is also supported by the projects CERN/FIS-PAR/0027/2019, CERN/FISPAR/0002/2017 and by national funds (OE), through FCT, I.P., in the scope of the framework contract foreseen in the numbers 4, 5 and 6 of the article 23, of the Decree-Law 57/2016, of August 29, changed by Law 57/2017, of July 19.

References

[1] B. P. Abbott *et al.*, "Observation of Gravitational Waves from a Binary Black Hole Merger," *Physical Review Letters*, vol. 116, no. 6, 061102, p. 061 102, Feb. 2016. DOI: 10.1103/PhysRevLett.116.061102. arXiv: 1602.03837 [gr-qc].

[2] LIGO Scientific Collaboration *et al.*, "Advanced LIGO," *Classical and Quantum Gravity*, vol. 32, no. 7, 074001, p. 074 001, Apr. 2015. DOI: 10.1088/0264-9381/32/7/074001. arXiv: 1411.4547 [gr-qc].

[3] F. Acernese and et al., "Advanced Virgo: a second-generation interferometric gravitational wave detector," *Class. Quant. Grav.*, vol. 32, no. 2, p. 024 001, 2015. DOI: 10.1088/0264-9381/32/2/024001. arXiv: 1408.3978 [gr-qc].

[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," *Physical Review X*, vol. 9, no. 3, 031040, p. 031 040, Jul. 2019. DOI: 10.1103/PhysRevX.9.031040. arXiv: 1811.12907 [astro-ph.HE].

[5] R. Abbott *et al.*, "GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run," *arXiv e-prints*, arXiv:2010.14527, arXiv:2010.14527, Oct. 2020. arXiv: 2010.14527 [gr-qc].

[6] J. Blackman *et al.*, "Numerical relativity waveform surrogate model for generically precessing binary black hole mergers,", vol. 96, no. 2, 024058, p. 024 058, Jul. 2017. DOI: 10.1103/PhysRevD.96.024058. arXiv: 1705.07089 [gr-qc].

[7] J. Veitch *et al.*, "Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library,", vol. 91, no. 4, 042003, p. 042 003, Feb. 2015. DOI: 10.1103/PhysRevD.91.042003. arXiv: 1409.7215 [gr-qc].

[8] G. Ashton *et al.*, "BILBY: A User-friendly Bayesian Inference Library for Gravitational-wave Astronomy,", vol. 241, no. 2, 27, p. 27, Apr. 2019. DOI: 10.3847/1538-4365/ab06fc. arXiv: 1811.02042 [astro-ph.IM].

[9] S. R. Green *et al.*, "Gravitational-wave parameter estimation with autoregressive neural network flows," *arXiv e-prints*, arXiv:2002.07656, arXiv:2002.07656, Feb. 2020. arXiv: 2002.07656 [astro-ph.IM].

[10] A. Elkins *et al.*, "Developing an app to interpret chest x-rays to support the diagnosis of respiratory pathology with artificial intelligence," *Journal of Medical Artificial Intelligence*, vol. 3, no. 0, 2020. [Online]. Available: http://jmai.amegroups.com/article/view/5545.

[11] F. F. Freitas *et al.*, "Exploring the standard model EFT in VH production with machine learning," *Phys. Rev. D*, vol. 100, no. 3, p. 035 040, 2019. DOI: 10.1103/PhysRevD.100.035040. arXiv: 1902.05803 [hep-ph].

[12] T. D. Gebhard *et al.*, "Convolutional neural networks: a magic bullet for gravitational-wave detection?" *Phys. Rev. D*, vol. 100, no. 6, p. 063 015, 2019. DOI: 10.1103/PhysRevD.100.063015. arXiv: 1904.08693 [astro-ph.IM].

[13] R. Biswas *et al.*, "Application of machine learning algorithms to the study of noise artifacts in gravitational-wave data,", vol. 88, no. 6, 062003, p. 062 003, Sep. 2013. DOI: 10.1103/PhysRevD.88.062003. arXiv: 1303.6984 [astro-ph.IM].

[14] R. E. Colgan *et al.*, "Efficient gravitational-wave glitch identification from environmental data through machine learning," *Phys. Rev. D*, vol. 101, p. 102 003, 10 May 2020. DOI: 10.1103/PhysRevD.101.102003. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.101.102003.

[15] G. Vajente *et al.*, "Machine-learning nonstationary noise out of gravitational-wave detectors," *Phys. Rev. D*, vol. 101, p. 042 003, 4 Feb. 2020. DOI: 10.1103/PhysRevD.101.042003. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.101.042003.

[16] A. Torres-Forné *et al.*, "Application of dictionary learning to denoise ligo's blip noise transients," *Phys. Rev. D*, vol. 102, p. 023 011, 2 Jul. 2020. DOI: 10.1103/PhysRevD.102.023011. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.102.023011.

[17] R. Cotesta *et al.*, "Frequency-domain reduced-order model of aligned-spin effective-one-body waveforms with higher-order modes," en, *Physical Review D*, vol. 101, no. 12, p. 124 040, Jun. 2020, ISSN: 2470-0010, 2470-0029. DOI: 10.1103/PhysRevD.101.124040.

[18] A. Nitz *et al.*, *Gwastro/pycbc: Pycbc release v1.16.11*, version v1.16.11, Oct. 2020. DOI: 10.5281/zenodo.4075326. [Online]. Available: https://doi.org/10.5281/zenodo.4075326.

[19] K. He *et al.*, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015. arXiv: 1512.03385. [Online]. Available: http://arxiv.org/abs/1512.03385.

[20] Y. Bengio *et al.*, "Learning long-term dependencies with gradient descent is difficult," *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 157–166, Mar. 1994, ISSN: 1045-9227. DOI: 10.1109/72.279181.

[21] L. N. Smith, "No more pesky learning rate guessing games," *CoRR*, vol. abs/1506.01186, 2015. arXiv: 1506.01186. [Online]. Available: http://arxiv.org/abs/1506.01186.

[22] ——, "A disciplined approach to neural network hyper-parameters: Part 1 - learning rate, batch size, momentum, and weight decay," *CoRR*, vol. abs/1803.09820, 2018. arXiv: 1803.09820. [Online]. Available: http://arxiv.org/abs/1803.09820.

[23] T. He *et al.*, "Bag of Tricks for Image Classification with Convolutional Neural Networks," *arXiv e-prints*, arXiv:1812.01187, arXiv:1812.01187, Dec. 2018. arXiv: 1812.01187 [cs.CV].

[24] R. Zhang, "Making Convolutional Networks Shift-Invariant Again," *arXiv e-prints*, arXiv:1904.11486, arXiv:1904.11486, Apr. 2019. arXiv: 1904.11486 [cs.CV].

[25] Y. Gal and Z. Ghahramani, "Bayesian Convolutional Neural Networks with Bernoulli Approximate Variational Inference," *arXiv e-prints*, arXiv:1506.02158, arXiv:1506.02158, Jun. 2015. arXiv: 1506.02158 [stat.ML].

[26] R. Abbott *et al.*, "GW190814: Gravitational Waves from the Coalescence of a 23 Solar Mass Black Hole with a 2.6 Solar Mass Compact Object,", vol. 896, no. 2, L44, p. L44, Jun. 2020. DOI: 10.3847/2041-8213/ab960f. arXiv: 2006.12611 [astro-ph.HE].