

Identification of stochastic gravitational wave backgrounds from cosmic strings using machine learning

Xianghe Ma^{1,2}, Borui Wang³, Nan Yang⁴, Jin Li^{1,2,*}, Brendan McCane⁵, Mengfei Sun,^{1,2} Jie Wu^{1,2}, Minghui Zhang,⁶ and Yan Meng^{7,†}

¹*Department of Physics, Chongqing University, Chongqing 401331, People's Republic of China*

²*Chongqing Key Laboratory for Strongly Coupled Physics, Chongqing University, Chongqing 401331, People's Republic of China*

³*Department of Earth and Sciences, Southern University of Science and Technology, Shenzhen 518055, People's Republic of China*

⁴*Department of Electronical Information Science and Technology, Xingtai Key Laboratory for Research and Application of Robot Intelligent Detection and Sorting Technology, Xingtai University, Xingtai 054001, China*

⁵*School of Computing, University of Otago, Otago 9016, New Zealand*

⁶*Department of Physics, Southern University of Science and Technology, Shenzhen 518055, People's Republic of China*

⁷*Department of Physics, Xingtai Key Laboratory for Research and Application of Robot Intelligent Detection and Sorting Technology, Xingtai University, Xingtai 054001, China*



(Received 9 March 2025; accepted 6 August 2025; published 26 September 2025)

Cosmic strings play a crucial role in enhancing our understanding of the fundamental structure and evolution of the Universe, unifying our knowledge of cosmology, and potentially unveiling new physical laws and phenomena. The advent and operation of space-based detectors provide an important opportunity for detecting stochastic gravitational wave backgrounds (SGWBs) generated by cosmic strings. However, the intricate nature of an SGWB poses a formidable challenge in distinguishing its signal from the complex noise and other SGWB sources by some traditional methods. Therefore, we attempt to identify SGWBs based on machine learning. Our findings show that the joint detection of Linear Symbols to Attention and Taiji significantly outperforms individual detectors, and even in the presence of numerous low SNR signals. The true positive rate remains exceptionally high with 95%. Our study demonstrates that the multiband joint analysis method significantly enhances the discernment capability for SGWBs from different origins, providing a novel technical approach to disentangle the components in composite SGWB signals. Although our discussion is based solely on simulated data, the relevant methods can provide data-driven analytical capabilities for future observations of SGWBs.

DOI: 10.1103/jvlp-kvpk

I. INTRODUCTION

The Laser Interferometer Gravitational Wave Observatory (LIGO)/Virgo Collaboration made a groundbreaking announcement with the first direct detection of gravitational waves, produced by the merger of binary black holes (BBHs) [1]. This discovery not only confirmed Einstein's general theory of relativity but also marked the dawn of gravitational wave astronomy. With continuous advancements in detector sensitivity, the number of observed gravitational wave events has steadily increased. Since the first detection in 2015 (GW150914), ground-based detectors have successfully recorded over 100 events, primarily in the frequency range of tens to

hundreds of hertz. However, a significant gap still exists in the detection of low-frequency gravitational waves. To fill this gap, space-based gravitational wave detectors like Linear Symbols to Attention (LISA) [2], the Taiji program [3], and the TianQin project [4] are being developed, offering new opportunities for low-frequency gravitational wave detection.

Space-based detectors are not only capable of detecting a wide range of compact binary coalescence events [5], but they also serve as powerful tools for investigating stochastic gravitational wave backgrounds (SGWBs) in the millihertz frequency band. The SGWBs consist of a superposition of numerous independent gravitational wave signals from various directions in the sky, originating from a variety of sources across the Universe. These sources are classified into cosmological and astrophysical categories, spanning different epochs in the history of the Universe.

*Contact author: cqujinli1983@cqu.edu.cn

†Contact author: 200221007@xttc.edu.cn

The cosmological sources of the SGWBs include primordial processes in the early Universe, such as inflation [6,7], cosmic strings [8–11], and phase transitions [12–15]. This paper primarily focuses on the SGWBs generated by cosmic strings. Meanwhile, we also discuss the performance of our machine learning model in identifying these different sources of SGWBs. The study of SGWBs is crucial, as it provides unique insights into the early Universe, fundamental physics, and astrophysical populations and processes. Detecting and characterizing the SGWBs can help validate inflation models, explore the physics of the early Universe, and illuminate the population and evolution of compact objects over cosmic time [16–18].

Due to the inherent noise characteristics of gravitational wave detectors [19,20], particularly the prevalent foreground noise in space-based detectors, identifying gravitational wave signals from the data is a significant challenge. Traditional signal identification methods, such as Bayesian inference [21,22] and matched filtering [23,24] have limitations when applied to the detection of SGWBs. For example, Bayesian inference requires sufficient prior information, which is difficult to obtain for SGWBs, and it involves considerable offline processing time. Matched filtering relies on correlating the received signal with known target templates; however, the stochastic gravitational wave background lacks defined waveform templates. As an alternative, we employ machine learning techniques to identify SGWBs, offering a promising approach that complements traditional methods.

Machine learning has made significant progress in various gravitational wave data analysis tasks, including GW detection [25–28], parameter estimation [29,30], glitch classification [31–33], noise reduction [34,35], and signal extraction [36–39]. In Ref. [40], Zhoujian Cao *et al.* applied machine learning techniques to identify SGWBs within a general model and achieved promising results. In contrast, this study focuses on the SGWBs generated by cosmic strings, considering a broader SNR range than the work in Ref. [40], and utilizes multidetector joint observations with foreground noise.

This paper is organized as follows: Sec. II describes the data simulation process, including the simulation of SGWB signals, the sensitivity curves for LISA and Taiji, and the foreground noise. Section III provides a brief introduction to the machine learning network used in this study and discusses some key hyperparameters. Section IV presents the results and analysis, including outcomes with detector noise and foreground noise. Finally, Sec. V concludes the paper and offers some remarks.

II. DATA SIMULATION

Data simulation primarily comprises two components: signal simulation and noise simulation. The following

subsections discuss the principles of these simulations respectively, while detailed data composition and structure are provided in Appendix A.

A. The stochastic gravitational wave background signals generated by cosmic strings

We adopt the model proposed in Refs. [41–43] to describe the gravitational waves generated by cosmic strings, employing highly accurate analytical approximation formulas for their representation [44,45]. The gravitational wave signals are influenced significantly by the cosmic string tension $G\mu$, which characterizes the size of the loops. The value of $G\mu$ ranges from 10^{-17} to 10^{-15} , and we adopt a uniform distribution for sampling within each order of magnitude. We introduce a free parameter α to denote the string loop size and define the total power emitted by the cosmic strings as $\Gamma = 50$. The value of α ranges from 10^{-3} to 10^0 , which are also sampled with a uniform distribution.

For the SGWBs from cosmic strings, the contributions from cosmic string loops can be categorized into three periods:

- (i) loops formed and decayed during the radiation era,
- (ii) loops formed during the radiation era and decayed during the matter era, and
- (iii) loops formed during the matter era.

For loops formed and decayed in the radiation region, the form of stochastic gravitational wave background is given by

$$\Omega_{\text{GW}}^r(f) = \frac{128}{9} \pi A_r \Omega_r \frac{G\mu}{\varepsilon_r} \left[\left(\frac{f(1 + \varepsilon_r)}{\frac{B_r \Omega_m}{\Omega_r} + f} \right)^{\frac{3}{2}} - 1 \right], \quad (1)$$

where

$$\varepsilon_r = \frac{\alpha}{\Gamma G\mu}, \quad A_r = \frac{\tilde{c}}{\sqrt{2}} F \frac{v_r}{\xi_r^3}, \quad B_r = \frac{2H_0 \Omega_r^{\frac{1}{2}}}{\nu_r \Gamma G\mu}. \quad (2)$$

The label r indicates the radiation era and in these equations, $v_r = 0.662$, $\xi_r = 0.271$, $\nu_r = 1/2$, $F = 0.1$ and \tilde{c} is a phenomenological parameter which can be set as $\tilde{c} = 0.23 \pm 0.04$ [46]. In our work, the evolution of the Universe is assumed to follow a standard Λ cold dark matter (Λ CDM) model, with its underlying parameters are $H_0 = 100h \text{ km}/(\text{s} \cdot \text{Mpc})$, $h = 0.678$, $\Omega_r = 8.397 \times 10^{-5}$, and $\Omega_m = 0.308$. For loops formed in the radiation region and decayed in the matter region, their contribution to SGWBs has the following form:

$$\Omega_{\text{GW}}^{\text{rm}}(f) = 32\sqrt{3}\pi(\Omega_m\Omega_r)^{\frac{3}{4}}H_0\frac{A_r(1+\epsilon_r)^{\frac{3}{2}}}{\Gamma f^{\frac{1}{2}}\epsilon_r} \times \left\{ \frac{\left(\frac{\Omega_m}{\Omega_r}\right)^{\frac{1}{4}}}{\left(B_m\left(\frac{\Omega_m}{\Omega_r}\right)^{\frac{1}{2}}+f\right)^{\frac{1}{2}}} \left[2 + \frac{f}{B_m\left(\frac{\Omega_m}{\Omega_r}\right)^{\frac{1}{2}}+f} \right] - \frac{1}{(B_m+f)^{\frac{1}{2}}} \left[2 + \frac{f}{B_m+f} \right] \right\}, \quad (3)$$

where

$$B_r = \frac{2H_0\Omega_m^{\frac{1}{2}}}{\nu_m\Gamma G\mu}, \quad (4)$$

where $\nu_m = 2/3$. The contribution of loops generated in the matter period to the SGWB generation by cosmic strings is given by

$$\Omega_{\text{GW}}^m(f) = 54\pi H_0\Omega_m^{\frac{3}{2}}\frac{A_m}{\Gamma}\frac{\epsilon_m+1}{\epsilon_m}\frac{B_m}{f} \times \left\{ \frac{2B_m+f}{B_m(B_m+f)} - \frac{1}{f}\frac{2\epsilon_m+1}{\epsilon_m(\epsilon_m+1)} + \frac{2}{f}\log\left(\frac{\epsilon_m+1}{\epsilon_m}\frac{B_m}{B_m+f}\right) \right\}, \quad (5)$$

where

$$\epsilon_m = \frac{\alpha}{\Gamma G\mu}, \quad A_m = \frac{\tilde{c}}{\sqrt{2}}F\frac{v_m}{\xi_m^3}. \quad (6)$$

The label m indicates the matter era and in these equations $v_m = 0.583$ and $\xi_m = 0.625$. Therefore, the SGWBs generated by cosmic strings can be well approximated as

$$\Omega_{\text{GW}}(f) = \Omega_{\text{GW}}^r(f) + \Omega_{\text{GW}}^{\text{rm}}(f) + \Omega_{\text{GW}}^m(f), \quad (7)$$

which can provide a good approximation for loops with $\alpha \geq \Gamma G\mu$.

B. Simulation of instrument noise

Leveraging the triangular geometry of LISA-like detectors, time-delay interferometry (TDI) techniques can be employed to combine phase differences with varying time delays, effectively canceling laser frequency noise [47,48]. For simplicity, we assume that the instrument noise consists of two primary components: test mass acceleration noise and optical path length fluctuations. These noise sources are considered identical for each spacecraft. Given that the arm lengths are equal, the LISA instrument effectively forms an equilateral triangle [49].

For computational convenience, we adopt approximations of the gravitational wave (GW) response in the A, E, and T channels, as provided in Ref. [50]:

$$R_A^i(f) = R_E^i(f) = \frac{9}{20}|W^i(f)|^2 \left[1 + \left(\frac{f}{4f_i/3} \right)^2 \right]^{-1}, \quad (8)$$

and

$$R_T^i(f) \simeq \frac{1}{4032} \left(\frac{f}{f_i} \right)^6 |W^i(f)|^2 \left[1 + \frac{5}{16128} \left(\frac{f}{f_i} \right)^8 \right]^{-1}, \quad (9)$$

where $i = \text{LISA, Taiji}$, $W^i(f) = 1 - e^{-2if/f_i}$, and for the LISA-like detector, $f_i = c/(2\pi L_i)$, with $L_{\text{LISA}} = 2.5 \times 10^6$ km, and $L_{\text{Taiji}} = 3 \times 10^6$ km.

We adopt the noise model outlined in the LISA science requirements document [19,51] and assume that the same model applies to the LISA-like detector, Taiji. This model presumes that the noise in all channels remains constant and identical [48], specifically accounting for two primary noise sources: acceleration noise and optical path disturbance noise, which are explicitly defined as follows [2,52]:

$$N_{\text{acc}}^i(f) = \frac{N_a^i}{(2\pi f)^4} \left(1 + \left(\frac{f_1}{f} \right)^2 \right) = \frac{(\sqrt{(\delta a_i)^2}/L_i)}{(2\pi f)^4} \left(1 + \left(\frac{f_1}{f} \right)^2 \right), \quad (10)$$

$$N_{\text{op}}^i(f) = N_o^i = \left(\sqrt{(\delta x_i)^2}/L_i \right)^2, \quad (11)$$

where

$$\sqrt{(\delta a_{\text{LISA}})^2} = 3 \times 10^{-15} \text{ m/s}^2, \quad (12)$$

$$\sqrt{(\delta x_{\text{LISA}})^2} = 1.5 \times 10^{-11} \text{ m}, \quad (13)$$

and [53]

$$\sqrt{(\delta a_{\text{Taiji}})^2} = 3 \times 10^{-15} \text{ m/s}^2, \quad (14)$$

$$\sqrt{(\delta x_{\text{Taiji}})^2} = 8 \times 10^{-12} \text{ m}, \quad (15)$$

where $i = \text{LISA, Taiji}$, $f_1 = 0.4$ mHz. These noise models can be transformed into interferometer noise through

$$N_X^i(f) = [4N_{\text{op}}^i(f) + 8[1 + \cos^2(f/f_i)]N_{\text{acc}}^i(f)]|W^i(f)|^2, \quad (16)$$

$$N_{XY}^i(f) = -[2N_{\text{op}}^i(f) + 8N_{\text{acc}}^i(f)]\cos(f/f_i)|W^i(f)|^2. \quad (17)$$

The noise models for the power spectral densities in the A, E, T channels are obtained by diagonalizing the covariance matrix of the X,Y,Z channels. The resulting diagonal entries are given by

$$N_A^i(f) = N_E^i(f) = N_X^i(f) - N_{XY}^i(f), \quad (18)$$

$$N_T^i(f) = N_X^i(f) + 2N_{XY}^i(f). \quad (19)$$

The noise spectral density for different channels can be derived from the noise power spectral density (PSD) $S(f)$ and the corresponding response function, $R(f)$ satisfies

$$S_j^i(f) = \frac{N_j^i(f)}{R_j^i(f)}, \quad (20)$$

where $j = A, E, T$. Based on the noise spectral density, the total equivalent energy density for a single LISA-like detector can be formulated as follows [50]:

$$\Omega^i = \frac{4\pi^2 f^3}{3H_0^2} \left(\sum_{j=A,E,T} \frac{1}{S_j^i(f)} \right)^{-1}. \quad (21)$$

In order to demonstrate the detection capability of a gravitational wave detector, a power-law integrated sensitivity (PLS) is proposed [54]. Based on a given observation time T_{ob} and SNR threshold ρ_m , the PLS of the detector is given by

$$\Omega_{\text{PLS}}(f) = \max_{\kappa} \left(\Omega_{\kappa} \left(\frac{f}{f_{\text{ref}}} \right)^{\kappa} \right), \quad (22)$$

$$\Omega_{\kappa} = \frac{\rho_m}{\sqrt{2T_{\text{ob}}}} \left(\int_{f_{\text{min}}}^{f_{\text{max}}} df \frac{\left(\frac{f}{f_{\text{ref}}} \right)^{2\kappa}}{\Omega^i(f)^2} \right)^{-\frac{1}{2}}, \quad (23)$$

where f_{ref} can be freely chosen without affecting the PLS result [54], and the index $\kappa \in [-8, 8]$. Based on previous studies of SNR for gravitational wave detectors [42], we assume $\rho_m = 10$ and $T_{\text{ob}} = 4$ years.

C. Simulation of foreground noise

In actual observations, the data consists not only of gravitational wave signals generated by cosmic strings but also of astrophysical foreground noise. This foreground noise primarily includes two components: double white dwarfs (DWDs) and inspiraling binary black holes/binary neutron stars (BNSs) based on the observations of LIGO and Virgo [55–57].

The gravitational wave model for double white dwarfs is a modulated signal based on the LISA orbital motion. Its energy spectral density can be approximated using the broken power-law model proposed in Refs. [47,56], which is given by

$$\Omega_{\text{DWD}}(f) = \frac{A_1 \left(\frac{f}{f_i} \right)^{\alpha_1}}{1 + A_2 \left(\frac{f}{f_i} \right)^{\alpha_2}}, \quad (24)$$

where $A_1 = 7.44 \times 10^{-14}$, $A_2 = 2.96 \times 10^{-7}$, $\alpha_1 = -1.98$, $\alpha_2 = -2.6$.

For the superposition of gravitational wave background produced by inspiraling BBHs/BNSs observed by LIGO and Virgo, it is given by

$$\Omega_{\text{astro}}(f) = \Omega_{\text{astro}} \left(\frac{f}{f_*} \right)^{\alpha_{\text{astro}}}, \quad (25)$$

where $f_* = 3$ mHz, $\Omega_{\text{astro}} = 4.44 \times 10^{-12}$, and $\alpha_{\text{astro}} = 2/3$.

D. SGWB signals from other sources

Beyond cosmic strings, the LISA and Taiji frequency bands are expected to host SGWB signals from other sources, such as inflation and first-order phase transitions. For ease of study, we use a double broken power law in first-order phase transitions and consider the power law (PL) signal in the inflation.

In the first-order phase transitions, the power spectrum takes the form [47]

$$\Omega_{\text{PT}}(f; \Omega_p, f_p, r_b, b) = \Omega_p M(f; f_p, r_b, b), \quad (26)$$

where Ω_p is the peak of the power spectrum, f_p is the frequency corresponding to Ω_p , and r_b describes the ratio between the two breaks in the spectrum. The parameter b defines the spectral slope between the two breaks. The spectral shape $M(f; f_p, r_b, b)$ is a double broken power law with a spectral slope 9 at low frequencies and 4 at high frequencies, as follows:

$$M(f; f_p, r_b, b) = \left(\frac{f}{f_p} \right)^9 \left(\frac{1 + r_b^4}{r_b^4 + (\frac{f}{f_p})^4} \right)^{(9-b)/4} \times \left(\frac{b+4}{b+4-m+m(\frac{f}{f_p})^2} \right)^{(b+4)/2}, \quad (27)$$

and

$$m = \frac{9r_b^4 + b}{r_b^4 + 1}, \quad (28)$$

where $\Omega_p = 2 \times 10^{-11}$, $f_p = 10^{-3}$, $r_b = 0.4$, and $b = 1$.

In the inflation, the power spectrum takes the form [58]

$$h^2 \Omega_{\text{PL}}(f) = h^2 \Omega_* \left(\frac{f}{f_*} \right)^{n_t}, \quad (29)$$

and

$$h^2 \Omega_* \simeq 1.5 \times 10^{-13} \frac{H_*^4 e^{4\pi\xi_*}}{M_{\text{pl}}^4 \xi_*^6}, \quad (30)$$

$$n_t \simeq -4\varepsilon_* + (4\pi\xi_* - 6)(\varepsilon_* - \eta_*), \quad (31)$$

where h is dimensionless Hubble constant, $f_* = 1$ mHz is pivot frequency, Ω_* parametrizes the amplitude of the

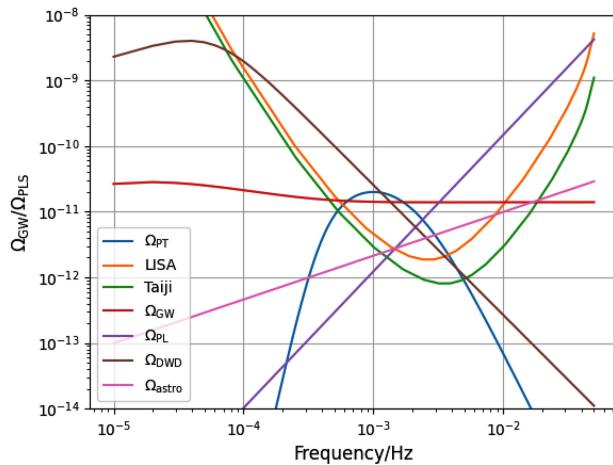


FIG. 1. Gravitational waves from different sources, foreground noise of gravitational wave, and PLS plots from different GW detectors for the case $\rho_m = 10$ and $T_{\text{ob}} = 4$ years. Ω_{GW} is the SGWB from the cosmic string with parameters $G\mu = 2 \times 10^{-14}$ and $\alpha = 0.1$.

spectrum at the frequency f_* , and n_t is the slope of spectrum. Our parameter values are chosen as $\xi_* = 4.474 \pm 0.003$, $\varepsilon_* = 0.0450 \pm 0.0003$, $H_*/M_{\text{pl}} = 10^{-5}$, and $\eta_* = 0$.

In Fig. 1, we present the relationship between the detector's sensitivity, foreground noise, and the equivalent energy density of gravitational waves from various sources. The equivalent energy density of the DWD signal is determined based on the arm length of the LISA detector.

The sensitivity curves of LISA and Taiji are quite similar; however, Taiji's longer arm length results in lower instrumental noise and superior detection capability. In the joint

observation of LISA and Taiji, SGWB signals produced by cosmic strings are assumed to be long-lasting. Therefore, it is unlikely that one detector would observe the signal while the other does not.

III. DEEP LEARNING

We employed a residual shrinkage network (ResNet) for this learning task, incorporating dropout and regularization to mitigate overfitting [59–61]. The ResNet is a deep learning model that combines residual networks with soft thresholding, aiming to enhance robustness against high-noise data. It is particularly well suited for processing time-domain signal data. Figure 2 illustrates the streamlined workflow of the network we utilized. The architecture of our network is characterized by the following key features: residual connections [62–64], soft thresholding [65–67], regularization [68,69], dropout [70,71], and multiscale feature extraction [72].

Each residual block includes a skip connection [73,74], which directly adds the input to the output. Skip connections help alleviate the vanishing gradient problem and enable the network to more effectively learn identity mappings. Soft thresholding is integrated into each residual block, where an adaptive threshold is computed based on the global average pooling of the mean absolute value of the input features. This operation effectively filters out noise, significantly enhancing signal recognition performance, particularly in low-SNR scenarios.

Our loss function is focal loss [75,76], a weighted cross-entropy loss function designed to enhance the model's

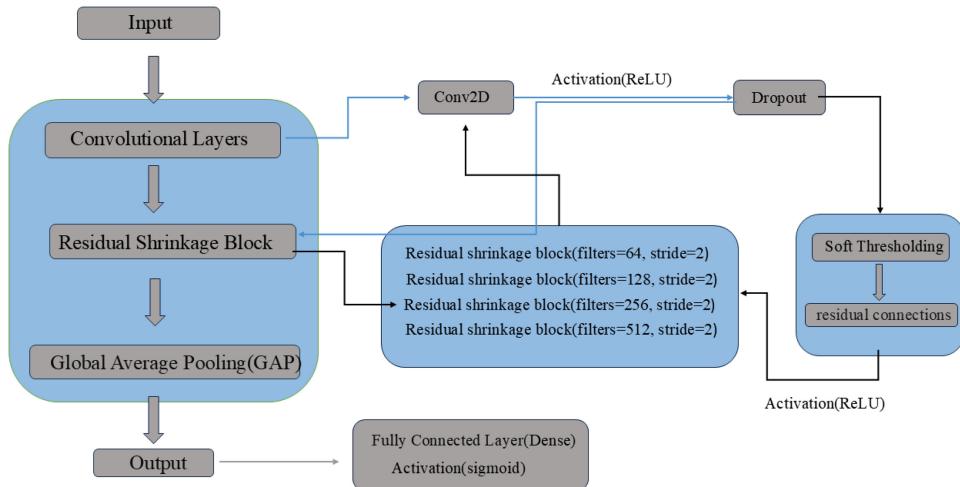


FIG. 2. Model architecture schematic representation of our machine learning model. We employed a ResNet architecture and incorporated Dropout and regularization at appropriate stages. An initial convolutional layer is added before the residual blocks to activate the input data and extract preliminary features. The residual block consists of four stacked residual shrinkage blocks, each containing two convolutional layers. Dropout is applied between the two convolutional layers, followed by soft-thresholding to suppress noise and highlight important features. A global average pooling layer is then used to compress the high-dimensional features into low-dimensional vectors, facilitating subsequent classification. Finally, a fully connected dense layer with a sigmoid activation function outputs the binary classification probabilities.

emphasis on samples that pose greater classification challenges, which is

$$L_{\text{Focal}} = -\frac{1}{N} \sum_{i=1}^N \alpha^{(i)} \cdot (1 - p_t^{(i)})^\gamma \cdot \log(p_t^{(i)}), \quad (32)$$

where, $\alpha^{(i)}$ represents the weighting coefficient for each sample, and $p_t^{(i)}$ denotes the predicted probability for the sample, expressed as follows:

$$\alpha^{(i)} = y_i \alpha^* + (1 - y_i)(1 - \alpha^*), \quad (33)$$

and

$$p_t^{(i)} = y_i \hat{y}_i + (1 - y_i)(1 - \hat{y}_i), \quad (34)$$

where, y_i represents the ground truth label, \hat{y}_i denotes the predicted probability for the positive class, α^* and γ are hyperparameters, and N indicates the total number of samples.

To mitigate the complexity of training and enhance overall efficiency, we employed a curriculum-based learning approach [77–79], which emphasizes a learning sequence from simple to complex. Specifically, this method involves adjusting the order of training samples based on their difficulty, starting with simpler or more easily understood samples and gradually introducing more complex ones. This approach helps the model better learn intricate patterns, thereby improving both training efficiency and performance.

In our study, the dataset consists of 200,000 samples, with an equal number of positive and negative samples. The positive samples are sorted by SNR in descending order. Starting from the highest SNR, every 20,000 positive samples, along with an equal number of negative samples, are selected together to form a subset, resulting in five subsets. Each subset thus contains 20,000 positive samples and 20,000 negative samples. For each subset, 20% of the samples are randomly selected as the validation set, while the remaining 80% are used for training.

Correspondingly, the training process is divided into five stages. In the first stage, only the subset with the highest SNR is used for training. In the second stage, the subset with a slightly lower SNR is added to the training set. This progression continues through the third and fourth stages. Finally, in the fifth stage, the entire training dataset is used for training [78,79]. This progressive training approach allows the model to initially learn fundamental features from high-SNR samples, gradually adapting to the complexities introduced by low-SNR samples, thereby enhancing overall learning performance. Table I presents some of the hyperparameters used in the network.

TABLE I. The hyperparameters of the network.

Optimizer	Adam
Learning rate	0.0003/0.0003/0.0003/0.0001/0.0001
Loss	Focal loss($\alpha^* = 0.25/0.25/0.2/0.2/0.2$)
Loss	Focal loss($\gamma = 1/2/2/2/2$)
Epochs	10/20/20/30/30
Batch size	64/64/64/32/32

IV. RESULTS AND ANALYSIS

An early stopping strategy was employed during the training process, which involved terminating the current phase of training and moving to the next phase when the loss showed minimal change over five consecutive epochs. In the first phase, the loss for all three models decreased rapidly to a low value. As each new phase introduced additional samples with lower SNR for training, a noticeable increase in loss occurred at the beginning of each phase. The curriculum learning strategy, which leverages high-SNR samples to establish a robust initial framework, ensures that the loss variations in subsequent phases remain moderate.

We analyzed the performance of individual detectors and joint detection in identifying the stochastic gravitational wave background separately. The results show that joint detection outperforms individual detectors, with Taiji achieving better results than LISA. When foreground noise is included, the detection accuracy decreases for both individual detectors and joint detection. However, joint detection still demonstrates superior performance compared to individual detectors. Additionally, in the cases with foreground noise, we analyzed how detection accuracy varies with SNR and identification between SGWBs from different sources.

A. Machine learning performance with instrumental noise

A receiver operating characteristics (ROC) graph is a technique for visualizing, organizing, and selecting classifiers based on their performance. ROC graphs have long been used in signal detection theory to depict the tradeoff between hit rates and false alarm rates of classifiers. An ROC curve is a two-dimensional depiction of classifier performance. To compare classifiers we may want to reduce ROC performance to a single scalar value representing expected performance. A common method is to calculate the area under the ROC curve, abbreviated AUC [80,81].

The ROC curve and confusion matrix are commonly used to evaluate the performance of machine learning models. In this section, we focus solely on the data containing the SGWB and instrumental noise. As shown in Fig. 3, the ROC curve for the LISA detector has the smallest area, with an AUC value of 0.8854. In contrast, machine learning with the Taiji detector and the joint detection of both detectors

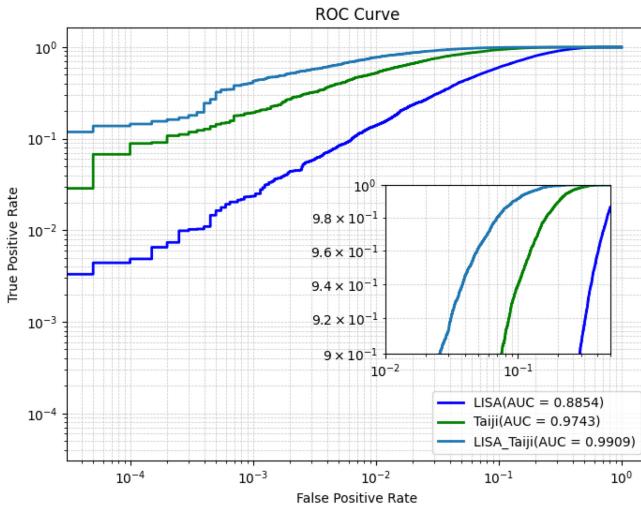


FIG. 3. ROC curves for LISA, Taiji, and the joint detection of both detectors under instrument noise.

achieved AUC values exceeding 0.97, demonstrating excellent detection performance.

As shown in Fig. 4, most of the misclassifications in LISA involve misidentifying noise-only samples as signal-only samples, which results in a false alarm rate (FAR) of 33.56% for LISA. Taiji and joint detection are much better in this regard, with 7.36% and 4.5%, respectively. We use detection rate, another term for true positive rate (TPR), to measure the probability of correctly recognizing a signal. Against the TPR, the combined detection is also much better than the single detector with 95.40%. It is clear that joint detection outperforms the individual detectors in SGWB identification.

As shown in Fig. 5, for LISA, this loss is always at a high level; for Taiji, this increase becomes evident in the thirdly phase; and for joint detection, the increase is only noticeable in the fourth phase. Furthermore, the final loss for joint detection is considerably lower than that for the individual detectors.

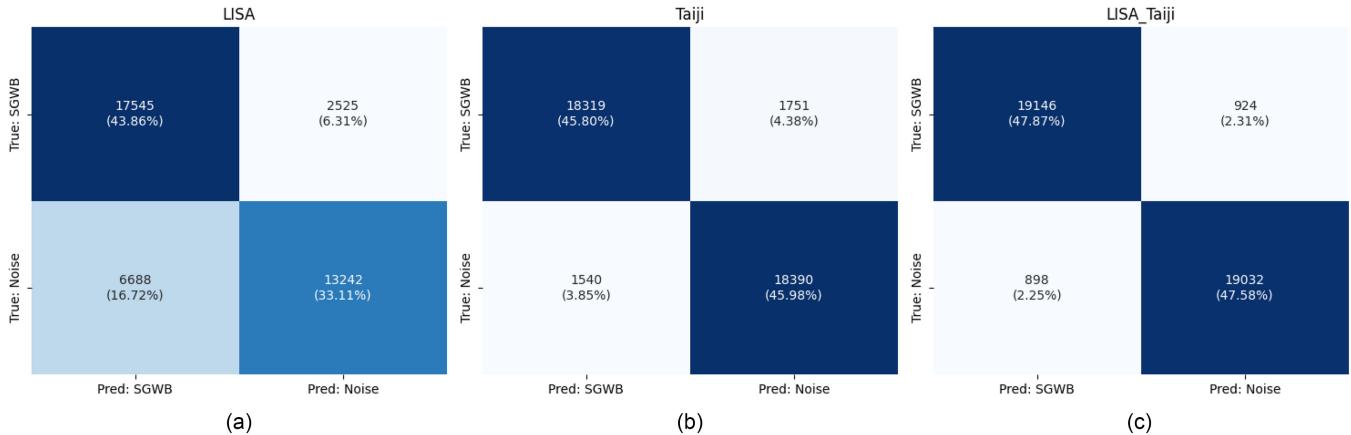


FIG. 4. Confusion matrices to the results of LISA, Taiji, and joint detection, with TPR of 87.42%, 91.69%, and 95.40%, and FAR of 33.56%, 7.36%, and 4.50%, respectively.

By analyzing the ROC curves, confusion matrices, and loss reduction plots, we conclude that joint detection outperforms individual detectors in identifying and detecting SGWB signals. Consequently, joint detection is likely to be a more effective and feasible approach for SGWB signal detection.

B. Machine learning performance with foreground noise

In our previous analysis, we concluded that joint detection provides higher accuracy compared to using a single detector. Therefore, the analysis in this section is based on joint detection. We further investigated the impact of SNR ranges on machine learning performance, considering the DWD and inspiraling BBH/BNS foreground noise based on the observations of LIGO and Virgo. Specifically, we use several datasets with different SNR ranges: 200,000 samples (SNR range: 20 to 76), 180,000 samples (SNR range: 22 to 76), 160,000 samples (SNR range: 24 to 76), 140,000 samples (SNR range: 27 to 76), and 120,000 samples (SNR range: 30 to 76). Since the SNR for the same positive sample can be correspondingly obtained in both the Taiji and LISA detectors, we refer to the SNR values under the LISA detector for joint detection throughout this analysis, in order to avoid any confusion.

Figure 6 underscores the importance of selecting appropriate SNR ranges for evaluating model performance, and shows how AUC values vary across different SNR intervals. With an increase in the lower bound of the SNR, the AUC of the ROC curve exhibits a monotonically increasing trend, demonstrating that the model achieves stronger generalization and classification performance under higher SNR conditions. Notably, within the SNR range of 30 to 76, the AUC values consistently exceed 0.99, indicating superior model performance in this range. Figure 7 compares the recognition performance across different SNR intervals. The model exhibits higher TPR and lower FAR in the 30–76 SNR range, indicating better detection performance.

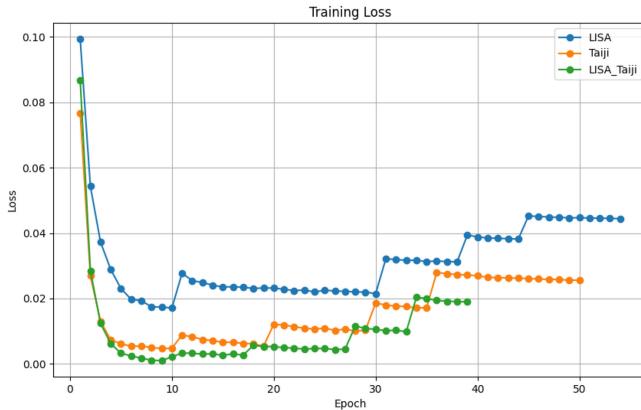


FIG. 5. Loss variations with epochs for the LISA, Taiji, and joint detection models. At lower SNRs, the loss for joint detection is consistently smaller than that of the individual detectors. The significant increase in loss at the beginning of a new phase occurs due to adding the lower SNR samples.

However, this does not imply that all samples with SNRs greater than 30 can be classified correctly. While high-SNR samples provide a solid foundation for feature learning and classification, they may still be misclassified due to other factors, such as noise complexity or subtle variations in signal characteristics. Furthermore, samples with SNRs below 30 pose a greater challenge to the model, as their signal features become more indistinguishable from noise, resulting in a higher misclassification rate.

These results indicate that within the SNR range of 30–76, even with the presence of foreground noise, machine learning with joint detection can still effectively identify SGWB signals in the data. As shown in Fig. 8, after adding the two types of foreground noise, only a small fraction of the samples remain unrecognized. When the SNR is between 30 and 34, the model achieves a classification accuracy of 77.99%. This accuracy increases to

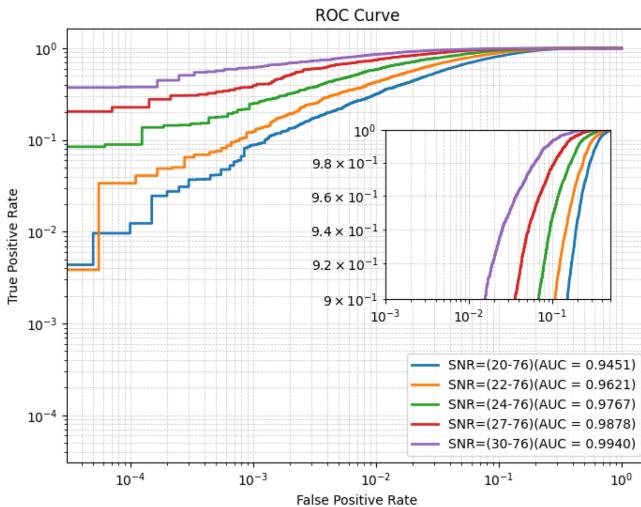


FIG. 6. ROC curves for different SNRs.

92.02% for SNRs between 34 and 38, and further improves to 98.27% for SNRs between 38 and 42. Notably, the model achieves 100% accuracy for all samples with SNRs between 42 and 76. Therefore, to keep the figure visually clear and uncluttered, Fig. 8 only presents results for the SNR range between 30 and 42. For parameter ranges where $G\mu \geq 10^{-16}$ and $\alpha \geq 0.1$, the samples can be correctly classified by our machine learning model. This suggests that SGWB signals from cosmic strings with these parameter values may be detectable using machine learning in future space-based gravitational wave joint detection efforts.

With this SNR range (30–76), we evaluated the performance of machine learning with three types of foreground noise: only DWD noise, only inspiral BBH/BNS noise based on LIGO and Virgo observations, and a mixture of both noise types. The total number of samples was reduced from 200,000 to 120,000, with positive and negative samples equally distributed. Figures 9 and 10 show that the AUC of the ROC curves for all three cases exceeds 0.99. Only a single type of foreground noise interference, all models consistently maintain a TPR above 98% and restrict the FAR to approximately 1%, fully demonstrating their excellent noise resistance and detection reliability.

In summary, the results demonstrate that restricting the SNR range significantly enhances the model's overall performance. This underscores the critical role of high-SNR samples in improving model accuracy, while highlighting the necessity of future efforts to refine algorithms and models to better handle low-SNR samples in complex signal-noise environments.

C. SGWB identification from different sources

Building upon previous findings regarding the SNR influence mechanism and the established optimal parameter range, this study employs cosmic-string SGWBs with $G\mu = 2 \times 10^{-14}$ and $\alpha = 0.1$ as reference samples mentioned in Appendix A, with particular focus on the binary classification between SGWBs from cosmic strings and those originating from first-order phase transitions or inflation. As illustrated in Fig. 1, the cosmic-string SGWB demonstrates remarkable spectral stability across the detector's sensitive frequency range. In contrast, SGWBs generated by either first-order phase transitions or inflation exhibit substantial spectral variations. Since they manifest distinct differences in different spectral ranges, we choose multiband analysis to identify them.

As shown in Fig. 11, the multiband analysis method enables effective identification of SGWB signals in both LISA and Taiji detectors. It should be noted that while LISA demonstrates limited performance in detecting SGWBs from first-order phase transitions, Taiji exhibits outstanding detection capabilities for SGWBs originating from both cosmic strings and first-order phase transitions. Although this study only utilizes data from two frequency

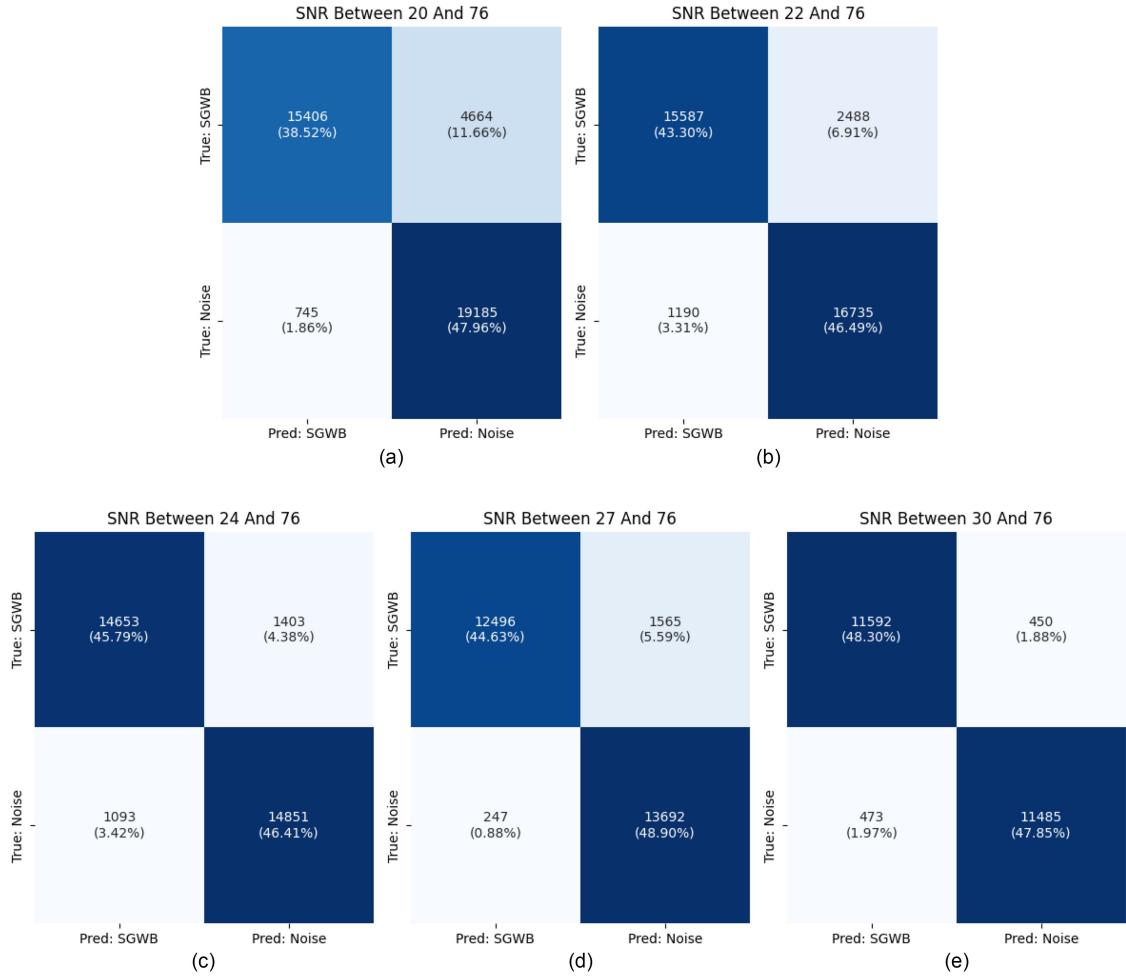


FIG. 7. Confusion matrix for different SNRs. (a) SNR between 20 and 76 with 76.76% TPR and 3.74% FAR, (b) SNR between 22 and 76 with 86.24% TPR and 6.64% FAR, (c) SNR between 24 and 76 with 91.26% TPR and 6.86% FAR, (d) SNR between 27 and 76 with 94.46% TPR and 5.50% FAR, (e) SNR between 30 and 76 with 96.26% TPR and 3.95% FAR.

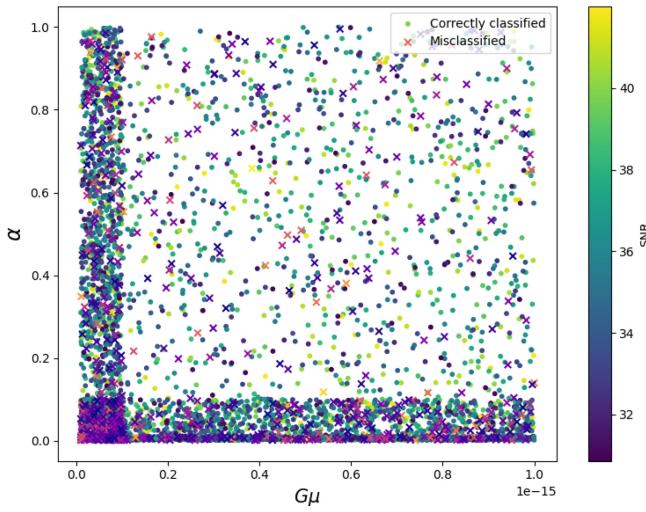


FIG. 8. Classification in the parameter space. Dots represent samples correctly classified by the model, forks represent misclassified samples, and color shades represent the magnitude of the SNR value, which corresponds to Fig. 7(e).

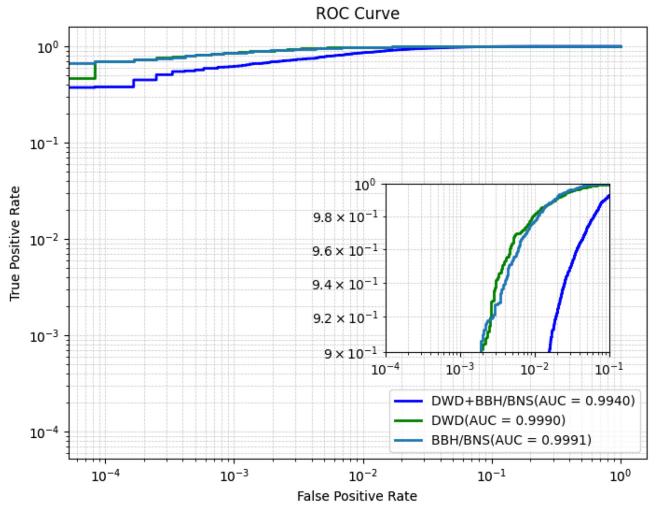


FIG. 9. ROC curves with different foreground noises in the joint detection. The analysis includes the simultaneous injection of the aforementioned two types of foreground noise, using samples with an SNR range of 30 to 76.

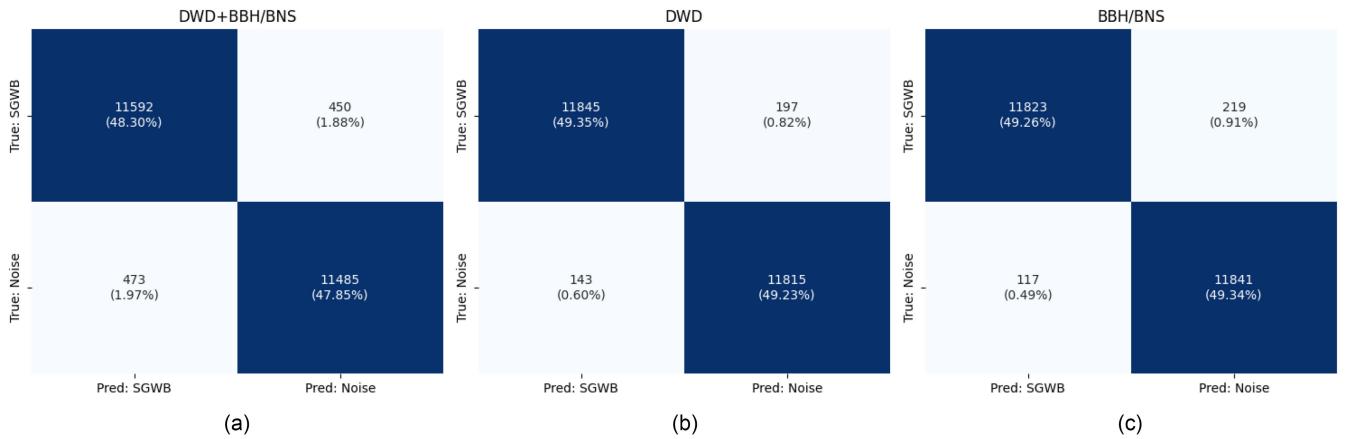


FIG. 10. Confusion matrix for different foreground noise. The sample selection is consistent with Fig. 11. (a) data with DWD foreground noise and inspiral BBH/BNS noise based on LIGO and Virgo observations, (b) data with DWD foreground noise, (c) data with inspiral BBH/BNS noise based on LIGO and Virgo observations.

bands, we are confident that similar results will be observed after the experimental data are available.

As a comparative analysis of the presented method's efficiency against the existing related research, we evaluate our neural network and compare its performance with the approach presented in Ref. [40], where LISA was used to detect the SGWB from inflation, achieving a TPR of 95% and a FAR of 1%. Our network achieved a TPR of 99% with a FAR of 1.4%, demonstrating its capability to detect such SGWB signals while maintaining a comparable FAR and achieving a significantly higher TPR.

V. SUMMARY AND DISCUSSION

In our research, the computational configuration consists of an NVIDIA RTX 3090 GPU and an Intel(R) Xeon(R) Silver 4214R CPU (2.40 GHz, 48-thread processor). We employed full analytical approximations to simulate the

SGWB generated by cosmic strings and modeled the instrumental noise of both the LISA and Taiji detectors based on their sensitivity curves. By leveraging residual shrinkage networks and curriculum learning strategies, we developed a robust model and demonstrated that multi-detector joint observations significantly enhance the detection of SGWB signals.

In traditional machine learning and deep learning training, training samples are typically shuffled, and the model encounters samples of varying difficulty throughout the process, with these difficulties being uniformly distributed. In contrast, curriculum learning mimics human learning by gradually guiding the model through the learning process according to the difficulty of the samples, enabling it to acquire knowledge more efficiently.

These results suggest that machine learning holds great promise for the effective detection of SGWB signals by space-based gravitational wave detectors. Furthermore, combining multiple detectors could further improve the performance of machine learning models, enabling more efficient and accurate identification of SGWB signals.

Moreover, foreground noise presents a significant challenge for space-based detectors, as many of the received data streams are likely to be substantially affected by such noise. To address this, we incorporated foreground noise into our study, including contributions from DWDs and inspiraling BBHs/BNSs, based on the observations of LIGO and Virgo. By analyzing the model's performance across different SNR ranges, we found that our machine learning model can effectively handle the challenges posed by foreground noise, providing a solid foundation for future SGWB signal detection efforts.

It should be emphasized that besides cosmic strings, various other astrophysical processes can also generate SGWBs. Our study demonstrates that the multiband joint analysis method significantly enhances the discernment capability for SGWBs from different origins, providing a

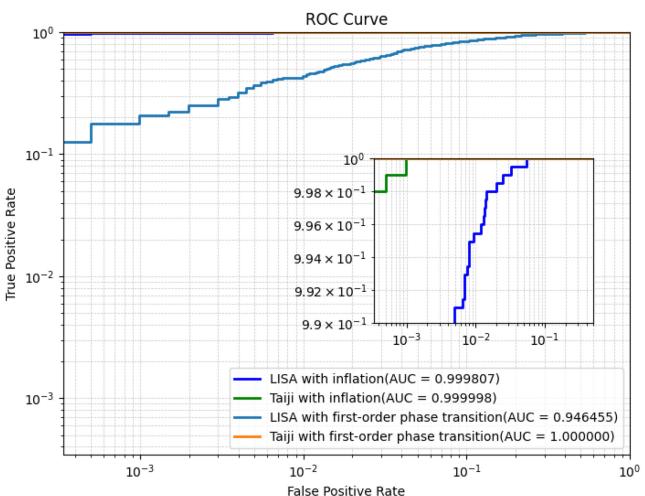


FIG. 11. ROC curves with SGWB identification from different sources.

novel technical approach to disentangle the components in composite SGWB signals.

The discussions presented here offer a meaningful attempt and valuable reference for future space-based gravitational wave data analysis powered by machine learning. Certainly, the observational data acquired by future gravitational wave detectors will exhibit substantially greater complexity than the simulated data analyzed in this study. The refined cosmic string models will impose more stringent constraints on gravitational wave production [82,83], thereby presenting significant challenges for the application of machine learning techniques in both data analysis and theoretical model interpretation. We plan to research it in the future work.

ACKNOWLEDGMENTS

This work was supported by the National Key Research and Development Program of China (Grant No. 2021YFC2203004), the Fundamental Research Funds for the Central Universities Project No. 2024IAIS-ZD009, the National Natural Science Foundation of China (Grants No. 12575072 and No. 12347101), and the Natural Science Foundation of Chongqing (Grant No. CSTB2023NSCQ-MSX0103).

DATA AVAILABILITY

The data that support the findings of this article are openly available [41].

APPENDIX: DATA COMPOSITION AND STRUCTURE

Our simulated data are time-domain sequences consisting of ten groups. The first six groups refer to gravitational waves originating from cosmic string and different kinds of noise ($n_i(t)$, $i = 1, 2, 3, 4, 5, 6$). In each group, the samples can be divided into two categories: one containing the SGWB from cosmic strings with noise, i.e., $h_{\text{SGWB}} + n_i(t)$; and the other only containing noise, i.e., $n_i(t)$. Two hundred thousand samples were included in each group, with 100,000 positive and 100,000 negative samples each, and the two groups were equalized. $n_{i=1,2,3}(t)$ stands for noise, respectively:

- $n_1(t)$ is the instrumental noise from LISA detector,
- $n_2(t)$ is the instrumental noise from Taiji detector, and
- $n_3(t)$ is the instrumental noise from LISA and Taiji detectors.

The third group, $h_{\text{SGWB}} + n_3(t)$, combines the data from the first two groups to simulate the effect of joint observation by LISA and Taiji. The remaining three groups are based on the joint observation instrumental noise with the additional foreground noises:

- $n_4(t)$ is $n_3(t)$ plus the foreground noise from double white dwarfs,

- $n_5(t)$ is $n_3(t)$ plus the foreground noise from gravitational wave background generated by binary black holes and binary neutron stars observed by LIGO and Virgo, and
- $n_6(t)$ consists of $n_4(t) + n_5(t)$.

Our analysis focuses on the frequency range of 1 mHz to 10 mHz, where both LISA and Taiji exhibit optimal detection sensitivity. Each data segment in the simulation spans 1 million seconds with a sampling rate of 3.6 mHz, resulting in 3,600 discrete data points per segment. Both the signal and noise components in our simulation are derived from the PSD ($S_{h/n}$), calculated using the following expression:

$$S_{h/n}(f) = \frac{3H_0^2}{4\pi^2} \frac{\Omega(f)}{f^3}. \quad (\text{A1})$$

We can generate both the SGWB signal and the noise using the noise_from_psd function in PyCBC [84], which synthesizes time-domain data or noise from a given PSD in the frequency domain. The generated signal and noise can then be combined for further analysis.

We construct the most typical TDI combination [85]: the Michelson combination (X,Y,Z). By linearly combining the X,Y,Z channels of the Michelson combination, the so-called optimal combination (A,E,T) can be obtained, defined as [86]

$$A = \frac{Z-X}{\sqrt{2}}, \quad E = \frac{X-2Y+Z}{\sqrt{6}}, \quad T = \frac{X+Y+Z}{\sqrt{3}}. \quad (\text{A2})$$

In an ideal situation, with equal arms and equal noise levels, this combination can eliminate the mutual correlation between signals and noise in different channels [86].

In Fig. 12, we present example samples of the generated data. As shown, the SGWB behaves like noise in the time

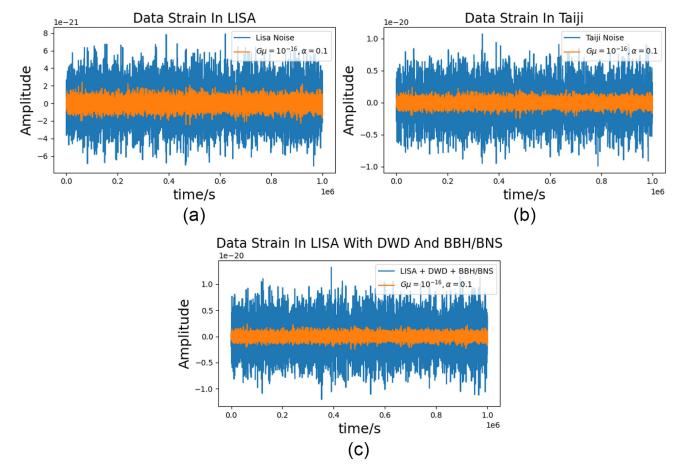


FIG. 12. Some data samples: (a) instrumental noise of LISA and the SGWB in the time-domain data, (b) instrumental noise of Taiji, (c) DWD noise and inspiral BBH/BNS based on the observations of LIGO and Virgo noise.

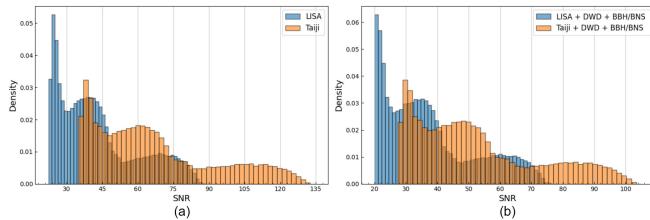


FIG. 13. SNR distributions of our samples. The blue column represents the SNR for the LISA detector, while the orange column represents the SNR for the Taiji detector. (a) shows the SNR with only the detector's intrinsic noise and (b) shows the data with both DWD and inspiral BBH/BNS based on the observations of LIGO and Virgo foreground noises.

domain, which complicates the process of signal identification. In Fig. 13, we illustrate their SNR distributions. For the same data, the minimum SNR of LISA is 22, while Taiji achieves a minimum SNR of 35. Similarly, the maximum SNR for Taiji is 134, whereas LISA only reaches a maximum SNR of 88. This difference arises because Taiji, with its longer arm lengths compared to LISA, has lower instrumental noise, resulting in a higher SNR. Additionally, the SNR distribution for LISA is more concentrated. However, the presence of foreground noise adds complexity to the data, further decreasing the SNRs. It is noteworthy that during data generation, we implemented an SNR threshold criterion: only samples with $\text{SNR} \geq 12$ after injecting foreground noise into LISA's measurements were retained for analysis. As demonstrated in Fig. 11, the representative cases shown in panels (a), (b), and (c) exhibit SNR values of 50, 48, and 37, respectively. The SNR is calculated using the formula provided in Ref. [87]:

$$\rho = 2 \left(\int_{f_{\min}}^{f_{\max}} \frac{|h(f)|^2}{S_n(f)} df \right)^{\frac{1}{2}}, \quad (\text{A3})$$

where $S_n(f)$ is the PSD of the noise and $h(f)$ is the signal.

In the final four groups, we investigate the classification performance for gravitational waves originating from different sources. For this part of the study, we used only single detectors. Each dataset was divided into two categories: one consists of SGWBs from cosmic strings with noise, i.e., $h_{\text{SGWB}} + n_{i=7,8}$; the other comprises SGWBs from other sources with noise, i.e., $h_{j=\text{PL,PT}} + n_{i=7,8}$, where h_{PL} is the SGWB from the power law model under inflation and h_{PT} is the SGWB from first-order phase transitions. The foreground noise is also consists of the double white dwarfs and gravitational wave background generated by BBHs and BNSs observed by LIGO and Virgo. $n_{i=7,8}(t)$ stands for noise, respectively:

- (i) $n_7(t)$ is the instrumental noise from LISA detector plus the foreground noise, and
- (ii) $n_8(t)$ is the instrumental noise from Taiji detector plus the foreground noise.

These two types of noise are then combined pairwise with two signals (h_{PL} and h_{PT}), resulting in four distinct datasets. While maintaining the same data generation protocol as described above, we implemented a composite-band detection strategy by dividing the spectrum into two distinct bands: the low-frequency band (0.5–5 mHz) and the high-frequency band (5–14 mHz). Twenty thousand samples were included in each group, with 10,000 positive ($h_{\text{SGWB}} + n_{i=7,8}$) and 10,000 negative ($h_{j=\text{PL,PT}} + n_{i=7,8}$) samples of the final four groups, and the two groups were equalized.

-
- [1] B. P. Abbott, R. Abbott, T. Abbott, M. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari *et al.*, Observation of gravitational waves from a binary black hole merger, *Phys. Rev. Lett.* **116**, 061102 (2016).
 - [2] P. Amaro-Seoane, H. Audley, S. Babak, J. Baker, E. Barausse, P. Bender, E. Berti, P. Binetruy, M. Born, D. Bortoluzzi *et al.*, Laser interferometer space antenna, [arXiv:1702.00786](https://arxiv.org/abs/1702.00786).
 - [3] Z. Luo, Z. Guo, G. Jin, Y. Wu, and W. Hu, A brief analysis to Taiji: Science and technology, *Results Phys.* **16**, 102918 (2020).
 - [4] J. Luo, L.-S. Chen, H.-Z. Duan, Y.-G. Gong, S. Hu, J. Ji, Q. Liu, J. Mei, V. Milyukov, M. Sazhin *et al.*, Tianqin: A space-borne gravitational wave detector, *Classical Quantum Gravity* **33**, 035010 (2016).
 - [5] M. Pitkin, S. Reid, S. Rowan, and J. Hough, Gravitational wave detection by interferometry (ground and space), *Living Rev. Relativity* **14**, 1 (2011).
 - [6] M. S. Turner, Detectability of inflation-produced gravitational waves, *Phys. Rev. D* **55**, R435 (1997).
 - [7] M. C. Guzzetti, N. Bartolo, M. Liguori, and S. Matarrese, Gravitational waves from inflation, *Riv. Nuovo Cimento* **39**, 399 (2016).
 - [8] S. Sarangi and S.-H. H. Tye, Cosmic string production towards the end of brane inflation, *Phys. Lett. B* **536**, 185 (2002).
 - [9] T. Damour and A. Vilenkin, Gravitational wave bursts from cosmic strings, *Phys. Rev. Lett.* **85**, 3761 (2000).
 - [10] M. J. Stott, T. Elghozi, and M. Sakellariadou, Gravitational wave bursts from cosmic string cusps and pseudocusps, *Phys. Rev. D* **96**, 023533 (2017).

- [11] X. Siemens, V. Mandic, and J. Creighton, Gravitational-wave stochastic background from cosmic strings, *Phys. Rev. Lett.* **98**, 111101 (2007).
- [12] E. Witten, Cosmic separation of phases, *Phys. Rev. D* **30**, 272 (1984).
- [13] A. Kosowsky, M. S. Turner, and R. Watkins, Gravitational waves from first-order cosmological phase transitions, *Phys. Rev. Lett.* **69**, 2026 (1992).
- [14] P. B. Dev and A. Mazumdar, Probing the scale of new physics by Advanced LIGO/Virgo, *Phys. Rev. D* **93**, 104001 (2016).
- [15] B. Von Harling, A. Pomarol, O. Pujolas, and F. Rompineve, Peccei-Quinn phase transition at LIGO, *J. High Energy Phys.* **01** (2020) 195.
- [16] C. Caprini and D. G. Figueira, Cosmological backgrounds of gravitational waves, *Classical Quantum Gravity* **35**, 163001 (2018).
- [17] S. Kuroyanagi, T. Chiba, and T. Takahashi, Probing the Universe through the stochastic gravitational wave background, *J. Cosmol. Astropart. Phys.* **11** (2018) 038.
- [18] N. Christensen, Stochastic gravitational wave backgrounds, *Rep. Prog. Phys.* **82**, 016903 (2018).
- [19] E. S. Phinney, P. L. Bender, and S. Buchman, LISA science requirements document, Max-Planck-Institut für Quantenoptik Report No. MPQ 233, 2002.
- [20] C. Liu, W.-H. Ruan, and Z.-K. Guo, Confusion noise from galactic binaries for Taiji, *Phys. Rev. D* **107**, 064021 (2023).
- [21] G. E. Box and G. C. Tiao, *Bayesian Inference in Statistical Analysis* (John Wiley & Sons, New York, 2011).
- [22] U. Von Toussaint, Bayesian inference in physics, *Rev. Mod. Phys.* **83**, 943 (2011).
- [23] G. Turin, An introduction to matched filters, *IRE Trans. Inf. Theory* **6**, 311 (1960).
- [24] L. Stanković and D. Mandic, Convolutional neural networks demystified: A matched filtering perspective-based tutorial, *IEEE Trans. Syst., Man Cyber.* **53**, 3614 (2023).
- [25] D. George and E. Huerta, Deep neural networks to enable real-time multimessenger astrophysics, *Phys. Rev. D* **97**, 044039 (2018).
- [26] H. Gabbard, M. Williams, F. Hayes, and C. Messenger, Matching matched filtering with deep networks for gravitational-wave astronomy, *Phys. Rev. Lett.* **120**, 141103 (2018).
- [27] X.-T. Zhang, C. Messenger, N. Korsakova, M. L. Chan, Y.-M. Hu, and J.-d. Zhang, Detecting gravitational waves from extreme mass ratio inspirals using convolutional neural networks, *Phys. Rev. D* **105**, 123027 (2022).
- [28] M. López, I. Di Palma, M. Drago, P. Cerdá-Durán, and F. Ricci, Deep learning for core-collapse supernova detection, *Phys. Rev. D* **103**, 063011 (2021).
- [29] H. Gabbard, C. Messenger, I. S. Heng, F. Tonolini, and R. Murray-Smith, Bayesian parameter estimation using conditional variational autoencoders for gravitational-wave astronomy, *Nat. Phys.* **18**, 112 (2022).
- [30] 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.* **127**, 241103 (2021).
- [31] R. E. Colgan, K. R. Corley, Y. Lau, I. Bartos, J. N. Wright, Z. Márka, and S. Márka, Efficient gravitational-wave glitch identification from environmental data through machine learning, *Phys. Rev. D* **101**, 102003 (2020).
- [32] M. Cavaglia, K. Staats, and T. Gill, Finding the origin of noise transients in LIGO data with machine learning, *Commun. Comput. Phys.* **25**, 963 (2019).
- [33] M. Razzano and E. Cuoco, Image-based deep learning for classification of noise transients in gravitational wave detectors, *Classical Quantum Gravity* **35**, 095016 (2018).
- [34] R. Ormiston, T. Nguyen, M. Coughlin, R. X. Adhikari, and E. Katsavounidis, Noise reduction in gravitational-wave data via deep learning, *Phys. Rev. Res.* **2**, 033066 (2020).
- [35] K. Mogushi, Reduction of transient noise artifacts in gravitational-wave data using deep learning, *arXiv:2105.10522*.
- [36] A. Torres-Forné, A. Marquina, J. A. Font, and J. M. Ibáñez, Denoising of gravitational wave signals via dictionary learning algorithms, *Phys. Rev. D* **94**, 124040 (2016).
- [37] W. Wei and E. Huerta, Gravitational wave denoising of binary black hole mergers with deep learning, *Phys. Lett. B* **800**, 135081 (2020).
- [38] H. Shen, D. George, E. Huerta, and Z. Zhao, Denoising gravitational waves using deep learning with recurrent denoising autoencoders, in *ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE, Piscataway, NJ, 2017); *arXiv:1711.09919*.
- [39] C. Chatterjee, L. Wen, F. Diakogiannis, and K. Vinsen, Extraction of binary black hole gravitational wave signals from detector data using deep learning, *Phys. Rev. D* **104**, 064046 (2021).
- [40] T. Zhao, R. Lyu, H. Wang, Z. Cao, and Z. Ren, Space-based gravitational wave signal detection and extraction with deep neural network, *Commun. Phys.* **6**, 212 (2023).
- [41] B.-R. Wang and J. Li, Ability of LISA, Taiji, and their networks to detect the stochastic gravitational wave background generated by cosmic strings, *Phys. Rev. D* **109**, 063520 (2024).
- [42] B.-R. Wang, J. Li, and H. Wang, Probing the gravitational wave background from cosmic strings with the alternative LISA-Taiji network, *Eur. Phys. J. C* **83**, 1010 (2023).
- [43] L. Sousa, P. P. Avelino, and G. S. Guedes, Full analytical approximation to the stochastic gravitational wave background generated by cosmic string networks, *Phys. Rev. D* **101**, 103508 (2020).
- [44] J. J. Blanco-Pillado, K. D. Olum, and B. Shlaer, Number of cosmic string loops, *Phys. Rev. D* **89**, 023512 (2014).
- [45] C. Ringeval, M. Sakellariadou, and F. R. Bouchet, Cosmological evolution of cosmic string loops, *J. Cosmol. Astropart. Phys.* **02** (2007) 023.
- [46] C. Martins and E. Shellard, Extending the velocity-dependent one-scale string evolution model, *Phys. Rev. D* **65**, 043514 (2002).
- [47] G. Boileau, N. Christensen, C. Gowling, M. Hindmarsh, and R. Meyer, Prospects for LISA to detect a gravitational-wave background from first order phase transitions, *J. Cosmol. Astropart. Phys.* **02** (2023) 056.
- [48] G. Wang and W.-B. Han, Alternative LISA-Taiji networks: Detectability of the isotropic stochastic gravitational wave background, *Phys. Rev. D* **104**, 104015 (2021).

- [49] M. Vallisneri and C. R. Galley, Non-sky-averaged sensitivity curves for space-based gravitational-wave observatories, *Classical Quantum Gravity* **29**, 124015 (2012).
- [50] T. L. Smith and R. R. Caldwell, LISA for cosmologists: Calculating the signal-to-noise ratio for stochastic and deterministic sources, *Phys. Rev. D* **100**, 104055 (2019).
- [51] J. Baker, J. Bellovary, P. L. Bender, E. Berti, R. Caldwell, J. Camp, J. W. Conklin, N. Cornish, C. Cutler, R. DeRosa *et al.*, The laser interferometer space antenna: Unveiling the millihertz gravitational wave sky, [arXiv:1907.06482](https://arxiv.org/abs/1907.06482).
- [52] W.-R. Hu and Y.-L. Wu, The Taiji program in space for gravitational wave physics and the nature of gravity, *Natl. Sci. Rev.* **4**, 685 (2017).
- [53] W.-H. Ruan, C. Liu, Z.-K. Guo, Y.-L. Wu, and R.-G. Cai, The LISA–Taiji network, *Nat. Astron.* **4**, 108 (2020).
- [54] E. Thrane and J. D. Romano, Sensitivity curves for searches for gravitational-wave backgrounds, *Phys. Rev. D* **88**, 124032 (2013).
- [55] Z.-C. Chen, F. Huang, and Q.-G. Huang, Stochastic gravitational-wave background from binary black holes and binary neutron stars and implications for LISA, *Astrophys. J.* **871**, 97 (2019).
- [56] G. Boileau, A. C. Jenkins, M. Sakellariadou, R. Meyer, and N. Christensen, Ability of LISA to detect a gravitational-wave background of cosmological origin: The cosmic string case, *Phys. Rev. D* **105**, 023510 (2022).
- [57] G. Wang, W.-T. Ni, W.-B. Han, and C.-F. Qiao, Algorithm for time-delay interferometry numerical simulation and sensitivity investigation, *Phys. Rev. D* **103**, 122006 (2021).
- [58] M. Braglia, G. Calcagni, G. Franciolini, J. Fumagalli, G. Nardini, M. Peloso, M. Pieroni, S. Renaux-Petel, A. Ricciardone, G. Tasinato *et al.*, Gravitational waves from inflation in LISA: Reconstruction pipeline and physics interpretation, *J. Cosmol. Astropart. Phys.* **11** (2024) 032.
- [59] M. Zhao, S. Zhong, X. Fu, B. Tang, and M. Pecht, Deep residual shrinkage networks for fault diagnosis, *IEEE Trans. Indus. Inf.* **16**, 4681 (2019).
- [60] Y. Li and H. Chen, Image recognition based on deep residual shrinkage network, in *2021 International Conference on Artificial Intelligence and Electromechanical Automation (AIEA)* (IEEE, Guangzhou, China, 2021), pp. 334–337.
- [61] X. Liu and Z. Hou, Improved deep residual shrinkage network for noisy image recognition, in *2024 5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT)* (IEEE, Nanjing, China, 2024), pp. 1408–1411.
- [62] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, in *Proceedings of the AAAI Conference on Artificial Intelligence* (AAAI, San Francisco, CA, USA, 2017), Vol. 31.
- [63] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, Las Vegas, NV, USA, 2016), pp. 770–778.
- [64] F. Yu, V. Koltun, and T. Funkhouser, Dilated residual networks, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, Honolulu, HI, USA, 2017), pp. 472–480.
- [65] D. L. Donoho, De-noising by soft-thresholding, *IEEE Trans. Inf. Theory* **41**, 613 (1995).
- [66] J. Joy, S. Peter, and N. John, Denoising using soft thresholding, *Int. J. Adv. Res. Electr. Electron. Instrum. Eng.* **2**, 1027 (2013).
- [67] K. Bredies and D. A. Lorenz, Linear convergence of iterative soft-thresholding, *J. Fourier Anal. Appl.* **14**, 813 (2008).
- [68] F. Girosi, M. Jones, and T. Poggio, Regularization theory and neural networks architectures, *Neural Comput.* **7**, 219 (1995).
- [69] Y. Tian and Y. Zhang, A comprehensive survey on regularization strategies in machine learning, *Inf. Fusion* **80**, 146 (2022).
- [70] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, Dropout: A simple way to prevent neural networks from overfitting, *J. Mach. Learn. Res.* **15**, 1929 (2014).
- [71] V. Tinto, Dropout from higher education: A theoretical synthesis of recent research, *Rev. Educ. Res.* **45**, 89 (1975).
- [72] M. Pauly, R. Keiser, and M. Gross, Multi-scale feature extraction on point-sampled surfaces, in *Computer Graphics Forum* (Wiley Online Library, Cambridge, MA, USA, 2003), Vol. 22, pp. 281–289.
- [73] A. E. Orhan and X. Pitkow, Skip connections eliminate singularities, [arXiv:1701.09175](https://arxiv.org/abs/1701.09175).
- [74] H. Wang, P. Cao, J. Wang, and O. R. Zaiane, Uctransnet: Rethinking the skip connections in u-net from a channel-wise perspective with transformer, in *Proceedings of the AAAI Conference on Artificial Intelligence* (AAAI, Palo Alto, CA, USA, 2022), Vol. 36, pp. 2441–2449.
- [75] J. Mukhoti, V. Kulharia, A. Sanyal, S. Golodetz, P. H. S. Torr, and P. K. Dokania, *Calibrating Deep Neural Networks using Focal Loss* (Curran Associates Inc., Red Hook, NY, USA, 2020).
- [76] X. Li, W. Wang, L. Wu, S. Chen, X. Hu, J. Li, J. Tang, and J. Yang, Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection, *Adv. Neural Inf. Process. Syst.* **33**, 21002 (2020).
- [77] M. Lisicki, A. Afkanpour, and G. W. Taylor, Evaluating curriculum learning strategies in neural combinatorial optimization, [arXiv:2011.06188](https://arxiv.org/abs/2011.06188).
- [78] G. Hacohen and D. Weinshall, On the power of curriculum learning in training deep networks, in *International Conference on Machine Learning* (PMLR, Long Beach, CA, USA, 2019), pp. 2535–2544.
- [79] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, Curriculum learning, in *Proceedings of the 26th Annual International Conference on Machine Learning* (Association for Computing Machinery, New York, NY, USA, 2009), pp. 41–48.
- [80] T. Fawcett, An introduction to roc analysis, *Pattern Recogn. Lett.* **27**, 861 (2006).
- [81] C. E. Metz, Basic principles of roc analysis, in *Seminars in Nuclear Medicine* (Elsevier, New York, 1978), Vol. 8, pp. 283–298.
- [82] M. Hindmarsh and J. Kume, Multi-messenger constraints on Abelian-Higgs cosmic string networks, *J. Cosmol. Astropart. Phys.* **04** (2023) 045.

- [83] J. Kume and M. Hindmarsh, Revised bounds on local cosmic strings from NANOGrav observations, *J. Cosmol. Astropart. Phys.* **12** (2024) 001.
- [84] A. Nitz, I. Harry, D. Brown, C.M. Biwer, J. Willis, T. Dal Canton, C. Capano, T. Dent, L. Pekowsky, S. De *et al.*, gwastro/PyCBC: v2. 1.2 release of PyCBC, Zenodo (2022).
- [85] J. Armstrong, F. Estabrook, and M. Tinto, Time-delay interferometry for space-based gravitational wave searches, *Astrophys. J.* **527**, 814 (1999).
- [86] T. A. Prince, M. Tinto, S. L. Larson, and J. Armstrong, LISA optimal sensitivity, *Phys. Rev. D* **66**, 122002 (2002).
- [87] K. Haris, A. K. Mehta, S. Kumar, T. Venumadhav, and P. Ajith, Identifying strongly lensed gravitational wave signals from binary black hole mergers, arXiv:1807.07062.