

Detection of Lensed Gravitational Waves in the Millihertz Band Using Frequency-Domain Lensing Feature Extraction Network

Tianlong Wang,^{1, 2, 3} Tianyu Zhao,^{2, *} Minghui Du,² Ziren Luo,^{2, 4, 5} Peng Dong,^{1, 4, 5, †} and Peng Xu^{2, 1, 4, 5, 6, ‡}

¹*School of Fundamental Physics and Mathematical Sciences, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences (UCAS), Hangzhou 310024, China*

²*Center for Gravitational Wave Experiment, National Microgravity Laboratory, Institute of Mechanics, Chinese Academy of Sciences, Beijing 100190, China*

³*University of Chinese Academy of Sciences, Beijing 100049, China*

⁴*Key Laboratory of Gravitational Wave Precision Measurement of Zhejiang Province, Hangzhou Institute for Advanced Study, UCAS, Hangzhou 310024, China*

⁵*Taiji Laboratory for Gravitational Wave Universe (Beijing/Hangzhou), UCAS, Beijing 100049, China*

⁶*Lanzhou Center of Theoretical Physics, Lanzhou University, Lanzhou 730000, China*

(Dated: December 29, 2025)

The space-based gravitational wave (GW) detectors are expected to observe lensed GW events, offering new opportunities for cosmology and fundamental physics. In the millihertz frequency band, the GW wavelength is often comparable to the Schwarzschild radius of the lens, where wave-optics effects are significant. Although traditional matched filtering is effective, the intense computational resources required motivate the search for more efficient alternatives to accelerate candidate event screening. To address this bottleneck, we introduce a Dual-Channel Lensing feature extraction eXtended Long Short-Term Memory Network (DCL-xLSTM). Unlike conventional recurrent architectures, DCL-xLSTM employs a matrix-valued memory structure and a memory-mixing mechanism to effectively capture amplitude diffraction patterns that span the entire millihertz frequency band. Trained on data generated by Point Mass (PM) and Singular Isothermal Sphere (SIS) models accounting for the transition from wave-optics to geometric-optics, the proposed method achieves an area under the curve (AUC) exceeding 0.99, maintaining a true positive rate (TPR) above 98% at a false positive rate (FPR) below 1%. The network is robust against variations in signal-to-noise ratio, lens type, and lens mass, establishing its viability as a high-efficiency tool for future space-based GW detection.

I. INTRODUCTION

The first detection of the GW150914 event initiated the era of gravitational wave (GW) astronomy [1]. Since then, the ground-based detector network has expanded the catalog to 200 confirmed GWs [2–5], enabling unprecedented tests of general relativity in the strong-field regime [6] and provided novel insights into astrophysical populations and merger rates of compact objects [7]. Gravitational lensing has been verified by electromagnetic (EM) observations for decades and has led to several groundbreaking findings in astrophysics [8–17].

The next frontier in GW observation lies in the millihertz frequency band, which will be accessible to future space-borne interferometers such as LISA [18], Taiji [19], and TianQin [20]. These observatories are expected to detect the mergers of massive black hole binaries (MBHBs) to high cosmological redshifts [14, 17, 21, 22]. At these cosmological distances, the probability of strong gravitational lensing becomes significant, rendering the detection of lensed events not merely a possibility, but an expectation [14, 23]. A key feature of GW lensing in the millihertz band is the critical role of wave-optics [14, 24].

The characteristic GW wavelength can be comparable to or larger than the Schwarzschild radius of the lens for lens masses $M_L \lesssim 10^8 M_\odot$, causing diffraction and interference effects to become significant [13, 22, 25, 26]. This contrasts to the geometric optical limit valid for ground-based detectors, where the wavelength is negligible compared to the lens scale, resulting in multiple images with time delays and magnifications [13, 14, 22].

Current state-of-the-art methods for identifying lensed GWs, such as Bayesian parameter estimation and joint parameter estimation (JPE) [13, 26–33], are known to be accurate, but often computationally expensive [27, 34]. A complete parameter estimation for even a single event can take hours [34, 35]. The computational cost presents a bottleneck for future GW surveys. Space-based and third-generation observatories are expected to detect $O(10^5)$ to $O(10^6)$ events annually [34–39]. Because the search for lensed pairs scales quadratically with the number of events ($O(N^2)$) [27], this volume of data would require $O(10^{10})$ to $O(10^{12})$ pairwise comparisons [35, 37, 38]. This challenge motivates the development of faster methods, such as deep learning [40], which can rapidly detect the large number of candidate pairs [11, 12, 25, 34, 35].

Early proof-of-principle studies demonstrated that deep learning methods could distinguish GW signals with high precision [41–43]. Recent studies have shown that deep learning frameworks have emerged as a promising high-speed alternative to identify lensed GW signals [11, 34, 35].

* zhaotianyu@imech.ac.cn

† dongpeng@ucas.ac.cn

‡ xupeng@imech.ac.cn

Kim et al. [11] utilized a VGG-19 network [44] on spectrograms, successfully identifying microlensing-induced "beating patterns" [45–47] and performing parameter estimation for lens properties. The method typically reframes the identification task as an image classification problem, analyzing 2D time-frequency representations such as spectrograms or Q-transforms [11, 12, 34, 35]. Goyal et al. [35] introduced a hybrid approach, combining a DENSENET201 model [48] for Q-transform analysis with an XGBoost algorithm [49] to incorporate skymap localization data. Their method achieved performance comparable to Bayesian techniques but reduced the computation time by orders of magnitude [35]. The SLICK pipeline, introduced by Magare et al. [34], further enhanced this by using a parallel network architecture to analyze both the Q-transform and the Sine-Gaussian maps simultaneously, finding that combined input significantly reduces false positives [34]. More recently, architectures have evolved to Vision Transformers (ViT) [50]. Li et al. [12] proposed the Squeeze-and-Excitation Multilayer Perceptron Data-efficient Image Transformer model, which classifies spectrogram pairs by explicitly modeling their morphological similarity [12]. Specialized models, such as the Wavelet Convolutional Detector, have also been developed to specifically target the diffraction patterns associated with microlensing by compact dark matter [25]. These collective works demonstrate the rapid maturation of deep learning as a powerful and efficient tool for GW data analysis. However, image-based approaches [11, 35] introduce additional computational steps during spectrogram generation and may limit the visibility of characteristic diffraction arising from oscillatory modulations in spectral amplitude [45, 46]. Standard spectrograms, subject to time-frequency resolution constraints, can sometimes under-resolve these fine-scale interference structures, effectively smoothing out the subtle signatures required for precise model identification [22]. Therefore, analyzing the raw strain data via a sequence-based modeling approach offers a robust alternative to retain the full fidelity of these effects.

In this work, we present a deep learning framework for the identification of lensed GWs in the millihertz band. First, we develop a classifier that encompasses the continuous transition from the wave-optics to the geometric-optics regime. By extending beyond asymptotic limits, our dataset and model are designed to accurately capture complex diffraction-induced amplitude modulation to ensure physical fidelity across the diverse lens masses relevant to LISA. Second, we adopt a direct sequence modeling approach that leverages the full resolution of the frequency-domain amplitude spectrum. Unlike 2D image-based methods, where fine-scale spectral features may be attenuated due to resolution constraints, our method analyzes whitened Time-Delay Interferometry (TDI) Channel A and E strain, which allows for explicit preservation of the high-frequency oscillatory modulations, providing a robust basis for model identification. Third, we employ a **Dual-Channel Lensing** feature ex-

traction eXtended Long Short-Term Memory Network (**DCL-xLSTM**). Compared with conventional LSTM, it introduces a matrix-based memory structure and a memory-mixing mechanism that allow for more retention of intricate diffraction details in long spectral sequences. This architecture enhances the model's ability to handle long-term dependencies beyond the capabilities of standard LSTMs, all while maintaining favorable linear computational complexity.

The remainder of this article is organized as follows. Section II details the methodology, including the physics of lensing, the waveform simulation pipeline, and the DCL-xLSTM network architecture. Section III presents the classification performance, analyzing the robustness of different lens models, masses, and signal-to-noise ratios. Finally, Section IV summarizes our findings and discusses their implications for future multi-messenger astronomy.

II. METHOD

A. Lens Models

Gravitational lensing distorts GW signals through mechanisms largely determined by the interplay between the GW wavelength λ_{GW} and the characteristic size of the lens (Schwarzschild radius R_s). This relationship can be characterized by the dimensionless frequency parameter w illustrated in Figure 1. Following the convention in the wave-optics literature [14, 22], we define w in terms of the redshifted lens mass $M_{Lz} = M_L(1 + z_L)$:

$$w = \frac{8\pi GM_{Lz} f}{c^3}, \quad (1)$$

as Figure 1 shows, the parameter divides the phenomena into two distinct regimes. In the wave-optics regime ($w \lesssim 1$), diffraction effects dominate, causing amplitude oscillations and phase shifts without the formation of discrete geometric images. In the geometric-optics regime ($w \gg 1$), the diffraction integral approximates a sum over discrete stationary points, manifesting as multiple images with magnifications and constant time delays. However, the critical transition regime ($0.1 < w < 10$) bridges these extremes, corresponding to the scenario where the GW wavelength is comparable to the Schwarzschild radius of the lens. In this domain, the validity of both the diffraction limit and the stationary phase approximation broken down. As shown in the right panel, the transition regime produces the first prominent peak in the amplification factor $|F(w)|$, which can be attributed to constructive interference. This regime also marks the crossover from a smooth, non-oscillatory diffraction-dominated behavior at $w \lesssim 1$ to the oscillatory interference fringes characteristic of the geometric-optics limit at $w \gg 1$.

At a fixed frequency f , the lensed waveforms $\tilde{h}_{+,\times}^L(f)$ relate to the unlensed waveforms through the complex amplification factor $F(f)$:

$$\tilde{h}_{+,\times}^L(f) = F(f) \tilde{h}_{+,\times}(f). \quad (2)$$

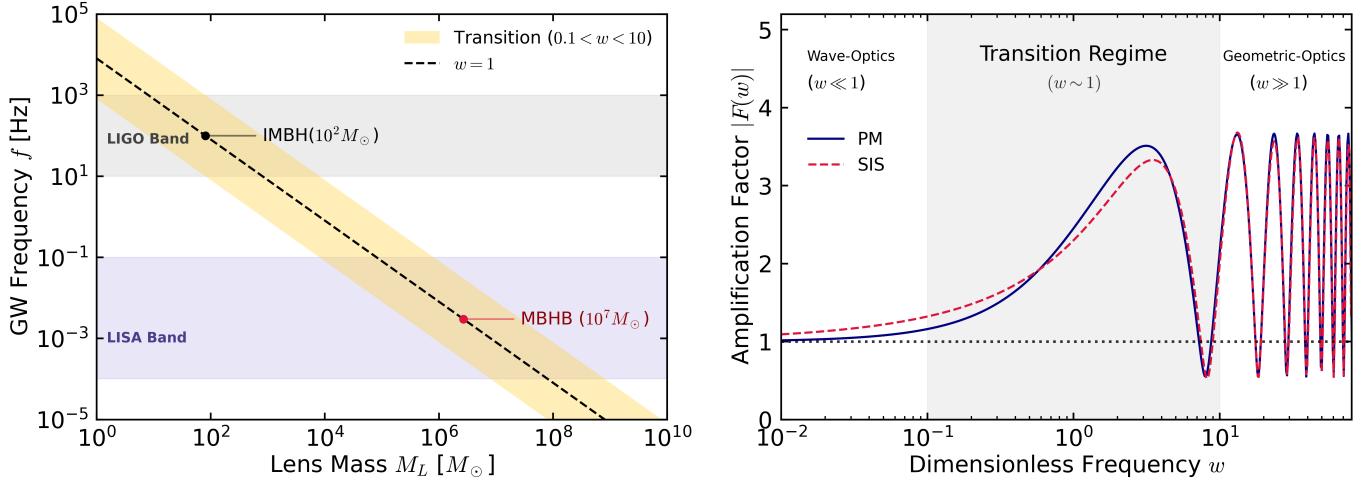


Figure 1. Overview of gravitational lensing regimes and signal amplification. **Left:** The lens mass (M_L) versus GW frequency (f) parameter space. The dashed line ($w = 1$) marks the transition between geometric and wave optics, with the shaded orange band ($0.1 < w < 10$). Sensitivity bands for LISA (blue) and LIGO (gray) are shown for references. **Right:** The frequency amplification factor $|F(w)|$ is a function of dimensionless frequency w for Point Mass (PM, solid) and Singular Isothermal Sphere (SIS, dashed) models with a source impact parameter $y = 0.3$. The gray shaded region corresponds to the transition regime shown in the left panel.

1. Point Mass Lens

The point mass lens represents the simplest point case, characterized by a density profile $\rho(r) = M_L \delta^3(r)$ where M_L denotes the lens mass, which is applicable to compact objects such as black holes. The amplification factor $F(w)$ is given by [22, 51]:

$$F(w) = \exp \left[\frac{\pi w}{4} + \frac{i w}{2} \left(\ln \frac{w}{2} - 2\phi_m(y) \right) \right] \times \Gamma \left(1 - \frac{i w}{2} \right) {}_1F_1 \left(\frac{i w}{2}, 1, y^2 \frac{i w}{2} \right), \quad (3)$$

where $\phi_m(y) = \frac{(x_m - y)^2}{2} - \ln x_m$ with $x_m = \frac{y + \sqrt{y^2 + 4}}{2}$. Here, $\Gamma(z)$ is the Euler gamma function and ${}_1F_1(a, b, z)$ is Kummer's confluent hypergeometric function. The parameter y represents the dimensionless position of the source, defined as $y = \frac{\gamma D_L}{\xi_0 D_S}$, where γ is the displacement of the source, D_L and D_S are the distances to the lens and to the source, respectively, and $\xi_0 = \sqrt{(4GM_L/c^2)D_L D_S}/c$ is the Einstein radius of the lens, as illustrated in Figure 2. The analytical formula is able to give rise to the well-known approximations.

2. Singular Isothermal Sphere Lens

The SIS model, described by the density profile $\rho(r) = \sigma_v^2/(2\pi r^2)$ with σ_v representing the velocity dispersion, which is a more complex representation suitable for galaxies or dark matter halos. The surface density is charac-

terized as: $\Sigma(\xi) = \frac{\sigma_v^2}{2\xi}$ with the Einstein radius ξ_0 serving as normalization constant with $\xi_0 = 4\pi\sigma_v^2 \frac{D_L D_{LS}}{D_S}$.

The general solution for the amplification factor, valid across all physical regimes from wave-optics to geometric-optics, is given by the following diffraction integral [22, 51]:

$$F(f) = -i w e^{i w y^2 / 2} \times \int_0^\infty dx x J_0(wxy) \exp [i w (\frac{1}{2} x^2 - x + \phi_m(y))], \quad (4)$$

where J_0 is the zeroth-order Bessel function, $\phi_m(y) = y + 1/2$ and the lens mass is defined by:

$$M_{Lz} = (1 + z_L) \frac{4\pi^2}{G c^2} \sigma_v^4 \frac{D_L D_{LS}}{D_S}, \quad (5)$$

according to [52, 53], we take $\sigma_v \simeq 20\text{--}40 \text{ km s}^{-1}$ for lenses of similar mass scale. For the parameter $w > 1$, the integrand becomes highly oscillatory. To ensure numerical stability and eliminate aliasing artifacts in the training data, we evaluate this integral using the Levin collocation method [54], which transforms the oscillatory quadrature into a non-oscillatory system of ordinary differential equations.

B. Waveform Simulation

We simulate lensed GWs originating from coalescing MBHB [55, 56]. Our simulation pipeline proceeds as follows: The frequency-domain waveforms $\tilde{h}_{+,\times}(f)$ are generated using the *IMRPhenomD* Model [57, 58]. The source

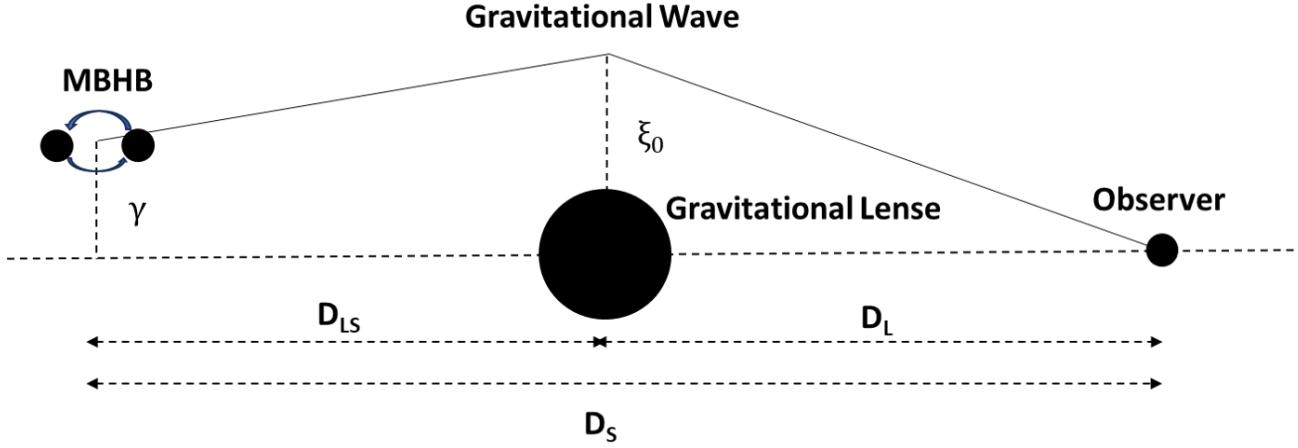


Figure 2. **A schematic diagram of gravitational lensing of GWs.** The signal from binary system is deflected by an intervening lens. Distances are shown: source-to-lens (D_{LS}), lens-to-observer (D_L). The impact parameter of the source relative to the lens axis is γ , and ξ_0 is the Einstein radius in the lens plane.

parameters are drawn from the distributions detailed in Table I. We apply the complex amplification factor $F(f)$ directly to the source waveform: $\tilde{h}_{+, \times}^L(f) = F(f)\tilde{h}_{+, \times}(f)$. The step incorporates the frequency-dependent amplitude modulations derived from the PM or SIS models. And the lensed waveforms are projected onto the LISA constellation using the *bbhx* [59, 60] software package, which computes the TDI observables A, E and T by accounting for the spacecraft's orbital motion and the frequency-dependent antenna response functions. Finally, Gaussian noise colored by the LISA noise PSD $S_n(f)$ is added to the signals projected by the detector. The SNR is computed after lensing and projection to reflect the observed signal strength.

The parameter space is characterized by eight physical parameters:

$$\Theta = \{\eta, M, t_c, D_L(z_S), \theta_S, \phi_S, \iota, \psi\}, \quad (6)$$

where $\eta \equiv m_1 m_2 / M^2$ is the symmetric mass ratio, $M \equiv m_1 + m_2$ denotes the total mass, and $D_L(z_S)$ represents the luminosity distance at the source redshift z_S . The angular parameters (θ_S, ϕ_S) specify sky localization in the detector coordinates, while ι and ψ determine the orbital inclination and polarization angles, respectively. Table I summarizes the parameter ranges adopted for our simulations.

The detector projection combines GW polarizations through the frequency-domain response:

$$\tilde{h}^{A,E,T}(f) = \sum_{lm} \mathcal{T}^{A,E,T}(f, t_{lm}(f)) \tilde{h}_{lm}(f), \quad (7)$$

where the time-frequency mapping follows from the stationary phase approximation:

$$t_{lm}(f) = t_{\text{ref}} - \frac{1}{2\pi} \frac{d\phi_{lm}(f)}{df}. \quad (8)$$

Table I. Parameter space for the simulated lensed GWs used in this study.

Parameter	Range	Units
1. GW Source Parameters		
Source Mass (M)	$[10^4, 10^6]$	M_\odot
Mass Ratio (η)	$[0.2, 0.8]$	—
Source Redshift (z_S)	$[0.1, 3.0]$	—
Sky Pos. (θ_S, ϕ_S)	$[0, \pi] \times [0, 2\pi]$	rad
Inclination (ι)	$[0, \pi]$	rad
Polarization (ψ)	$[0, \pi]$	rad
Coalesce Time (t_c)	$[-3600, 3600]$	s
2. Lens Parameters		
Lens Model	PM and SIS	—
M_L	$[10^6, 10^8]$	M_\odot
Lens Redshift (z_L)	$[0.1, 3.0]$	—
Impact Param. (y)	$[0.1, 5.0]$	—
3. Simulation Parameters		
Waveform Model	IMRPhenomD	—
Noise Model	LISA PSD	—
SNR	$[20, 70]$	—

The uncorrelated TDI channels are constructed as follows [61]:

$$A = \frac{1}{\sqrt{2}}(Z - X), \quad (9)$$

$$E = \frac{1}{\sqrt{6}}(X - 2Y + Z), \quad (10)$$

$$T = \frac{1}{\sqrt{3}}(X + Y + Z), \quad (11)$$

with $\mathcal{T}^{A,E,T}$ encoding both the antenna pattern and LISA's orbital motion. The implementation uses GPU-accelerated batch processing of harmonic modes and cubic spline interpolation for the response functions [59, 60].

The noise characteristics of the frequency-domain are quantified by the one-sided power spectral density $S_n(f)$ [62].

$$P_{\text{OMS}}(f) = (15 \text{ pm})^2 \left[1 + \left(\frac{2 \text{ mHz}}{f} \right)^4 \right] \left(\frac{2\pi f}{c} \right)^2 \text{ Hz}^{-1}, \quad (12)$$

$$P_{\text{acc}}(f) = (3 \text{ fm/s}^2)^2 \left[1 + \left(\frac{0.4 \text{ mHz}}{f} \right)^2 \right] \times \left[1 + \left(\frac{f}{8 \text{ mHz}} \right)^4 \right] \frac{1}{(2\pi f)^4}. \quad (13)$$

For the A and E TDI channels, PSDs can be derived as derived as [62]

$$S_{A0,E0}(f) = 2[2(3 + 2\cos x + \cos 2x)P_{\text{acc}}(f) + (2 + \cos x)P_{\text{OMS}}(f)], \quad (14)$$

$$S_{A1,E1}(f) = 4\sin^2 x S_{A0,E0}(f), \quad (15)$$

$$S_{A2,E2}(f) = 4\sin^2 2x S_{A0,E0}(f), \quad (16)$$

where $x = 2\pi f L/c$, $L = 2.5 \text{ Gm}$ is the length of the LISA arms and c is the speed of light. Here, the subscripts 0, 1, and 2 denote the 0th-, 1st-, and 2nd-generation TDI combinations (i.e., different TDI orders) for the A and E channels. We adopt $S_{A2,E2}(f)$ in this work.

The SNR ρ for a waveform $h(t)$ is computed using the inner product:

$$\rho = \sqrt{4 \operatorname{Re} \int_{f_{\min}}^{f_{\max}} \frac{|\tilde{h}(f)|^2}{S_N(f)} df}, \quad (17)$$

where the integration interval $[f_{\min}, f_{\max}]$ corresponds to the sensitive band of LISA. Noise realizations are scaled to achieve target SNR while preserving the statistical properties of the LISA noise model.

C. Dual-Channel Lensing Feature Extraction xLSTM

Recurrent neural networks (RNNs) update a hidden state sequentially as

$$\mathbf{h}_t = \phi(\mathbf{W}\mathbf{x}_t + \mathbf{U}\mathbf{h}_{t-1} + \mathbf{b}), \quad (18)$$

where \mathbf{x}_t is the input at step t , \mathbf{h}_t is the hidden state, $\phi(\cdot)$ is a nonlinear activation, and $\mathbf{W}, \mathbf{U}, \mathbf{b}$ are trainable parameters. Long Short-Term Memory (LSTM) networks [63] introduce a gated memory state \mathbf{c}_t and update

$$\begin{aligned} \mathbf{f}_t &= \sigma(\mathbf{W}_f \mathbf{x}_t + \mathbf{U}_f \mathbf{h}_{t-1} + \mathbf{b}_f), \\ \mathbf{i}_t &= \sigma(\mathbf{W}_i \mathbf{x}_t + \mathbf{U}_i \mathbf{h}_{t-1} + \mathbf{b}_i), \\ \mathbf{o}_t &= \sigma(\mathbf{W}_o \mathbf{x}_t + \mathbf{U}_o \mathbf{h}_{t-1} + \mathbf{b}_o), \\ \tilde{\mathbf{c}}_t &= \tanh(\mathbf{W}_c \mathbf{x}_t + \mathbf{U}_c \mathbf{h}_{t-1} + \mathbf{b}_c), \\ \mathbf{c}_t &= \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t, \\ \mathbf{h}_t &= \mathbf{o}_t \odot \tanh(\mathbf{c}_t), \end{aligned} \quad (19)$$

where $\sigma(\cdot)$ is the sigmoid function and \odot denotes element-wise multiplication.

We adopt the xLSTM architecture [64], which strengthens LSTM-style gating via exponential gates and employs two cell variants: sLSTM (vector-valued memory) and mLSTM (matrix-valued memory). For both variants, gate pre-activations are parameterized by

$$\tilde{\mathbf{g}}_t = \mathbf{W}_g \mathbf{x}_t + \mathbf{U}_g \mathbf{h}_{t-1} + \mathbf{b}_g, \quad g \in \{i, f, o\}, \quad (20)$$

and exponential gating is applied element-wise, e.g.

$$\mathbf{i}_t = \exp(\tilde{\mathbf{i}}_t), \quad \mathbf{f}_t = \exp(\tilde{\mathbf{f}}_t), \quad \mathbf{o}_t = \sigma(\tilde{\mathbf{o}}_t). \quad (21)$$

To stabilize exponential gates, xLSTM introduces a stabilizer state \mathbf{m}_t and uses stabilized gates $\tilde{\mathbf{i}}'_t, \tilde{\mathbf{f}}'_t$ defined by

$$\begin{aligned} \mathbf{m}_t &= \max(\log \mathbf{f}_t + \mathbf{m}_{t-1}, \log \mathbf{i}_t), \\ \tilde{\mathbf{i}}'_t &= \exp(\log \mathbf{i}_t - \mathbf{m}_t), \\ \tilde{\mathbf{f}}'_t &= \exp(\log \mathbf{f}_t + \mathbf{m}_{t-1} - \mathbf{m}_t), \end{aligned} \quad (22)$$

where $\max(\cdot, \cdot)$, $\log(\cdot)$, and $\exp(\cdot)$ are applied element-wise.

The sLSTM retains a vector-valued memory $\mathbf{c}_t \in \mathbb{R}^{d_h}$ and introduces an explicit normalizer state $\mathbf{n}_t \in \mathbb{R}^{d_h}$:

$$\begin{aligned} \mathbf{c}_t &= \mathbf{f}'_t \odot \mathbf{c}_{t-1} + \mathbf{i}'_t \odot \tilde{\mathbf{c}}_t, \\ \mathbf{n}_t &= \mathbf{f}'_t \odot \mathbf{n}_{t-1} + \mathbf{i}'_t, \\ \mathbf{h}_t &= \mathbf{o}_t \odot \frac{\mathbf{c}_t}{\max(\mathbf{n}_t, \mathbf{1})}, \end{aligned} \quad (23)$$

where $\mathbf{1}$ is the all-ones vector and the division is element-wise.

The mLSTM replaces \mathbf{c}_t by a matrix memory $\mathbf{C}_t \in \mathbb{R}^{d \times d}$. Using a value vector $\mathbf{v}_t \in \mathbb{R}^d$, the memory update takes an outer-product form

$$\mathbf{C}_t = \mathbf{f}'_t \mathbf{C}_{t-1} + \mathbf{i}'_t (\mathbf{v}_t \mathbf{v}_t^\top), \quad \mathbf{v}_t = \mathbf{W}_v \mathbf{x}_t + \mathbf{U}_v \mathbf{h}_{t-1}, \quad (24)$$

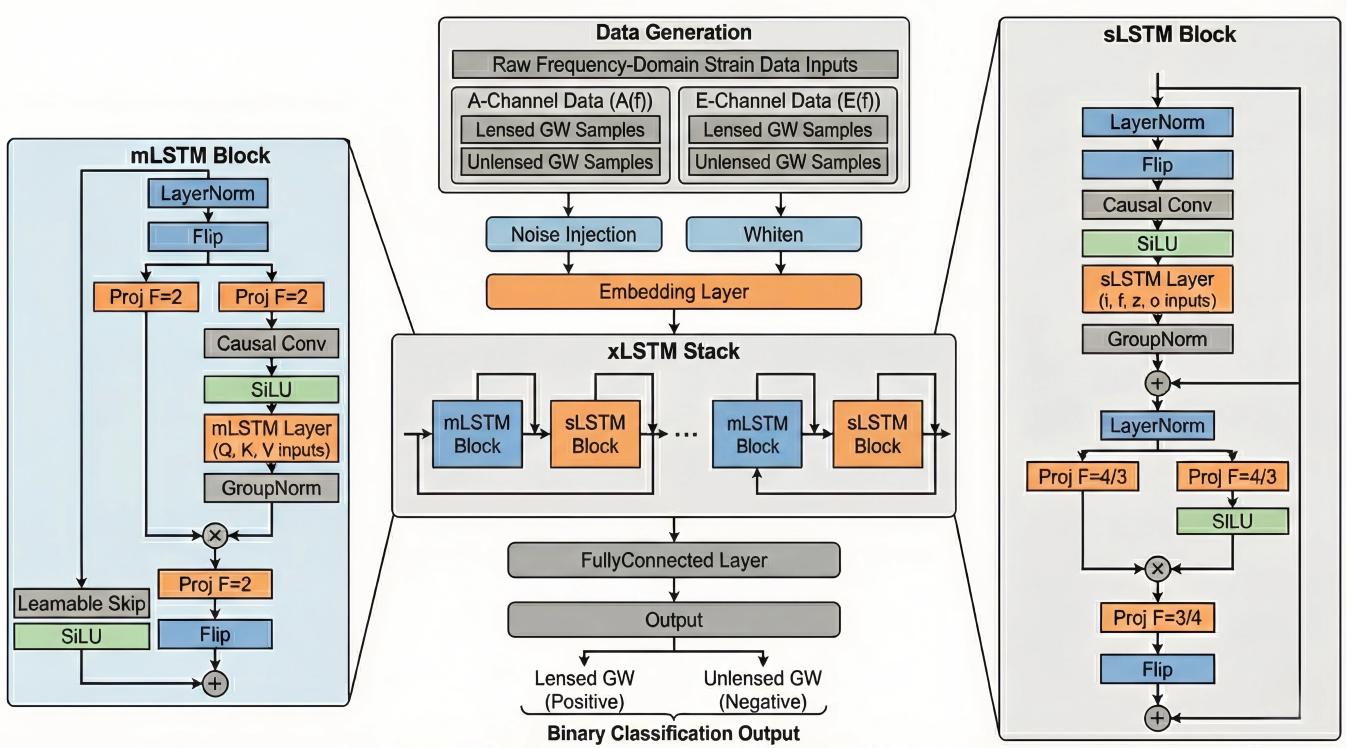


Figure 3. The architecture of the DCL-xLSTM for GW classification. Frequency-domain strain amplitudes from the A and E TDI channels are preprocessed and sampled at 2048 points to form a dual-channel input sequence $\{\mathbf{x}_t\}$, where $\mathbf{x}_t = (|A(f_t)|, |E(f_t)|)$. The sequence is processed by a stack of mLSTM and sLSTM blocks, which extract long-range spectral features and cross-channel correlations characteristic of lensing. The final hidden representation is passed through a fully connected layer to produce a probability for lensed versus unlensed gravitational-wave events.

where $(\cdot)^\top$ denotes transpose.

The hidden state is obtained by a normalized retrieval using a query vector $\mathbf{q}_t \in \mathbb{R}^d$ and a normalization vector $\mathbf{n}_t \in \mathbb{R}^d$.

$$\begin{aligned} \mathbf{h}_t &= \mathbf{o}_t \odot \frac{\mathbf{C}_t \mathbf{q}_t}{\max(|\mathbf{n}_t^\top \mathbf{q}_t|, 1)}, \\ \mathbf{q}_t &= \mathbf{W}_q \mathbf{x}_t, \\ \mathbf{n}_t &= \mathbf{W}_n \mathbf{x}_t + \mathbf{U}_n \mathbf{h}_{t-1}, \end{aligned} \quad (25)$$

where $|\cdot|$ denotes the absolute value.

The input consists of strain amplitude spectra from the A and E TDI channels. Each GW event is transformed into whitened frequency-domain amplitudes

$$\begin{aligned} \mathbf{A} &= \{|A(f_1)|, \dots, |A(f_{2048})|\}, \\ \mathbf{E} &= \{|E(f_1)|, \dots, |E(f_{2048})|\}, \end{aligned}$$

sampled at 2048 frequency points. The A and E channels respond differently to the same GW due to their distinct interferometric combinations and antenna pattern functions. To exploit this complementarity, we construct a two-dimensional feature vector at each frequency index t :

$$\mathbf{x}_t = [|A(f_t)|, |E(f_t)|] \in \mathbb{R}^2,$$

forming a dual-channel sequence $\{\mathbf{x}_t\}_{t=1}^{2048}$ that is fed into the first block. This representation enables the network to

learn cross-channel correlations and identify coherent amplitude modulations characteristic of gravitational-wave lensing.

The final hidden state of the stack is passed through a fully connected layer to yield a lensing probability

$$p = P(\text{lensed} \mid A(f), E(f)).$$

A GW event is classified as lensed if $p > 0.5$, and as unlensed otherwise. In an extended formulation, the network may additionally output channel-wise probabilities p_A and p_E ; a joint decision rule is then defined by

$$p_{\text{joint}} = p_A p_E, \quad p_{\text{joint}} > 0.5 \Rightarrow \text{lensed},$$

reflecting the approximate statistical independence of noise in the A and E channels. The detailed architecture of the proposed network is illustrated in Figure 3, which presents the overall pipeline.

III. RESULTS

To evaluate classification performance, we constructed a balanced dataset of 16,000 samples, consisting of equal numbers of lensed ($N = 8,000$) and unlensed ($N = 8,000$) waveforms. We divided the samples into two distinct

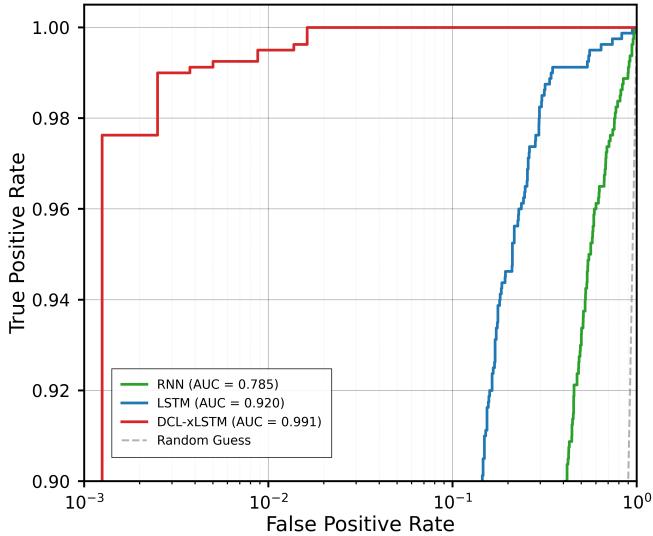


Figure 4. Receiver operating characteristic (ROC) curves for the binary classification task on the combined dataset. The DCL-xLSTM model (red solid line, AUC = 0.991) demonstrates performance significantly outperforming the LSTM (blue dashed line, AUC = 0.920) and the RNN (green dashed line, AUC = 0.785). The gray dashed line represents the random classifier baseline (AUC = 0.5). The x-axis (False Positive Rate) is plotted on a logarithmic scale to highlight performance at low FPRs, which is critical for rare event detection.

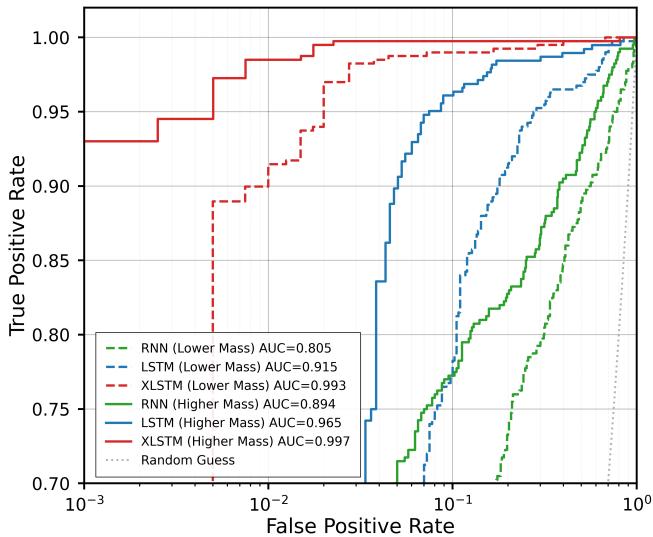


Figure 5. Comparative ROC curves for GW signals classification. The plot illustrates the performance of DCL-xLSTM (red), LSTM (blue), and RNN (green) models against the Higher Mass (solid lines) and Lower Mass (dashed lines) datasets. While all models exhibit improved sensitivity for higher lens masses (solid curves), the DCL-xLSTM model displays stability, showing minimal performance degradation between mass regimes (AUC decreases only from 0.997 to 0.993). In contrast, the LSTM and RNN models show more significant performance gaps between the two datasets, highlighting the superior generalization capability of the DCL-xLSTM architecture.

mass ranges to test the network’s sensitivity across different diffraction conditions. The *High Mass* group ($M_L \in [10^7, 10^8] M_\odot$) represents a regime with strong wave-optics effects, where signal distortions are clearly visible. In contrast, the *Low Mass* group ($M_L \in [10^6, 10^7] M_\odot$) corresponds to the onset of diffraction, where the lensing features are subtle and the waveform deviations are lower. This approach ensures that the model is tested against both clear and faint lensing signatures within the transition region. The lensing effects were generated using two standard lens models: PM model [11] and SIS model [24]. To simulate realistic observation conditions, all signals were whitened and injected into Gaussian noise based on the LISA’s noise model. The optimal SNR was sampled from a uniform distribution of 30 to 70, allowing us to assess performance across a wide range of signal strengths. The dataset was randomly divided into training (70%), validation (15%), and testing (15%) sets, with strict separation to prevent data leakage. We compared the proposed xLSTM-based classifier with the RNN and LSTM models. The primary metric for performance is the area under the receiver operating characteristic curve (AUC), which measures the ability of the network to distinguish between classes independent of specific decision thresholds.

A. General Classification Performance

The classification performance in the dataset is evaluated in Figure 4. The DCL-xLSTM model achieves near-perfect separability between lensed and unlensed classes, with an AUC of 0.991. A practical advantage of the DCL-xLSTM classifier is its performance at low false positive rate (FPR). At an FPR of 10^{-3} , it maintains a true positive rate (TPR) exceeding 0.98, which is critical for detecting rare lensing events with high confidence.

In contrast, standard LSTM (AUC = 0.920) and RNN (AUC = 0.785) exhibit significantly degraded performance in this low-FPR regime, with their TPR falling substantially below that of DCL-xLSTM.

The performance is consistent with the models’ architectural capacity to capture the long-range, complex dependencies inherent in the representation of lensed waveforms. The better performance of DCL-xLSTM can be attributed to its matrix-valued memory and exponential gating mechanisms, which provide the necessary representational power to model the intricate correlations arising from the hybrid lensing physics. Based on the result, DCL-xLSTM thus serves as the optimal foundation for the detailed analyzes that follow.

B. Robustness Across Lensing Regimes

To assess the robustness of our approach under different diffractive conditions, we categorized the classification results according to lens mass. The higher lens mass

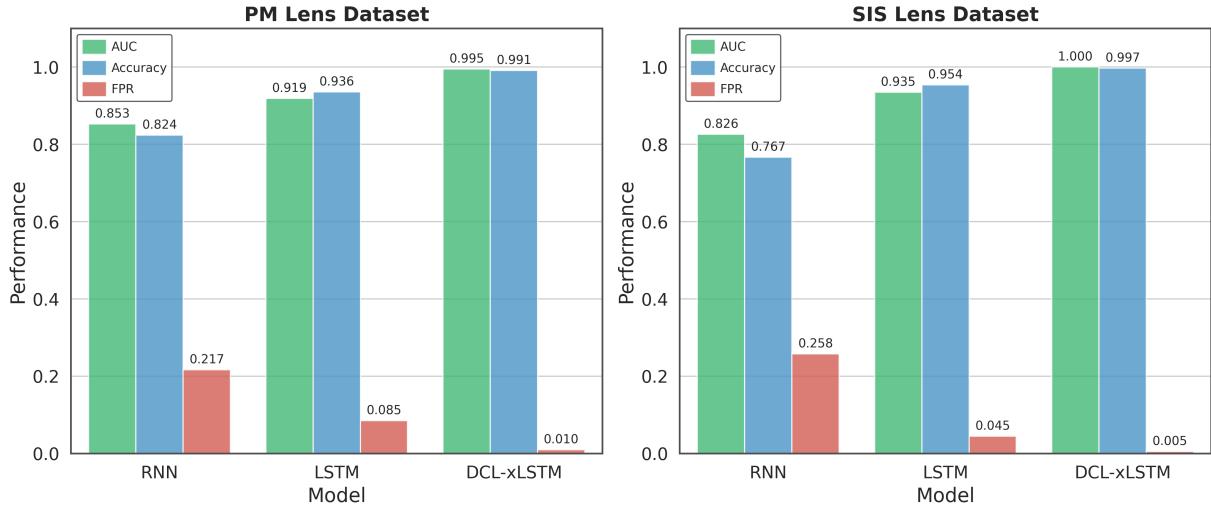


Figure 6. **Performance metrics (AUC, Accuracy, FPR) on different lens models.** **Left:** PM lenses. **Right:** SIS lenses. The DCL-xLSTM model achieves near-perfect AUC and accuracy while maintaining a very low FPR (PM: 0.010, SIS: 0.005), outperforming LSTM and RNN across all metrics.

dataset ($M_L \in [10^7, 10^8]M_\odot$) highlights the regime in which wave-optics effects become pronounced, characterized by distinct modulations of amplitude. In contrast, the lower lens mass dataset ($M_L \in [10^6, 10^7]M_\odot$) addresses the onset of diffraction. In this range, lensing signatures are inherently more subtle, resulting in waveform deviations that are less conspicuous than the higher mass counterparts. The trends depicted in Figure 5 illustrate how the strength of the diffractive features influences the classification efficacy.

In the higher lens mass regime, the pronounced spectral distortions induced by strong wave-optics effects provide clear discriminative features. Consequently, all recurrent architectures operate effectively in this domain, with the RNN and LSTM achieving satisfactory sensitivity. A more revealing divergence appears in the lower mass regime. As the lens mass decreases, the diffractive signatures become inherently more subtle, making them less distinguishable from the detector noise. Under these conditions, the performance of the baseline RNN and LSTM models begins to decrease. In contrast, the DCL-xLSTM model retains stability and maintains a high AUC even when signal deviations are lower. This resilience suggests that the matrix memory structure is particularly effective in capturing fine-grained transfer functions associated with the onset of diffraction, which may be overlooked by scalar-memory architectures.

To ensure the framework’s applicability across different lens types, we extended our evaluation to include datasets generated with both PM and SIS lens profiles. As illustrated in Figure 6, the DCL-xLSTM architecture exhibits remarkable consistency between these varying physical models. Although the standard LSTM maintains competitive AUC scores, it struggles with false positives, exhibiting an FPR approximately $8.5\times$ (PM) and $9\times$

(SIS) higher than that of the DCL-xLSTM. The baseline RNN faces greater challenges in this context, with AUC values dropping below 0.86 and the FPR exceeding 0.21, suggesting that simpler recurrent structures may fail to capture the lensed features from noise in diverse density profiles. The high stability and low false alarm rate of the DCL-xLSTM underscore its potential for GW analysis under varied astrophysical conditions.

C. Detection Sensitivity Across SNR

The sensitivity of a detection algorithm to varying noise conditions is important for practical applications. We evaluated the classification accuracy of the RNN, LSTM, and DCL-xLSTM models at SNR values of 30, 40, 50, 60, and 70. We also included SNR = 20 to assess performance at lower SNR and to analyze the trend of accuracy as the SNR increases.

The results, summarized in Figure 7, show that the classification accuracy of three models improves as the SNR increases. Across the two lens models, the accuracy trends with SNR are consistent and the model ranking is stable. When the lens masses are comparable, differences between lens models appear to be modest, which is plausible for a binary classification task where performance is driven mainly by how clearly the lensing signature emerges above the noise. In the higher lens mass regime, all models achieve high accuracy (> 0.90) at moderate SNR levels (≥ 50). The DCL-xLSTM model saturates this comparable performance level at a lower SNR (~ 30) compared to LSTM and RNN. In the lower lens mass regime, the lensing signature is a weaker spectral distortion, which makes its characteristic features inherently more subtle and less distinguishable from noise. The in-

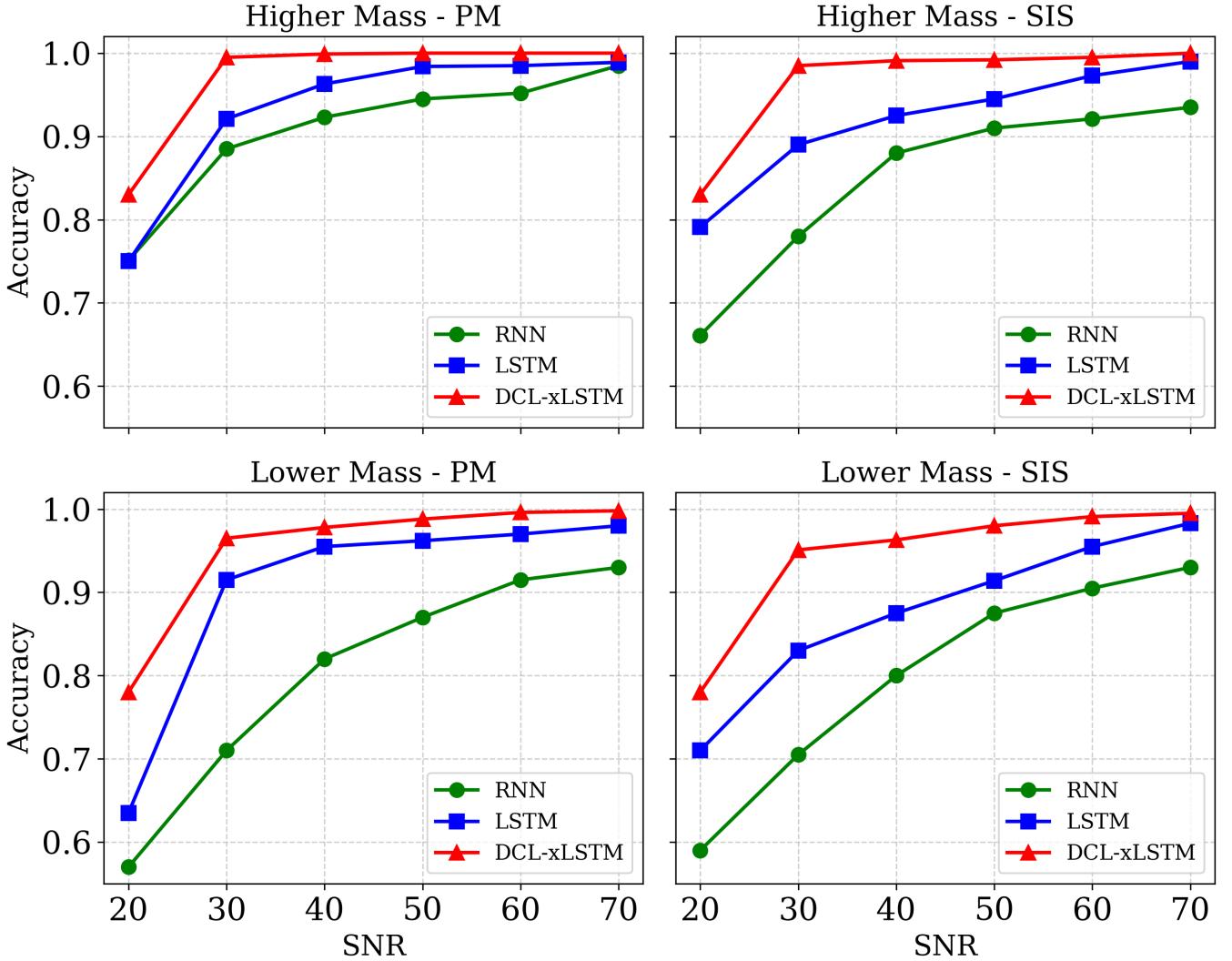


Figure 7. Classification accuracy versus SNR for the RNN (green circles), LSTM (blue squares), and DCL-xLSTM (red triangles) models. The four panels correspond to different combinations of source mass (higher/lower) and lens model (PM/SIS). The DCL-xLSTM model maintains the highest accuracy across all SNR levels and physical scenarios, with the performance advantage being most pronounced at low SNR.

herent difficulty leads to a more significant performance gap between models, especially at lower SNR (20-30). Within this regime, DCL-xLSTM tends to retain a clear accuracy advantage over both LSTM and RNN, with the difference most pronounced at the lowest SNR. This pattern is consistent with improved robustness to noise and a stronger ability to capture distortion-related structure under unfavorable observational conditions.

D. Ablation Study

Given the consistently strong performance of DCL-xLSTM in the previous sections, we further examine the contributions of its core architectural components. The xLSTM-based architecture integrates two types of memory

blocks: sLSTM and mLSTM. To isolate their individual and synergistic effects, we conducted an ablation study comparing three variants: an sLSTM-only model, an mLSTM-only model, and the full hybrid model. The hybrid model, which was selected as the best-performing configuration after an extensive hyperparameter search, was used in all prior experiments.

The results of this ablation study are summarized in Figure 8. All three variants achieve high and comparable accuracy under low-SNR conditions and in wave-optics effects scenarios, indicating that the distinct lensing features can be learned by either architectural paradigm. A significant performance gap emerges in the more challenging regimes, specifically under low-SNR conditions with less wave-optics effects. In these cases, the hybrid model consistently outperforms both single-component

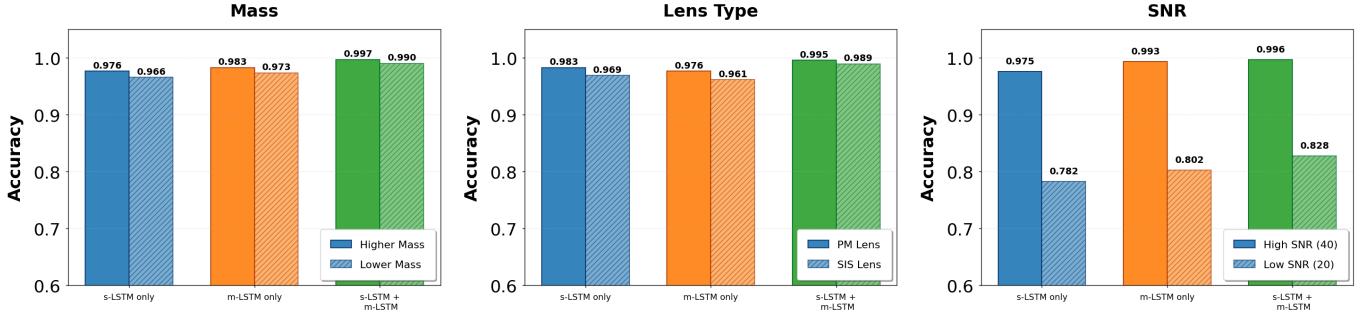


Figure 8. **Performance comparison of DCL-xLSTM architectural variants across different physical conditions.** The hybrid (s+m-LSTM) model leads the performance among all architectural variants, particularly in challenging low-SNR and complex lensing scenarios.

variants. For example, at an SNR of 20, the hybrid model achieves an accuracy approximately 4% higher than the single-component variants. The result provides empirical evidence for functional complementarity between the sLSTM and mLSTM blocks when processing noisy, complex signals.

The sLSTM block excels at modeling fine-grained temporal dependencies through its scalar memory and enhanced gating mechanisms. The mLSTM block captures large-scale global patterns efficiently via its parallelizable matrix-valued memory. The superior performance of the hybrid model under diverse conditions demonstrates its ability to take advantage of these complementary strengths, validating its selection as the optimal architecture for the classification of lensed GW signals.

IV. CONCLUSION AND DISCUSSION

We have developed and validated a deep learning framework for the identification of lensed GW signals in the millihertz band. The proposed DCL-xLSTM architecture consistently outperforms standard recurrent models (LSTM and RNN), achieving high classification performance in challenging scenarios involving mixed datasets. The DCL-xLSTM model maintains robust sensitivity across a range of lens masses and SNR, including regimes near the transition between wave-optics and geometric-optics, where classification tasks becomes more difficult. These results suggest that the architecture may serve as a useful tool for future space-based GW lensing studies.

While the proposed framework already demonstrates strong classification capabilities, there are still many directions that could further enhance its practical efficacy under more realistic conditions. In this work, we have adopted *IMRPhenomD* waveforms for data generation. Moving forward, incorporating additional waveform templates such as those including higher-order modes, spin precession, and EOB-based models may help to broaden physical coverage and improve generalization. Another aspect worth exploring is the inclusion of confusion noise, particularly the unresolved Galactic binaries expected to

dominate the millihertz band. Taking into account this factor during training could help the model better distinguish lensing-induced features from foreground structures. Additionally, while current experiments assume stationary Gaussian noise, incorporating more realistic noise models that reflect mission operations may improve the framework performance and reliability. These directions represent incremental yet valuable extensions to build upon the current work.

In future work, we plan to explore two extensions to broaden the scope of the present study. First, the framework can be extended beyond the single-plane, spherically symmetric lens assumption to incorporate more realistic lens configurations, including multi-plane lensing and composite lenses composed of multiple compact objects and extended mass distributions. Tools such as the GLoW package would support this expansion and allow the classifier to be tested with more scenarios. Second, the transverse motion of the lens and source is an attractive next step. Incorporating the dynamical effects into the training pipeline would broaden the set of measurable lensing observables and enable joint inference of lens parameters and effective transverse velocity. These developments would improve the realism of the simulations and strengthen the detection of lensed signals in millihertz GW observations.

ACKNOWLEDGMENTS

This work was supported by the National Key Research and Development Program of China (Grant Nos. 2021YFC2201901, 2021YFC2203004, 2020YFC2200100, and 2021YFC2201903) and by the International Partnership Program of the Chinese Academy of Sciences (Grant No. 025GJHZ2023106GC). The waveform projection to LISA TDI observables was performed using the *bbhx* package. Numerical computations were performed with the CPU and GPU computing resources provided by the National Microgravity Laboratory (NML), Institute of Mechanics, Chinese Academy of Sciences.

-
- [1] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), Observation of gravitational waves from a binary black hole merger, *Phys. Rev. Lett.* **116**, 061102 (2016).
- [2] R. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration), Gwtc-3: Compact binary coalescences observed by ligo and virgo during the second part of the third observing run, *Phys. Rev. X* **13**, 041039 (2023).
- [3] T. L. S. Collaboration, the Virgo Collaboration, the KAGRA Collaboration, A. G. Abac, I. Abouelfettouh, and et al., *Gwtc-4.0: Population properties of merging compact binaries* (2025), arXiv:2508.18083 [astro-ph.HE].
- [4] T. L. S. Collaboration, the Virgo Collaboration, the KAGRA Collaboration, A. G. Abac, I. Abouelfettouh, and et al., *Gwtc-4.0: Updating the gravitational-wave transient catalog with observations from the first part of the fourth ligo-virgo-kagra observing run* (2025), arXiv:2508.18082 [gr-qc].
- [5] T. L. S. Collaboration, the Virgo Collaboration, the KAGRA Collaboration, A. G. Abac, I. Abouelfettouh, and et al., *Gwtc-4.0: An introduction to version 4.0 of the gravitational-wave transient catalog* (2025), arXiv:2508.18080 [gr-qc].
- [6] R. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration), Tests of general relativity with gwtc-3, arXiv preprint arXiv:2112.06861 (2021), arXiv:2112.06861 [gr-qc].
- [7] R. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration), Population of merging compact binaries inferred using gravitational waves through gwtc-3, *Phys. Rev. X* **13**, 011048 (2023).
- [8] C. S. Kochanek, *The saas fee lectures on strong gravitational lensing* (2004), arXiv:astro-ph/0407232 [astro-ph].
- [9] X.-Y. Lin, X.-J. Wang, H. Zhou, Z. Li, K. Liao, and Z.-H. Zhu, *Constraints on compact dark matter population from micro-lensing effect of gravitational wave for the third-generation gravitational wave detector* (2025), arXiv:2508.13577 [astro-ph.CO].
- [10] B. Liu, Z. Li, and Z.-H. Zhu, Complementary constraints on dark energy equation of state from strongly lensed gravitational wave, *Monthly Notices of the Royal Astronomical Society* **487**, 1980–1985 (2019).
- [11] K. Kim, J. Lee, R. S. H. Yuen, O. A. Hannuksela, and T. G. F. Li, Identification of lensed gravitational waves with deep learning, *Astrophys. J.* **915**, 119 (2021).
- [12] D. Li, T. Liu, A. Liu, C. Wen, J. Wang, K. Liao, and J. Cui, Identification of strongly lensed gravitational wave events using squeeze-and-excitation multi-layer perceptron data-efficient image transformer, arXiv preprint arXiv:2508.19311 (2025), arXiv:2508.19311 [astro-ph.IM].
- [13] J. Janquart, M. Wright, S. Goyal, *et al.*, Follow-up analyses to the o3 ligo-virgo-kagra lensing searches, *Mon. Not. R. Astron. Soc.* **526**, 3832 (2023).
- [14] S. Savastano, *Lensing of Gravitational Waves: Novel Phenomenology and Applications in the Strong and Weak Regimes*, Ph.D. thesis, Humboldt-Universität zu Berlin (2024).
- [15] K. Liao, X.-L. Fan, X. Ding, M. Biesiada, and Z.-H. Zhu, Precision cosmology from future lensed gravitational wave and electromagnetic signals, *Nature Communications* **8**, 10.1038/s41467-017-01152-9 (2017).
- [16] K. Liao, M. Biesiada, and Z.-H. Zhu, Strongly lensed transient sources: A review, *Chinese Physics Letters* **39**, 119801 (2022).
- [17] P. Cremonese, D. F. Mota, and V. Salzano, Characteristic features of gravitational wave lensing as probe of lens mass model (2021), arXiv:2111.01163 [astro-ph.CO].
- [18] P. Amaro-Seoane, H. Audley, S. Babak, and et al., *Laser interferometer space antenna* (2017), arXiv:1702.00786 [astro-ph.IM].
- [19] W.-R. Hu and Y.-L. Wu, The taiji program in space for gravitational wave physics and the nature of gravity, *National Science Review* **4**, 685 (2017), <https://academic.oup.com/nsr/article-pdf/4/5/685/31566708/nwx116.pdf>.
- [20] J. Luo, L.-S. Chen, H.-Z. Duan, Y.-G. Gong, S. Hu, J. Ji, Q. Liu, J. Mei, V. Milyukov, M. Sazhin, C.-G. Shao, V. T. Toth, H.-B. Tu, Y. Wang, Y. Wang, H.-C. Yeh, M.-S. Zhan, Y. Zhang, V. Zharov, and Z.-B. Zhou, Tianqin: a space-borne gravitational wave detector, *Classical and Quantum Gravity* **33**, 035010 (2016).
- [21] G. Cusin and N. Tamanini, Characterization of lensing selection effects for lisa massive black hole binary mergers, *Monthly Notices of the Royal Astronomical Society* **504**, 3610–3618 (2021).
- [22] R. Takahashi and T. Nakamura, Wave effects in gravitational lensing of gravitational waves from chirping binaries, *Astrophys. J.* **595**, 1039 (2003).
- [23] M. Çalışkan, L. Ji, R. Cotesta, E. Berti, M. Kamionkowski, and S. Marsat, Observability of lensing of gravitational waves from massive black hole binaries with lisa, *Physical Review D* **107**, 10.1103/physrevd.107.043029 (2023).
- [24] H. Villarrubia-Rojo, S. Savastano, M. Zumalacárcel, L. Choi, S. Goyal, L. Dai, and G. Tambalo, Glow: novel methods for wave-optics phenomena in gravitational lensing, arXiv preprint arXiv:2409.04606 (2024), arXiv:2409.04606 [gr-qc].
- [25] A. Liu, T. Liu, D. Li, C. Wen, J. Wang, K. Liao, J. Cui, and H. Zhou, Identifying microlensing by compact dark matter through diffraction patterns in gravitational waves with machine learning, arXiv preprint arXiv:2509.04538 (2025), arXiv:2509.04538 [astro-ph.IM].
- [26] Y. Yuan, M. Du, X. yi Lin, P. Xu, and X. Fan, Bayesian analysis of wave-optics gravitationally lensed massive black hole binaries with space-based gravitational wave detector (2025), arXiv:2509.01888 [astro-ph.HE].
- [27] X. Shan, B. Hu, X. Chen, and R.-G. Cai, An interference-based method for the detection of strongly lensed gravitational waves, *Nature Astronomy* **9**, 916–924 (2025).
- [28] Z. Cao, L.-F. Li, and Y. Wang, Gravitational lensing effects on parameter estimation in gravitational wave detection with advanced detectors, *Phys. Rev. D* **90**, 062003 (2014).
- [29] D. Sun and X. Fan, Pattern of lensed chirp gravitational wave signal and its implication on the mass and position of lens (2019), arXiv:1911.08268 [gr-qc].
- [30] A. Chen, P. Cremonese, J. M. Ezquiaga, and D. Keitel, Invariance transformations in wave-optics lensing: implications for gravitational-wave astrophysics and cosmology (2024), arXiv:2408.03856 [astro-ph.CO].

- [31] Z. Gao, K. Liao, L. Yang, and Z.-H. Zhu, Identifying strongly lensed gravitational waves with the third-generation detectors, *Monthly Notices of the Royal Astronomical Society* **526**, 682–690 (2023).
- [32] J. Janquart, O. A. Hannuksela, K. Haris, and C. Van Den Broeck, A fast and precise methodology to search for and analyse strongly lensed gravitational-wave events, *Monthly Notices of the Royal Astronomical Society* **506**, 5430–5438 (2021).
- [33] X.-y. Lin, J.-d. Zhang, L. Dai, S.-J. Huang, and J. Mei, Detecting strong gravitational lensing of gravitational waves with tianqin, *Physical Review D* **108**, 10.1103/physrevd.108.064020 (2023).
- [34] S. Magare, A. More, and S. Choudary, Slick: Strong lensing identification of candidates kindred in gravitational wave data, arXiv preprint [arXiv:2403.02994](#) (2024), [arXiv:2403.02994 \[astro-ph.HE\]](#).
- [35] S. Goyal, K. Haris, S. J. Kapadia, and P. Ajith, Rapid identification of strongly lensed gravitational-wave events with machine learning, *Phys. Rev. D* **104**, 124057 (2021).
- [36] M. Motuz, K. Jani, W. Smith, and K. Ruiz-Rocha, Predicting and Mapping Binary Black Hole Mergers for Current and Next-Gen Gravitational Wave Detectors, *Bulletin of the AAS* **57** (2025), <https://baas.aas.org/pub/2025n2i303p07>.
- [37] R. W. Kiendrebeogo, A. M. Farah, E. M. Foley, Gray, and et al., Updated observing scenarios and multimessenger implications for the international gravitational-wave networks o4 and o5, *The Astrophysical Journal* **958**, 158 (2023).
- [38] M. Çalışkan, J. M. Ezquiaga, O. A. Hannuksela, and D. E. Holz, Lensing or luck? false alarm probabilities for gravitational lensing of gravitational waves, *Phys. Rev. D* **107**, 063023 (2023).
- [39] H. Zhou, Z. Li, K. Liao, and Z. Huang, Constraints on compact dark matter from lensing of gravitational waves for the third-generation gravitational wave detector, *Monthly Notices of the Royal Astronomical Society* **518**, 149–156 (2022).
- [40] Y. Lecun, Y. Bengio, and G. Hinton, Deep learning, *Nature* **521**, 436 (2015).
- [41] *Frontiers of Physics* **20**, 45301 (2025).
- [42] A. Singh, L. Li, O. Hannuksela, T. G. F. Li, and K. Kim, Discrimination of lensed gravitational waves using deep learning, *Am. J. Undergrad. Res.* **16**, 5 (2019).
- [43] E. Cuoco, J. Powell, and et al., Enhancing gravitational-wave science with machine learning, *Machine Learning: Science and Technology* **2**, 011002 (2020).
- [44] K. Simonyan and A. Zisserman, Very deep convolutional networks for large-scale image recognition (2015), [arXiv:1409.1556 \[cs.CV\]](#).
- [45] O. Bulashenko and H. Ubach, Lensing of gravitational waves: universal signatures in the beating pattern, *Journal of Cosmology and Astroparticle Physics* **2022** (07), 022.
- [46] S. Hou, X.-L. Fan, K. Liao, and Z.-H. Zhu, Gravitational wave interference via gravitational lensing: Measurements of luminosity distance, lens mass, and cosmological parameters, *Physical Review D* **101**, 10.1103/physrevd.101.064011 (2020).
- [47] S. Hou, P. Li, H. Yu, M. Biesiada, X.-L. Fan, S. Kawamura, and Z.-H. Zhu, Lensing rates of gravitational wave signals displaying beat patterns detectable by decigo and b-decigo, *Physical Review D* **103**, 10.1103/physrevd.103.044005 (2021).
- [48] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, Densely connected convolutional networks (2018), [arXiv:1608.06993 \[cs.CV\]](#).
- [49] T. Chen and C. Guestrin, Xgboost: A scalable tree boosting system, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, 2016) p. 785–794.
- [50] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, An image is worth 16x16 words: Transformers for image recognition at scale (2021), [arXiv:2010.11929 \[cs.CV\]](#).
- [51] N. Matsunaga and K. Yamamoto, The finite source size effect and wave optics in gravitational lensing, *Journal of Cosmology and Astroparticle Physics* **2006** (01), 023–023.
- [52] M. Hilker, H. Baumgardt, L. Infante, M. Drinkwater, E. Evstigneeva, and M. Gregg, Dynamical masses of ultra-compact dwarf galaxies in fornax, *Astronomy & Astrophysics* **463**, 119–130 (2006).
- [53] M. Hilker and S. Mieske, The properties of ultra-compact dwarf galaxies and their possible origin, [arXiv: Astrophysics \(2004\)](#).
- [54] D. Levin, Fast integration of rapidly oscillatory functions, *J. Comput. Appl. Math.* **67**, 95 (1996).
- [55] S. Marsat and J. G. Baker, Fourier-domain modulations and delays of gravitational-wave signals (2018), [arXiv:1806.10734 \[gr-qc\]](#).
- [56] S. Marsat, J. G. Baker, and T. D. Canton, Exploring the bayesian parameter estimation of binary black holes with lisa, *Physical Review D* **103**, 10.1103/physrevd.103.083011 (2021).
- [57] S. Husa, S. Khan, M. Hannam, M. Pürrer, F. Ohme, X. J. Forteza, and A. Bohé, Frequency-domain gravitational waves from nonprecessing black-hole binaries. i. new numerical waveforms and anatomy of the signal, *Physical Review D* **93**, 10.1103/physrevd.93.044006 (2016).
- [58] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. J. Forteza, and A. Bohé, Frequency-domain gravitational waves from nonprecessing black-hole binaries. ii. a phenomenological model for the advanced detector era, *Physical Review D* **93**, 10.1103/physrevd.93.044007 (2016).
- [59] M. L. Katz, S. Marsat, A. J. Chua, S. Babak, and S. L. Larson, Gpu-accelerated massive black hole binary parameter estimation with lisa, *Physical Review D* **102**, 10.1103/physrevd.102.023033 (2020).
- [60] M. L. Katz, Fully automated end-to-end pipeline for massive black hole binary signal extraction from lisa data, *Physical Review D* **105**, 10.1103/physrevd.105.044055 (2022).
- [61] T. A. Prince, M. Tinto, S. L. Larson, and J. W. Armstrong, Lisa optimal sensitivity, *Phys. Rev. D* **66**, 122002 (2002).
- [62] O. Burke, S. Marsat, J. R. Gair, and M. L. Katz, Mind the gap: addressing data gaps and assessing noise mis-modeling in lisa (2025), [arXiv:2502.17426 \[gr-qc\]](#).
- [63] S. Hochreiter and J. Schmidhuber, Long short-term memory, *Neural Computation* **9**, 1735 (1997).
- [64] M. Beck, K. Pöppel, M. Spanring, A. Auer, O. Prudnikova, M. Kopp, G. Klambauer, J. Brandstetter, and S. Hochreiter, xlstm: Extended long short-term memory (2024), [arXiv:2405.04517 \[cs.LG\]](#).