An Multi-step Prediction Algorithm for Analysis of Gravitational Waves Based on Deep Learning

Chunlin Luo
Shanghai Jiao Tong University
Department of Electronic Engineering
And Shanghai Frontiers Science Center for Gravitational Wave Detection
Shanghai 200240, China
e-mail: luochunlin@sjtu.edu.cn

Yuewei Zhang
Shanghai Jiao Tong University
Department of Electronic Engineering
And Shanghai Frontiers Science Center for Gravitational Wave Detection
Shanghai 200240, China
e-mail: yueweizhang@sjtu.edu.cn

Jie Zhu
Shanghai Jiao Tong University
Department of Electronic Engineering
And Shanghai Frontiers Science Center for Gravitational Wave Detection
Shanghai 200240, China
e-mail: zhujie@sjtu.edu.cn

***Abstract*—The detection of Gravitational Wave (GW) events necessitates a vast number of precise GW templates. Improving the accuracy and effectiveness of template waveform synthesis remains a significant challenge. This problem is interesting because accurate templates are critical for identifying GW signals from astronomical events such as Binary Black Hole (BBH) mergers. Previous methods have struggled with cumulative errors and limited parameter ranges, affecting template reliability. We propose a Multi-step prediction method based on deep learning to enhance waveform prediction accuracy. Our approach achieves over 99.6% mean template matching accuracy on a test set of 100,000 waveforms and performs well across a broader parameter range. The main conclusion is that our method significantly improves prediction accuracy and efficiency, facilitating better GW event detection.**

Keywords-Gravitational Wave; Waveform Prediction; Multi-step Prediction; Deep Learning;

I. INTRODUCTION

Since Albert Einstein predicted that GW would exist. A new era in astrophysical study began with the announcement in 2016 by the LIGO Scientific Collaboration of the first detection of GW [1]. An increasing number of countries and organizations have been joining in the research of GW. It is primarily concerned with a number of things, including target for GW identification [2]-[4], parameter estimate [5][6], denoising [7][8], and rapid waveform production [9][10]. The improvement of waveform modeling's production speed and waveform accuracy cannot be disregarded as a fundamental tool for GW astrophysics. In recent years, very promising progress has been made in the generation of GW [11][12]. Traditional optimization methods require specialized knowledge of physics and astronomy, which is difficult for novice GW research enthusiasts. And there is limited scope for optimizing the use of computational resources and the accuracy of waveform modeling.

With the rise of deep learning, various neural network models such as Convolutional Neural Networks (CNN) [13], Deep Neural Networks (DNN) [14], Recurrent Neural Networks(RNN) [15], Long and Short-Term Memory (LSTM) [16], Transformer [17], Generating Adversarial Networks (GAN) [18], etc. have excellent performance in various fields such as video, image, and speech. Deep learning has great potential for optimizing and investigating issues in the study of GW astrophysics. Waveform prediction is an essential part of the GW field. It can provide proposals on GW modeling, as well as on GW detection [2] and parameter estimation. It also helps to understand the physical process of GW formation from a deep learning level. There have been related studies applying deep learning methods to waveform generation and prediction. The conditional GAN (cGAN) model has been used to rapidly produce a large number of random waveforms [19]. A single-step prediction method has used the transformer model to forecast the remaining GW, etc [20]. These attempts have yielded promising results.

Existing methods often suffer from cumulative errors and limited parameter ranges, affecting waveform reliability. Additionally, they require significant computational resources and domain-specific expertise, making them less accessible to new researchers. An ideal solution would minimize these errors, extend parameter ranges, and optimize computational efficiency without extensive specialized knowledge. Our contribution addresses these gaps with a multi-step prediction method based on deep learning, using the LSTM neural network. This method predicts subsequent merger and ringdown phases from a known pre-inspiral waveform segment, reducing errors and improving efficiency. The goal is to enhance waveform prediction accuracy and efficiency, making GW template generation more reliable and accessible.

The structure of GW is relatively simple, and there is a one-to-one mapping between structure and parameters. A known waveform segment containing sufficient information corresponds to the only GW with an inspire-merge-ringdown process. With a known waveform sequence, the remaining waveform sequence can be predicted without relying on information from other time nodes. This situation qualifies for multi-step prediction, which means that the unknown waveform sequence can be directly predicted by inputting a segment of the waveform sequence. We briefly compare the traditional CNN, LSTM, and Bi-directional Long Short-Term Memory (BiLSTM) models. Mainly considering the running

time and prediction accuracy, we use the LSTM neural network structure which is more straightforward and performs well in time series prediction. Input pre-inspire phase waveform containing sufficient information to the neural net and predict the post-inspire, merge, and ringdown parts. The predicted results are matched to the real waveforms and the matches were collated statistically.

In the next section, we will introduce the methods and models used in the experiments. First, the basic structure and principles of the multi-step forecasting method will be described. Then, three traditional deep learning models are compared to select the appropriate one in terms of model training time and variation of loss values. We've picked LSTM and briefly described its components and guiding ideas. The basic structure and principles of the model will be introduced. There is also the data preparation session before the experiment. In Section 3, we will predict the test dataset and match the results to the real waveform and statistically obtain the match distribution, etc. Then, we summarize the advantages and disadvantages of the experimental approach and look ahead in Section 4.

II. BACKGROUND

This section provides an overview of the experimental background. We utilize the IMRPhenomD waveform generator to simulate the merger process of aligned-spin BBH, selecting and generating waveforms based on parameters such as mass and spin. The number of predicted points is set to 230, covering various stages of the merger process. For model selection, we compare the loss functions and runtime of three traditional models: CNN, LSTM, and BiLSTM. Experimental results demonstrate that LSTM exhibits significant advantages in time series prediction, particularly in terms of loss function convergence speed and runtime. Therefore, given the available computational resources and data collection, LSTM is chosen as the optimal model.

A. Dataset

Our dataset is generated with the IMRPhenomD [21] waveforms from the open-source tools PyCBC and LALsuite [22] and simulate the inspire, merge, and ringdown of the aligned spin BBH merge model in the time domain. Since waveforms of longer durations of lesser quality produced by BBH were not useful for training, the masses (m_1, m_2) are distributed in the range [30,70] $\boldsymbol{M}_{\odot}$. The sampling rate is set to 1024 Hz and the waveform length is set at 1s to shorten the training period [23]. The distribution of the spins (s_1, s_2) fall between [-0.4, 0.4]. The waveform's amplitude is multiplied by 10^{22} since it is too modest in magnitude for the model to be trained at its original level. The capacity of the model to forecast waveforms without blindness was taken into account when generating the training data set. The mass is incremented in steps of $0.8\boldsymbol{M}_{\odot}$ and the spin is incremented in steps of 0.1. The test set is randomly generated using the parameter range of the training set with 100,000 waveforms. Set the number of predicted points to 230, including the post-inspire, merge, and ringdown components.

B. Model Comparison

An appropriate model should be chosen from a limited dataset by contrasting its loss function and running time. Three traditional models—CNN, LSTM, and BiLSTM—are initially trained on a small test dataset of 9,000 waveforms that is randomly produced. For waveform prediction, the loss function may be used to determine the loss value by calculating the difference between the real value and the predicted value. The model's ability to forecast outcomes may be correctly reflected by the trend of loss value and the comparison of loss values of the three models is shown in Figure 1.

TABLE I. RUNTIME OF CNN, LSTM, AND BILSTM.

No	Model	Run Time/s
1	CNN	1734.4
2	LSTM	3690.4
3	BiLSTM	8229.4

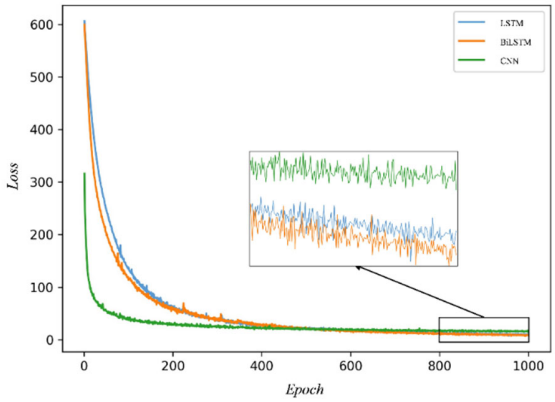

Figure 1. Loss of CNN, LSTM, BiLSTM

The loss values produced from the training model demonstrate how rapidly CNN converges. The loss function for our models is MSE to quickly fit the waveform points. However, the ultimate outcomes are not as good as LSTM and BiLSTM which demonstrates LSTM and BiLSTM have a distinct advantage for time series prediction. Table 1 displays the running times of the three models. When LSTM and BiLSTM are compared in terms of running times, the BiLSTM consumes twice the time while producing results that are almost identical to those of the LSTM. Given the available computer power and a sufficient data collection, we choose LSTM.

III. MODEL AND METHOD

In this section, we present our model and method. Due to the superior performance of the LSTM model, we employ the LSTM model in our approach. And we propose the

utilization of multi-step prediction method for waveform prediction, which significantly enhances both computational speed and accuracy of waveform prediction.

A. LSTM

LSTM is a recurrent neural network. Recurrent neural network is a recurrent network structure that updates the current state based on the current input and past state, which has a more favorable result than CNN and DNN in processing time series data. LSTM is a network structure that solves the long-term time-dependent problem of recurrent neural networks. It is well-known for deep learning research and excels in a variety of areas, including text and language connected to time series prediction [24].

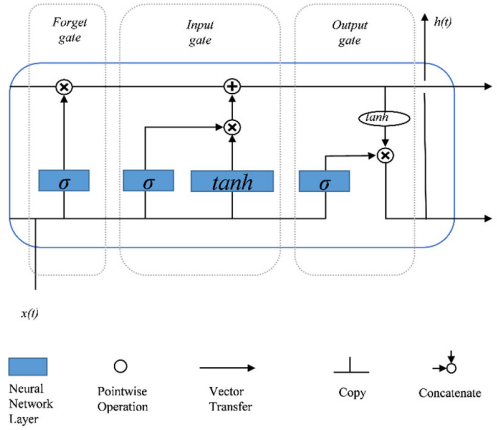

Figure 2. Schematic structure of LSTM.

Figure 2 depicts the LSTM's schematic structure [25]. A forget gate, an input gate, and an output gate are the three components that make up an LSTM neuron. The LSTM network's most distinctive structure, the forget gate, decides whether to keep and discard unit state information. The related mathematical expressions are as follows:

$$f_t = \sigma\left(W_{fh}h_{t-1} + W_{fx}x_t + b_f\right) \quad (1)$$

where W_{fh}, W_{fx} and the subsequent W_{ih}, W_{ix}, $W_{\tilde{c}h}$ and $W_{\tilde{c}x}$ are weights. b_f, b_i,, and $b_{\tilde{c}}$, are the biases. When the value of f_t, of the forgetting gate is set to 1, it retains all the information. On the other hand, when f_t is set to 0, all data is deleted.

The input gate selects the data to be deposited in the new cell as well as the input to the network. The following are the mathematical expressions for the relevant parameters:

$$i_t = \sigma(W_{ih}h_{t-1} + W_{ix}x_t + b_i) \quad (2)$$
$$\tilde{c}_t = \tanh(W_{\tilde{c}h}h_{t-1} + W_{\tilde{c}x}x_t + b_{\tilde{c}}) \quad (3)$$

where i_t represents the input gate's state. It can decide what data can be kept in the cell.

The output gate uses the current state for the output. The mathematical expressions which we compute the output and cell state are as follows:

$$c_t = f_t c_{t-1} + i_t \tilde{c}_t \quad (4)$$
$$o_t = \sigma(W_{oh}h_{t-1} + W_{ox}x_t + b_o) \quad (5)$$
$$h_t = o_t \bullet \tanh(c_t) \quad (6)$$

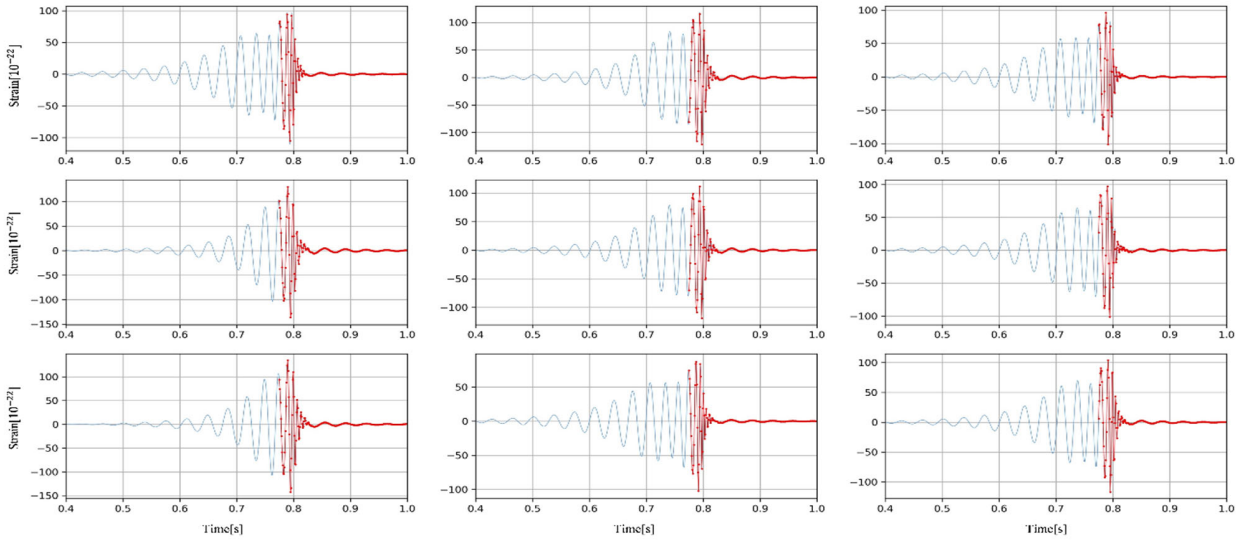

Figure 3. Waveform prediction comparison chart. The left side of the blue curve is input when there is no intersection part with the red one. The right side of the blue curve intersecting the red curve is the predicted target. The red part is the prediction part.

where h_t represents the output at moment t. c_t is the cell state of the LSTM at moment t.

B. Multi-step Prediction

Time series forecasting can be divided into single-step forecasting and multi-step forecasting depending on the

dimensionality of the output. The output dimension of single-step forecasting is one-dimensional and will be passed as input for forecasting the next step. The process continues until all forecasts are made. The error in the current prediction is then passed on to all subsequent points. And it goes without saying that these errors will mount.

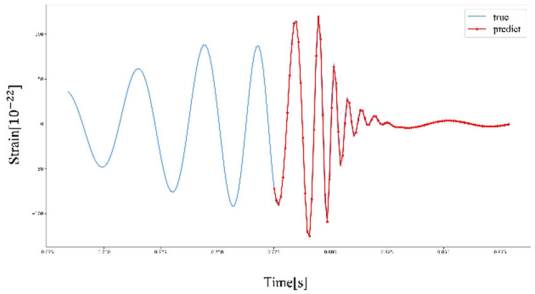

Figure 4. Waveform prediction enlargement chart.

Multi-step prediction, on the other hand, has an output dimension greater than one dimension. When handling time-dependent issues, it performs worse than single-step prediction. Multi-step prediction can be generally divided into multi-model multi-step prediction, multi-output, and seq2seq strategies [26]. We use multi-step prediction with multiple outputs. The remaining waveform is directly anticipated and output after a segment of the waveform sequence is entered. The neural network will adjust the number of neurons in the output layer according to the projected waveform's predicted number of points.

IV. RESULTS

This section presents the experimental results of our study. Here, we showcase the predictive performance of the models and their robustness under varying parameters.

A. *Demonstrate the Predictions of The Model*

Input known waveform segments and make predictions for the remaining segments using the taught deep learning model. We present nine waveforms from the test set at random in Figure 3, together with the input waveform, the predicted waveform, and the actual waveform. The subfigure of Figure 3 was expanded in Figure 4 to better display the intricacies. The predicted partial waveform sequence was compared to the actual waveform sequence, and the match rate was 99.91%.

The predicted matches for 100,000 waveforms were statistically done and are shown in the histogram of subplot (a.2) in Figure 5. The mean value of the match at the mass range of [30,70] $\boldsymbol{M}_{\odot}$, spin range of [-0.4,0.4] is 99.88% and the median is 99.95%.

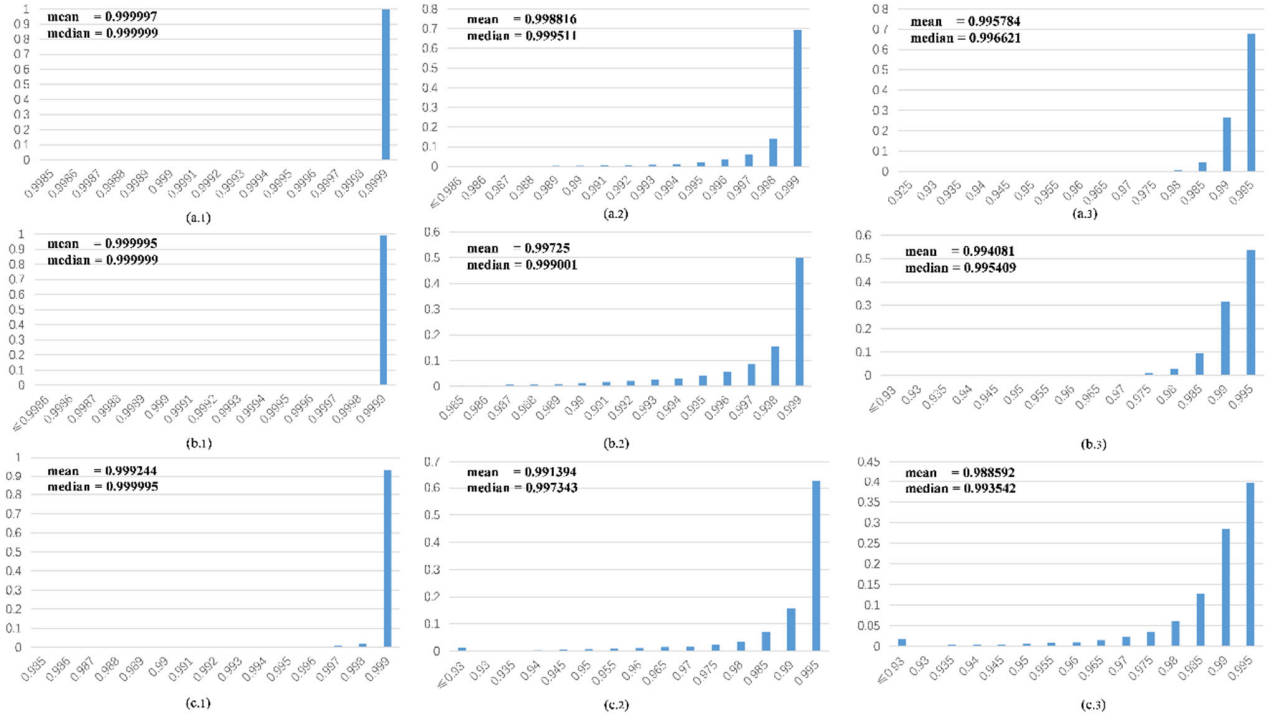

Figure 5. Model prediction matching statistics. a, b, and c represent the range of SPIN, which are $|s_{1,2}| \leq 0.4$, $|s_{1,2}| \leq 0.6$, and $|s_{1,2}| \leq 0.8$, respectively. 1, 2, and 3 represent models with 200, 230, and 260 predicted points

B. *Generalization Capability of The Model*

In this model, the three crucial variables are mass, spin, and sequence length. This section we will examine how well the model may be generalized in light of these three factors.

With three sequence lengths, we trained three models. They are 200 points, 230 points, and 260 points respectively.

Predictions were made for 100,000 waveforms using a uniform method of generating training and test datasets. The criteria used to generate the dataset when studying spin and mass later are the same. Statistics were made for each predicted waveform match, and the mean and median of the matches were calculated and the statistical results are shown in subplots a.1 and a.3 of Figure 5.

Expand the range of spin to generate the test dataset in $|s_{1,2}| \leq 0.6$ and $|s_{1,2}| \leq 0.8$. To investigate the predictive ability of the model in the range of the non-training set, three models with different sequence lengths in $|s_{1,2}| \leq 0.4$ were used to predict the test dataset in $|s_{1,2}| \leq 0.6$ and 0.8. The statistical results are displayed in Figure 5.

Broadening the sampling range of the BBH masses to [25, 75] $\boldsymbol{M}_{\odot}$ and [20, 80] $\boldsymbol{M}_{\odot}$. Fixed spins in $|s_{1,2}| \leq 0.4$ and sequence length of 230 points. We took the model to predict these 100,000 waveforms and matched them. The mean and median were obtained after tallying all of the matches and the outcomes are shown in Table 2.

TABLE II. MATCHING OF MODEL'S PREDICTIONS WITH WAVEFORMS OF DIFFERENT MASSES

Spin(s_1, s_2)	**Mass(m_1, m_2) / $M_{\odot}$**	**Mean**	**Medium**
[-0.4, 0.4]	[30, 70]	0.9988	0.9995
[-0.4, 0.4]	[25, 75]	0.9973	0.9990
[-0.4, 0.4]	[20, 80]	0.9934	0.9980

The statistics in Figure 5 and Table II indicate that the model can achieve more than 99% matching when varying sequence length, spin, and mass. In predicting 200 points, the model was able to generate waveforms with almost 100% overlap.

V. CONCLUSION AND FUTURE WORK

The proposed method and model aim to combine deep learning and find suitable methods to improve the accuracy and computational efficiency of the GW construction. It can provide suggestions for the construction of GW template library and provide methods for problems that may be encountered in the GW detection process. We observed many time domain waveforms of BBH mergers and found predictable trends in each phase of the waveform. The waveform consists of many segments stitched together and the segments are interconnected. Predicting the remaining waveform from a known segment of the waveform is precisely the principle of multi-step prediction. We proposed using a multi-step prediction method to learn and predict the time-domain waveform of a BBH merger with aligned spins. We tested the prediction effectiveness of the method at 200, 230, and 260 points. And the predicted segments under the same template matched well with the waveform of the target.

We also showed the generalization ability of the model in terms of mass and spin distribution. Model predictions were compared for spins($|s_{1,2}|$) $\leq$ 0.4, 0.6, and 0.8, and for mass($m_{1,2}$) at [30,70] $\boldsymbol{M}_{\odot}$, [25,75] $\boldsymbol{M}_{\odot}$ and [20,80] $\boldsymbol{M}_{\odot}$. The statistical results show that the model still performs well in prediction with 2 times the spin range and 1.5 times the mass range of the training data set. The match between the predicted value and the waveform can exceed 99%.

The GW of the BBH merger is a simple and traceable waveform in the time domain. It is suitable for forecasting using a multi-step forecasting method. Our model needs to specify the number of input and output points. The computational efficiency and prediction of the model are very high due to the absence of cumulative errors and double counting. The model can be subsequently improved by scaling down the number of input points and increasing the number of prediction points. It is also possible to train multiple models to cover various stages and combine the predictions to obtain the complete waveform.

REFERENCES

[1] B. P. Abbott, et al. "Observation of gravitational waves from a binary black hole merger," Phys. Rev. Lett, vol. 116, p. 061102, Feb 2016. [Online]. Available from: https://link.aps.org/doi/10.110

[2] Y. Zhang, et al. "Deep learning model based on a bidirectional gated recurrent unit for the detection of gravitational wave signals," Physical Review D, vol. 106, no. 12, p. 122002, 2022.

[3] B. Beheshtipour and M. A. Papa, "Deep learning for clustering of continuous gravitational wave candidates," Phys. Rev. D, vol. 101, p. 064009, Mar 2020. [Online]. Available from: https://link.aps.org/doi/10.1103/PhysRevD.101.0640-09

[4] M. B. Schäfer, A. H. Nitz, "From one to many: A deep learning coincident gravitational-wave search", Physical Review D, 2022, 105(4): 043003.

[5] S. R. Green and J. Gair, "Complete parameter inference for GW150914 using deep learning," Machine Learning: Science and Technology, vol. 2, no. 3, p. 03LT01, jun 2021. [Online]. Available from: https://doi.org/10.1088/26322153/abfaed

[6] M. Dax, S. R. Green, J. Gair, J. H. Macke, A. Buonanno, and B. Schölkopf, "Real-time gravitational wave science with neural posterior estimation," Phys. Rev. Lett., vol. 127, p. 241103, Dec 2021. [Online]. Available from: https://link.aps.org/doi/10.110

[7] H. Shen, D. George, E. A. Huerta, and Z. Zhao, "Denoising gravitational waves with enhanced deep recurrent denoising auto-encoders," in ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019, pp. 3237–3241.

[8] W. Wei and E. Huerta, "Gravitational wave denoising of binary black hole mergers with deep learning," Physics Letters B, vol. 800, p. 135081, 2020. [Online]. Available from: https://www.sciencedire-ct.com/science/article/pii/S0370269319308032

[9] C. García-Quirós, et al. "Multimode frequency-domain model for the gravitational wave signal from nonprecessing black-hole binaries," Phys. Rev. D, vol. 102, p. 064002, Sep 2020. [Online]. Available from: https://link.aps.org/doi/10.1103/

[10] N. E. M. Rifat, S. E. Field, G. Khanna, and V. Varma, "Surrogate model for gravitational wave signals from comparable and large-mass-ratio black hole binaries," Phys.

Rev. D, vol. 101, p. 081502, Apr 2020. [Online]. Available from: https://link.aps.org/doi/10.1103/PhysRevD.101.081502

[11] A. J. K. Chua, C. R. Galley, and M. Vallisneri, "Reduced-order modeling with artificial neurons for gravitational-wave inference," Phys. Rev. Lett., vol. 122, p. 211101, May 2019. [Online]. Available from: https://link.aps.org/doi/10.1103-/PhysRevLett.122.211101

[12] S. Khan and R. Green, "Gravitational-wave surrogate models powered by artificial neural networks," Phys. Rev. D, vol. 103, p. 064015, Mar 2021. [Online]. Available from: https://link.-aps.org/doi/10.1103/P-hysRevD.103.064015

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Commun. ACM, vol. 60, no. 6, p. 84–90, may 2017. [Online]. Available from: https://doi.org/10.1145/3065386

[14] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," Science, vol. 313, no. 5786, pp. 504–507, 2006. [Online]. Available from: https://www.science.org/doi/-abs/10.1126/science.1127647

[15] W. Zaremba, I. Sutskever, and O. Vinyals, "Recurrent neural network regularization," CoRR, vol. abs/1409.2329, 2014. [Online]. Available from: http://arxiv.org/abs/1409.2329

[16] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[17] A. Vaswani, et al. "Attention is all you need," in Advances in Neural Information Processing Systems, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017. [Online]. Available from: https://proceedings.neurips.cc-/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aaPaper.pdf

[18] I. Goodfellow, et al. "Generative adversarial networks," Commun. ACM, vol. 63, no. 11, p. 139–144, oct 2020. [Online]. Available from: https://doi.org/10.1145/3422622

[19] J. McGinn, C. Messenger, M. J. Williams, and I. S. Heng, "Generalised gravitational wave burst generation with generative adversarial networks," Classical and Quantum Gravity, vol. 38, no. 15, p. 155005, Jun 2021. [Online]. Available from: https://doi.org/10.1088/13616-382/ac09cc

[20] A. Khan, E. A. Huerta, and H. Zheng, "Interpretable ai forecasting for numerical relativity waveforms of quasicircular, spinning, nonprecessing binary black hole mergers," Physical Review D, vol. 105, no. 2, 1 2022. [Online]. Available from: https://www.osti.gov/biblio/1840697

[21] S. Khan, et al. "Frequency-domain gravitational waves from nonprecessing black-hole binaries. ii. a phenomenological model for the advanced detector era," Phys. Rev. D, vol. 93, p. 044007, Feb 2016. [Online]. Available from: https://-link.aps.org/doi/10.1103/PhysRevD.93.044007

[22] LIGO Scientific Collaboration, "LIGO Algorithm Library - LALSuite," free software (GPL), 2018.

[23] J. McGinn, C. Messenger, M. J. Williams, and I. S. Heng, "Generalised gravitational wave burst generation with generative adversarial networks," Classical and Quantum Gravity, vol. 38, no. 15, p. 155005, Jun 2021. [Online]. Available from: https://doi.org/10.1088/13616-382/ac09cc

[24] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[25] Y. Yu, X. Si, C. Hu, and J. Zhang, "A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures," Neural Computation, vol. 31, no. 7, pp. 1235–1270, 07 2019. [Online]. Available from: https://doi.org/10.1162/neco_-a_01199

[26] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in Advances in Neural Information Processing Systems, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Wein- berger, Eds., vol. 27. Curran Associates, Inc., 2014. [Online]. Available from: https://proceedings.neurips.cc/paper/2014/file/a14ac55a4f2-7472c5d894ec1c3c743d2Paper.pdf