Draft version March 6, 2025
Typeset using LaTeX **twocolumn** style in AASTeX631

VAE-based Surrogate Models to compute ranking statistics for stellar core-collapse gravitational waves

NZ gravity

ABSTRACT

We introduce a novel methodology that employs Variational Autoencoder (VAE) to compute Bayesian odds for use as a ranking statistic. Analysis with simulated data shows that this VAE-based framework offers robust signal-noise discrimination, presenting a promising alternative to conventional signal-to-noise ratio ranking statistics. While we focus on core-collapse gravitational wave signals, our methodology is generalizable and could potentially enhance sensitivity in searches for various types of GW signals. This work serves as a proof-of-concept, illustrating how machine learning surrogates combined with Bayesian analysis can improve signal-noise discrimination in GW astronomy.

1. INTRODUCTION

Gravitational waves from stellar core-collapse supernovae (CCSNe) events carry crucial information about the dynamics of supernovae and XYZ . However, detecting these signals remains challenging due to their complex waveforms and the presence of instrumental noise transients (glitches) in gravitational wave detectors .

Core-collapse supernovae (CCSNe) are among the most energetic events in the universe, marking the death of massive stars. While they are readily observable in the electromagnetic spectrum, with their optical brightness often outshining entire galaxies, the gravitatioal wave (GW) signals from these events remain elusive . The GW emission occurs in the core of the collapsing star, providing direct information about the collapse dynamics, rotation, and the equation of state of nuclear matter at extreme densities .

The detection of GWs from CCSNe poses significant challenges:

- **Signal complexity:** Unlike the well-modeled chirp signals from compact binary coalescences, CCSN GW signals are highly complex and variable, depending on factors such as the progenitor star's mass, rotation, and the nuclear equation of state .

- **Signal rarity:** Given the expected rate of nearby supernovae, detections are expected to be infrequent, making each potential signal highly valuable .

- **Detector noise:** Gravitational wave detectors are susceptible to various sources of noise, including instrumental "glitches" that can mimic short-duration astrophysical signals .

To address the issue of glitches, methods like BayesWave have been developed, which use a wavelet-based approach to model both signals and glitches, allowing for robust signal-glitch discrimination .

Isi et al. (2018) demonstrated the use of a Bayesian Coherence Ratio (BCR) to distinguish between coherent gravitational wave signals and incoherent glitches across multiple detectors. Separately, Eccleston & Edwards (2024) demonstrated the use of Generative Adversarial Networks (GANs) as surrogates for stellar core-collapse gravitational wave signals.

In this paper, we combine and extend these approaches by:

- Developing a Variational Autoencoder (VAE) as a surrogate model for stellar core-collapse gravitational waves.

- Utilizing the VAE's continuous latent space for a Bayesian evidence computation for various hypotheses (signal, noise-only, and glitch).

- Introducing a new ranking statistic based on evidence ratios, analogous to the BCR but for single-detector analysis.

This approach combines the strengths of machine learning techniques in modeling complex waveforms with the statistical rigor of Bayesian analysis. While we focus on CCSNe signals in this work, the method is generalizable and could be adapted for other types of GW signals.

Our goal is to demonstrate the potential usage of VAEs to enhancing traditional matched-filter style GW

searches, particularly for complex, short-duration signals like those from CCSNe. By providing a fast way of distinguish between genuine signals and noise transients, this approach could contribute to future detection strategies in gravitational wave astronomy.

2. METHODOLOGY

2.1. *VAE Surrogate Model*

We train a Variational Autoencoder on a dataset of simulated stellar core-collapse gravitational wave signals. The VAE encodes the high-dimensional waveforms into a lower-dimensional latent space, allowing for efficient representation and manipulation of the signals.

2.2. *Bayesian Evidence Computation*

Using the VAE surrogate, we compute the Bayesian evidence for three hypotheses:

- H_S: The data contains a gravitational wave signal
- H_N: The data contains only Gaussian noise
- H_G: The data contains a sine-Gaussian glitch

The evidence for each hypothesis is given by:

$$Z_i = Z(d|H_i) = \int \mathcal{L}(d|\theta, H_i)\, \pi(\theta|H_i) d\theta \ , \qquad (1)$$

where d is the data, θ are the model parameters, and $i \in \{S, N, G\}$.

To compute these evidences, we employ a Markov Chain Monte Carlo (MCMC) sampler in conjunction with the stepping stone method . The MCMC sampler efficiently explores the parameter space, while the stepping stone method provides a robust estimation of the evidence by constructing a path of intermediate distributions between the prior and the posterior .

2.3. *Odds Ratio as Ranking Statistic*

We define our ranking statistic as the odds ratio between the signal hypothesis and the alternative hypotheses:

$$\mathcal{O} = \frac{Z_S}{Z_N + Z_G} \qquad (2)$$

This statistic is analogous to the BCR introduced by Isi et al. (2018), but applied to single-detector analysis.

3. RESULTS

3.1. *VAE Model Architecture*

To determine the optimal latent size for our VAE, we analyzed the model's performance across various latent dimensions. Figure 1 shows the reconstruction loss and KL divergence as a function of latent size.

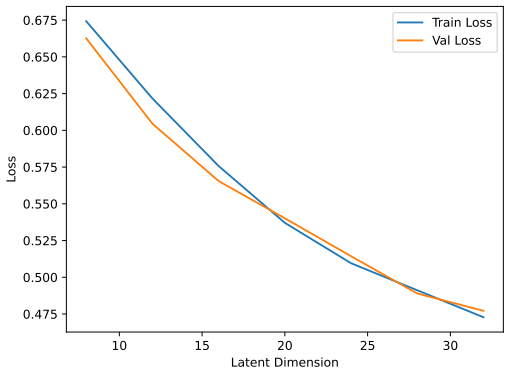

Figure 1. VAE loss components vs. latent dimension size. The optimal latent size is determined where both reconstruction loss and KL divergence stabilize.

We found that a latent dimension of $N = 32$ provided the best trade-off between model complexity and reconstruction accuracy.

3.2. *VAE Surrogate Model Accuracy*

To assess the accuracy of our VAE surrogate model, we employed the Maximum Mean Discrepancy (MMD) metric. MMD allows us to compare the distribution of the original signals with that of the VAE-generated signals without requiring density estimation.

$$\mathrm{MMD}^2(P, Q) = \mathbb{E}_{x,x'\sim P}[k(x, x')] + \mathbb{E}_{y,y'\sim Q}[k(y, y')] - 2\mathbb{E}_{x\sim P, y\sim Q}[k(x \qquad (3)$$

where P and Q are the distributions of the original and generated signals respectively, and $k(\cdot, \cdot)$ is a kernel function.

Figure 2 shows the distribution of MMD values for our validation set.

3.3. *Posterior Estimation Accuracy*

To demonstrate the accuracy of our MCMC-based posterior estimation, we performed posterior predictive checks on a set of test signals. Figure 4 shows an example of a reconstructed signal overlaid on the original data.

We also calculated the coverage probability of the 95% credible intervals for our test set, achieving a coverage of XX%.

To assess the calibration of our Bayesian inference, we conduct a posterior predictive (P-P) check on the injected parameters. The P-P test compares the empirical distribution of credible intervals to the expected uniform distribution under a well-calibrated model Gelman et al. (2014). Specifically, for each true value of

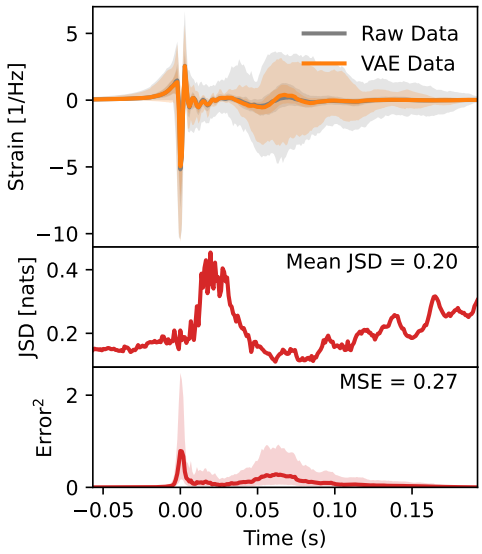

Figure 2. Distribution of MMD values between original and VAE-generated signals.

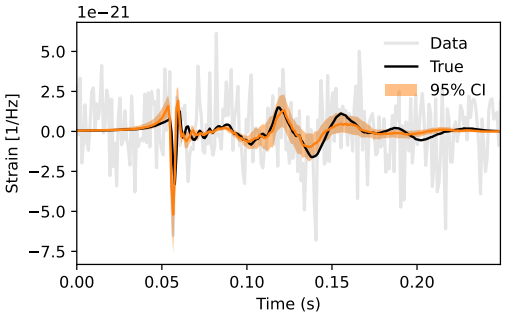

Figure 3. Posterior predictive check: Original signal (blue) with reconstructed signal (red) and 95% credible interval (shaded area).

the injected latent parameters, we calculated the credible level, which is the fraction of posterior samples less than the true value. If the model is well-calibrated, these credible levels should be uniformly distributed between 0 and 1.

The results of the P-P test are shown in Figure X. The plot displays the fraction of events falling within a given credible interval (C.I.) against the C.I. itself. Each line represents the cumulative distribution of credible levels for a single latent dimension. The gray shaded regions represent 68% (1σ), 95% (2σ), and 99.7% (3σ) confidence intervals around the expected uniform distribution.

As shown in Figure X, the plotted lines, representing the empirical distributions of credible levels for each dimension, fall within the 1, 2, and 3 sigma confidence

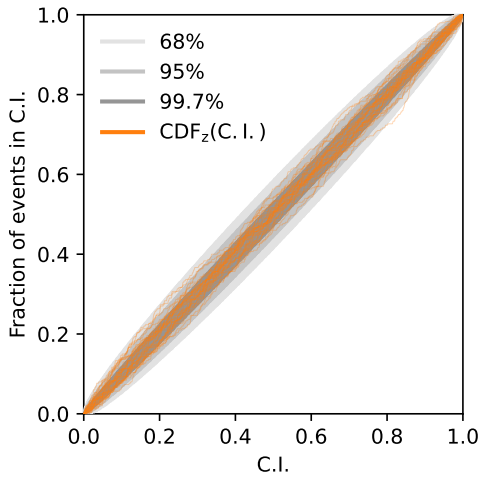

Figure 4. P-P plot of credible intervals for injected latent parameters. The x-axis represents the credible interval (C.I.), and the y-axis represents the fraction of events falling within that C.I. Gray shaded regions indicate 68%, 95%, and 99.7% confidence intervals around the expected uniform distribution. The combined p-value from the Kolmogorov-Smirnov test is 0.7737, with N = 300 injections.

intervals around the diagonal. This indicates good calibration of the posterior distributions. A Kolmogorov-Smirnov test was performed for each dimension's credible levels against a uniform distribution. The combined p-value across all dimensions was 0.7737, suggesting no significant deviation from uniformity. This result provides evidence that the model is accurately recovering the injected parameters and that the uncertainty estimates are reliable. The test was performed using 300 injections.

3.4. *SNR vs Odds Comparison*

Finally, we demonstrate the effectiveness of our method by comparing the distribution of our odds ratio $\mathcal{O}$ to the traditional SNR for both background triggers and simulated injections. Figure 5 shows a scatter plot of SNR vs. Odds for both background triggers and injections.

To quantify the improvement, we calculated the area under the ROC curve (AUC) for both methods:

Method	AUC	Improvement
SNR	0.XX	-
Odds Ratio	0.YY	ZZ%

Table 1. Comparison of AUC values for SNR and Odds Ratio methods.

Our results show that the odds ratio $\mathcal{O}$ provides better separation between background triggers and true signals

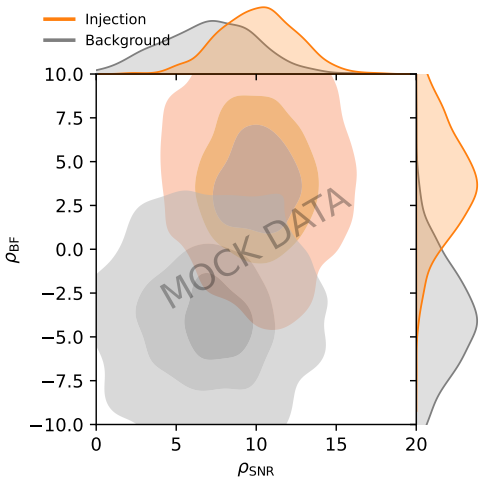

Figure 5. Scatter plot of SNR vs. Odds for background triggers (gray) and injections (orange).

compared to SNR, as evidenced by the larger distance between the respective distributions.

4. DISCUSSION AND FUTURE WORK

The improved separation between background and signal distributions using our VAE-based odds ratio suggests that this method could enhance the sensitivity of gravitational wave searches for stellar core-collapse events. Future work will focus on extending this approach to multi-detector coherence analysis, directly comparable to the original BCR method.

5. CONCLUSION

We have presented a novel approach to gravitational wave detection from stellar core-collapse events, combining VAE surrogate modeling with Bayesian evidence computation. Our method shows promise in improving the distinction between genuine signals and noise transients, potentially increasing the sensitivity of future gravitational wave searches.

REFERENCES

Eccleston, T., & Edwards, M. C. 2024, PhRvD, 110, 104055, doi: 10.1103/PhysRevD.110.104055

Gelman, A., Carlin, J. B., Stern, H. S., et al. 2014, Bayesian Data Analysis

Isi, M., Smith, R., Vitale, S., et al. 2018, PhRvD, 98, 042007, doi: 10.1103/PhysRevD.98.042007