

Detecting the third family of compact stars with normalizing flows

Valéria Carvalho^{1,*}, Márcio Ferreira^{1,†}, Constança Providência^{1,‡} and Michał Bejger^{2,3,§}

¹*CFisUC, Department of Physics, University of Coimbra, P-3004–516 Coimbra, Portugal*

²*INFN Sezione di Ferrara, Via Saragat 1, 44122 Ferrara, Italy*

³*Nicolaus Copernicus Astronomical Center, Polish Academy of Sciences, Bartycka 18, 00-716 Warsaw, Poland*



(Received 12 March 2024; accepted 23 April 2024; published 20 May 2024)

We explore the anomaly detection framework based on the normalizing flow (NF) models introduced in Morawski and Bejger [Phys. Rev. C **106**, 065802 (2022)] to detect the presence of a large (destabilizing) dense matter phase transition in neutron star (NS) observations of masses and radii, and relate the feasibility of detection with parameters of the underlying mass-radius sequence, which is a functional of the dense matter equation of state. Once trained on simulated data featuring continuous $M(R)$ solutions (i.e., no phase transitions), NF is used to determine the likelihood of a first-order phase transition in a given set of $M(R)$ observations featuring a discontinuity, i.e., to perform the anomaly detection. Different mock test sets, featuring two branch solutions in the $M(R)$ diagram, were parametrized by the NS mass at which the phase transition occurs, M_c , and the radius difference between the heaviest hadronic star and lightest hybrid star, ΔR . We analyze the impact of these parameters on the NF performance in detecting the presence of a first-order phase transition. Among the results, we report that given a set of 15 stars with radius uncertainty of 0.2 km, a detection of a two-branch solution is possible with 95% accuracy if $\Delta R > 0.4$ km.

DOI: [10.1103/PhysRevD.109.103032](https://doi.org/10.1103/PhysRevD.109.103032)

I. INTRODUCTION

Discovering the properties of the dense nuclear matter realized inside neutron stars (NS) is yet a fundamental unsolved question in nuclear physics. The present constraints on the high density nuclear matter equation of state (EOS) are originated from NS observations. The existence of hybrid stars, where a phase transition from nuclear matter to deconfined quark matter in the star's core occurs, is yet to be validated. The nature of the hypothetical phase transition depends on the surface tension between both phases, which could be either a sharp (Maxwell construction) for large values, or a smooth mixed phase transition for low values [1,2]. Assuming a large surface tension value, the transition between nuclear and quark phases is thus realized via a first-order phase transition. For some hybrid stars models, depending on the energy density discontinuity between both phases, a disconnected branch may appear in the $M(R)$ diagram [1–6]. The value of the energy density discontinuity dictates the stability of the hybrid star: for low values, the quark core is able to counteract the gravitational pressure from the nuclear mantle; otherwise, a too strong energy density discontinuity

destabilizes the hybrid star. The present work aims to analyze the detectability of such first-order phase transitions, specifically when the stable branch of hybrid stars is separated from the hadronic branch, and to assess the criteria for which the detectability statement is credible, such as number of observations, observational errors, phase transition parameters, by investigating the qualitative features induced in the mass-radius diagram.

Generative machine learning models are promising tools in tackling several challenges in high energy physics (see Ref. [7]), e.g., in anomaly detection tasks in particle physics problems [8]. Among the different generative models, normalizing flow (NF) models [9,10] provide an analytic description of the underlying data distribution, simultaneously enabling efficient sampling and probability density estimation. In astrophysics, NFs have been employed for rapid gravitational-wave parameter estimation parameters from gravitational-wave data [11,12], showing the remarkable versatility of this probabilistic generative model. The recent study [13] combines NF with Hamiltonian Monte Carlo methods to infer the complete posterior distribution of the EOS and nuisance parameters directly from telescope observations. Additionally, a novel framework for detecting possible phase transitions based on NS mass-radius observations was introduced in [14]. The authors employed a flow-based model for anomaly detection, specifically testing NF models on multiple samples of six EOSs exhibiting different phase transitions.

*val.mar.dinis@uc.pt

†marcio.ferreira@uc.pt

‡cp@uc.pt

§bejger@camk.edu.pl

The evaluation involved the analyses of the latent space to determine if the model successfully detected anomalies.

The present work extends the NF methodology introduced in [14], by applying it to generic $M(R)$ diagrams instead of specific hybrid EOSs. The generic $M(R)$ diagrams are constructed to simulate phase transitions at a specific NS mass and with a given phase radius difference between the heaviest hadronic star and lightest hybrid star. Furthermore, we use directly the NF ability to perform probability density estimation on input vectors, determining whether anomalies are present in the input vectors.

The structure of this paper unfolds as follows: In Sec. II, we provide an introduction to NF. Section III introduces the dataset used, the underlying theory, and the process of implementing mock observations. Section IV scrutinizes the obtained results, elucidating the metrics and datasets employed. Finally, Sec. V summarizes our findings and concludes the study.

II. NORMALIZING FLOWS

NFs, extensively reviewed in [15], constitute a probabilistic generative model based on invertible transformations, which has been explored in physics, e.g., [11,12,14]. Their goal is to model the intricate data probability distribution $p_x(\mathbf{x})$ by transforming random variables \mathbf{z} from a simple base distribution $p_z(\mathbf{z})$ —typically a multivariate standard normal distribution—through a nonlinear yet invertible and differentiable bijective transformation $\mathbf{f}_\theta: \mathbb{R}^N \rightarrow \mathbb{R}^N$. These transformations are defined by neural networks (see Ref. [7] for a comprehensive review with applications in physics) and parametrized by θ . The transformation can be expressed as

$$\mathbf{x} = \mathbf{f}_\theta(\mathbf{z}), \quad \text{where } \mathbf{z} \sim p_z(\mathbf{z}). \quad (1)$$

Applying the change of variables formula from probability theory allows us to compute the probability distribution:

$$p_x(\mathbf{x}) = p_z(\mathbf{z}) \left| \det \frac{\partial \mathbf{f}_\theta(\mathbf{z})}{\partial \mathbf{z}} \right|^{-1} = p_z(\mathbf{f}_\theta^{-1}(\mathbf{x})) \left| \det \frac{\partial \mathbf{f}_\theta^{-1}(\mathbf{x})}{\partial \mathbf{x}} \right|. \quad (2)$$

The second term represents the absolute value of the Jacobian determinant, where $\partial \mathbf{f}_\theta^{-1}(\mathbf{x})/\partial \mathbf{x}$ is an $N \times N$ matrix. This determinant captures the change in volume resulting from the transformation applied to the probability space. The transformation itself serves to morph the base distribution into the target distribution.

The mapping \mathbf{f}_θ can be expressed as a composition of invertible functions: $\mathbf{f}_\theta = \mathbf{f}_{\theta_K} \circ \dots \circ \mathbf{f}_{\theta_2} \circ \mathbf{f}_{\theta_1}$. Consequently, the target variable is obtained through the composition

$$\mathbf{x} = \mathbf{f}_\theta(\mathbf{z}) = \mathbf{f}_{\theta_K} \circ \dots \circ \mathbf{f}_{\theta_2} \circ \mathbf{f}_{\theta_1}(\mathbf{z}). \quad (3)$$

The relation between \mathbf{x} and \mathbf{z} is then $\mathbf{z} \xrightarrow{\mathbf{f}_{\theta_1}} h_1 \xrightarrow{\mathbf{f}_{\theta_2}} \dots \xrightarrow{\mathbf{f}_{\theta_{K-1}}} h_{K-1} \xrightarrow{\mathbf{f}_{\theta_K}} \mathbf{x}$, where $h_0 = \mathbf{z}$ and $h_K = \mathbf{x}$. The schematic representation of a flow-based model is in Fig. 1. The logarithm of the probability distribution can then be defined as

$$\log p_x(\mathbf{x}) = \log p_z(\mathbf{f}_{\theta_K}^{-1}(h_{K-1})) + \sum_{k=1}^K \log \left| \det \frac{\partial \mathbf{f}_{\theta_k}^{-1}(h_{k-1})}{\partial h_{k-1}} \right|. \quad (4)$$

The term $\log p_z(\mathbf{f}_{\theta_K}^{-1}(h_{K-1}))$ captures the logarithm of the base distribution's probability, while the summation over the K terms represents the cumulative contribution of Jacobian determinants from each invertible function in the composition. The overall likelihood of the data under the NF model is tractable and leads to the training objective (loss function) being minimizing:

$$\mathcal{L}(\mathcal{D}) = -\frac{1}{|\mathcal{D}|} \sum_{\mathbf{x} \in \mathcal{D}} \log p_x(\mathbf{x}), \quad (5)$$

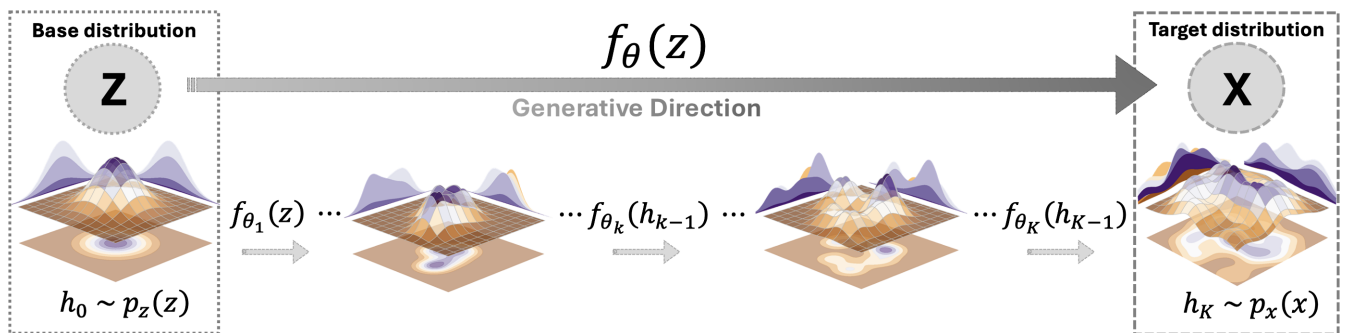


FIG. 1. Schematic representation of a flow-based model illustrating the application of a transformation $\mathbf{f}_\theta(\mathbf{z})$ across K coupling layers, denoted as $\mathbf{f}_{\theta_K} \circ \dots \circ \mathbf{f}_{\theta_2} \circ \mathbf{f}_{\theta_1}(h_0)$. This process operates on samples from a base distribution $p_z(\mathbf{z})$, yielding samples of the target distribution $p_x(\mathbf{x})$ in the generative direction. The opposite direction is known as the normalizing direction $\mathbf{f}_\theta^{-1}(\mathbf{x})$. The dataset elements $x_i \sim X$ are composed of N mass-radius observations, $x_i = [(M_1, R_1), (M_2, R_2), \dots, (M_N, R_N)]$; see Sec. III for details.

which is the negative log-likelihood (NLL) for the training dataset \mathcal{D} .

While various implementations of NF exist, we select the coupling neural spline flow [16] for the present work. In the context of coupling transforms, the input variable is divided into two parts, $[x_{1:k-1}, x_{k:K}]$, where \mathbf{f}_θ is applied to the second part, resulting in an output of $[x_{1:k-1}, \mathbf{f}_\theta(x_{k-1:K})]$. Values are shuffled between each transformation using a permutation layer. For the actual transformation, we utilized a rational-quadratic spline function. For the actual implementation, we have used the PyTorch [17] supplemented with the NFLOWS library [18].

III. DATASET

The possible existence of a first-order phase transition between nuclear and quark matter in NS remains an open question. Owing to the uncertainty on the value of the surface tension at the interface, the phase transition between two pure phases, say nucleonic and quark, could be either sharp (so-called Maxwell construction, featuring a large discontinuity in density profile) or a smooth one (via the “mixed phase”, so-called Gibbs construction) [19]. Assuming a high surface tension scenario, and thus a sharp phase transition, the appearance of a disconnected branch of hybrid stars depends on the energy density discontinuity value $\Delta\epsilon$ at the transition [1].

The present work investigates the hypothesis of the existence of hybrid star branches, and aims to analyze their detectability from a set of NS observations. Since we are only interested in accessing the degree of detectability of such branches regardless of the details of both nuclear and quark matter phases, we constructed a mock dataset that simulates a continuity of solutions. We used, as a starting point, a set of hadronic EOSs and emulate the appearance of a second branch solution in the $M(R)$ diagram as illustrated in Fig. 2. We selected a NS mass M_c and shifted the upper section by ΔR , which simulates the difference in radius between the hadronic star and a hybrid star (with quark matter core). This construction allows one to analyze the impact of $(M_c, \Delta R)$ in the likelihood of observing such scenarios given a set of $M(R)$ observations. Notice that this hybrid star mass-radius curve construction may be considered conservative because in a realistic description of a quark branch it is expected that (i) the second stable branch may start at $M < M_c$ as in [4] and (ii) the slope of the mass-radius of the second branch will probably be quite different from the one of the hadronic mass-radius behavior. We will test our model with a mass-radius curve obtained from microscopic models.

The key insight of using NF in the present task is the following: once NF models are trained with samples from $M(R)$ diagrams originated by hadronic EOSs (red dots in Fig. 2), we are capable of estimating the likelihood of any other $M(R)$ diagrams, such as the sample from a two-branch solution (black triangles in Fig. 2)—a well trained

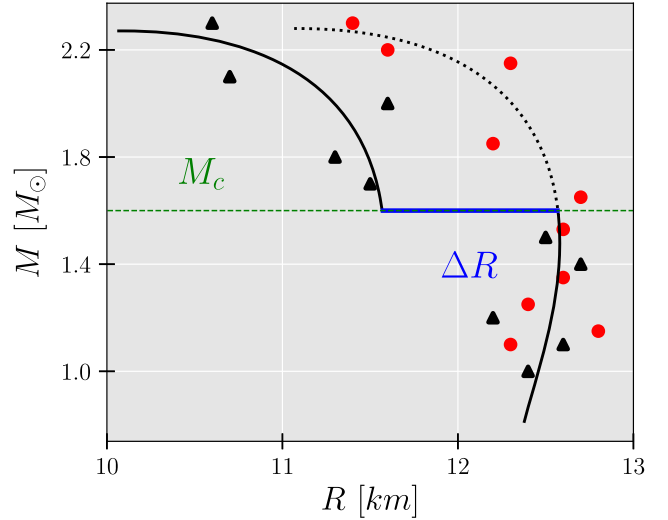


FIG. 2. To simulate the existence of two branches in a given TOV solution, we have defined a critical mass M_c , and the upper part solution of a hadronic EOS was shifted by ΔR . To illustrate the sampling procedure described in Sec. III A, we display a sample from the purely hadronic EOS (red dots) and from the hybrid $M(R)$ solution (black triangles).

model attributes low likelihood to any samples that deviate considerably from the statistical properties of the training set.

The base set of hadronic EOSs consists of 25 287 nuclear models based on a relativistic mean field description, which we divide in 80% train and 20% test, where minimal constraints were imposed [20]: several nuclear saturation properties, the existence of a $2M_\odot$ NS, and a consistent low-density pure neutron matter with N³LO calculations in chiral effective field theory. The specific dataset used in our study corresponds to Set 0 from the article [20], and the dataset has been made available in [21] (Fig. 14 in [20] shows the full $M(R)$ posterior of this dataset).

A. Observation mock data

The generation of the mock datasets used to train the NF models closely follows our previous works [22,23]. We sample N mass values, M_n^0/M_\odot , from $\mathcal{U}(1, M_{\max})$, where $n = 1, \dots, N$, since no NS for masses below $1M_\odot$ are expected from stellar evolution. Then, we determine both $\sigma_{n,M}$ and $\sigma_{n,R}$ from $\mathcal{U}(0, \sigma_M)$ and $\mathcal{U}(0, \sigma_R)$, respectively. The final $\{M_n, R_n\}$ observations are obtained from $R_n \sim \mathcal{N}(R(M_n^0), \sigma_{n,R}^2)$ and $M_n \sim \mathcal{N}(M_n^0, \sigma_{n,M}^2)$. Basically, we are creating a mock observation set with N stars whose observations are Gaussian scattered around the true Tolman-Oppenheimer-Volkoff (TOV) solution, with specific standard deviations for each star, $\sigma_{n,M}$ and $\sigma_{n,R}$. We generate $n_s = 120$ mock observation sets by repeating the described procedure 120 times for each EOS.

In total, we have generated three different datasets, the properties of which are detailed in Table I. While radius

TABLE I. Generation parameters for each dataset.

Dataset	$\sigma_M [M_\odot]$	σ_R [km]
Set 0	0	0
Set 1	0.1	0.2
Set 2	0.136	0.626

uncertainties of ~ 0.1 km or less are expected from future observations (e.g., Einstein Telescope [24], Cosmic Explorer [25]), we employed more conservative uncertainty parameters for Set 1: $\sigma_R = 0.2$ km and $\sigma_M = 0.1M_\odot$. Set 2 was derived using the methodology described in our prior article [23]. For each set, we trained two models with $N = 5$ and 15 to investigate how the number of stars affects the model performance in detecting a second branch in the TOV solution.

Let us stress that the NF models were trained in datasets in which only a single hadronic branch exists, i.e., they were trained with $\Delta R = 0$. After training, we applied these NF models in test datasets specified by $(M_c, \Delta R)$ in order to access the model performance in detecting the existence of a second branch. Contrarily to $n_s = 120$ mock observations used in the training sets, all test sets are composed of $n_s = 1$ mock observations to mimic a real-world scenario where we have access to a single mock observation of the “true” EOS.

B. Training procedure

The training process was conducted for three distinct datasets (see Table I) and two different input sizes, $N = 5$ and 15, as elaborated in Sec. III A, resulting in the training of six distinct models. Throughout the training, the dataset for training was partitioned into 80% for actual training and 20% for validation. A test on data normalization revealed superior results with standardization than with the method employed in the work [26] for our case. Our chosen architecture is a concatenation of blocks featuring an invertible linear transformation using lower–upper (LU) decomposition [27] alongside a rational quadratic spline transform [16] employing eight bins. To obtain the parameters of the spline and the respective \mathbf{f}_θ , we utilized a multilayer perceptron. Various configurations were explored, including changes in the number of hidden layers, neurons per layer, and activation functions. Optimal results were achieved with three hidden layers, each comprising 15 neurons, and the Softplus function as the activation function. For input instances with five stars ($N = 5$), i.e., five $\{M_i, R_i\}$ pairs, seven coupling transformations were implemented, while for 15 stars ($N = 15$), i.e., 15 $\{M_i, R_i\}$ pairs, we utilized ten. The training setup involved a learning rate of 0.001, utilizing the ADAM optimizer [28] with the AMSgrad improvement [29]. The models undergo training for 500 epochs with a batch size of 1024.

IV. RESULTS

After training the NF with NS data with no phase transitions, i.e., with purely hadronic branches, we assess the models’ performance in several mock datasets featuring varying degrees of phase transitions, as detailed in Sec. III. Specifically, we conducted tests with three critical masses (M_c) representing phase transitions at 1.2 , 1.5 , and $1.8M_\odot$, and with radius shifts (ΔR) ranging from 0.0 to 2.4 km, with intervals of 0.2 km (see Fig. 2). We ensured that each sample from the $M(R)$ diagrams contains at least one star in each branch, i.e., hadronic and hybrid. This procedure was applied across all two-branch solution diagrams within the different $(M_c, \Delta R)$ test sets, containing 2529 instances each.

How can we precisely quantify anomalies in our NF model? Leveraging the ability of NF to provide direct probability density estimation on target inputs, we anticipate that $\{M_n, R_n\}$ inputs from two-branch solutions yield lower density values, signaling the presence of an anomaly (phase transition). Calculating the NLL (see Eq. 5) on the test sets, anomalies are anticipated to exhibit substantially high NLL values. To quantify anomalies across different ΔR values, we approach it as a binary classification problem: the positive class represents the presence of an anomaly, while the absence of an anomaly is the negative class.

We establish the anomaly detection threshold by fixing the false positive rate (FPR)¹ at 1%. The FPR, a metric that quantifies the model’s tendency to incorrectly classify negative instances as positive, thus defines the condition for threshold determination on the test sets for the cases with $\Delta R = 0$ (indicating no anomaly). Since we have three sets (see Table I) and two input vector sizes, $N = 5$ and 15, we have trained a total of six NF models, each with its corresponding threshold anomaly.

The process of anomaly detecting is illustrated in Fig. 3, where we show the $N = 5$ (left panel) and $N = 15$ (middle panel) NF models trained on Set 1, using $M_c = 1.5M_\odot$. The gray histograms indicate the NLL of the test sets, and the vertical lines indicate the corresponding threshold values (FPR = 1%). Note that both the gray histograms and the thresholds are defined using the test sets of the one-branch solutions (hadronic EOS), i.e., $\Delta R = 0$. The other colored histograms represent the NLL of the different ΔR sets. For instance, the detection rate of any two-branch solution with $\Delta R \geq 0.8$ km is 100% for $N = 15$ stars (middle panel). The right panel represents a comparison between the two NF models on the $\Delta R = 2.4$ km (purple) sets. As expected, having access to 15 stars is more informative in detecting a two-branch solution than just

¹The FPR is calculated as $FPR = FP/(FP + TN)$, where FP stands for false positive (the number of actual negatives predicted as positives), and TN corresponds to true negative (the number of actual negatives predicted as negatives).

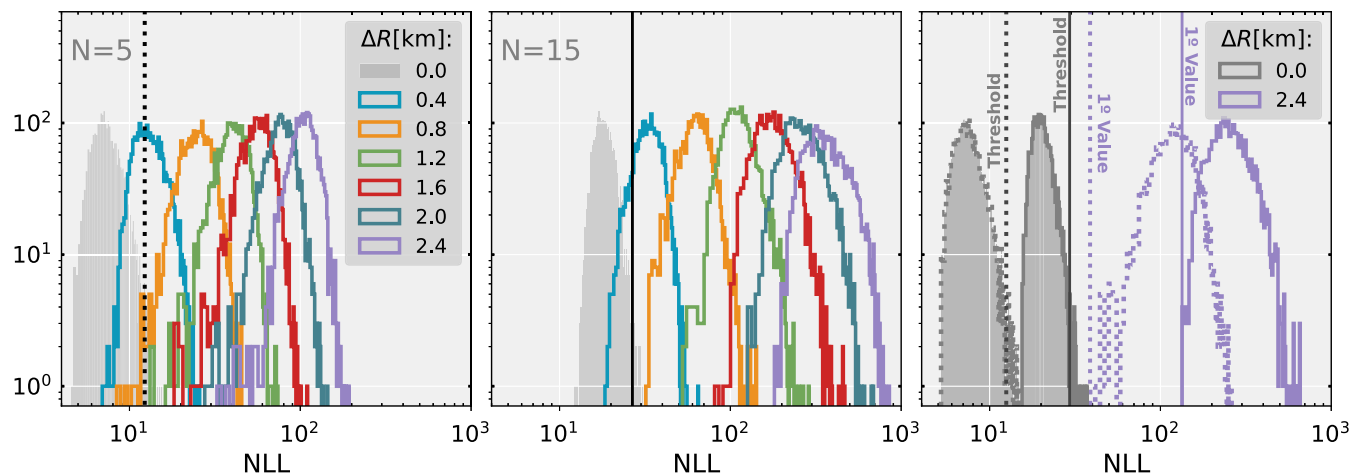


FIG. 3. NLL histograms for the $N = 5$ (left) and $N = 15$ (middle) NF models on the different ΔR sets (colors) using $M_c = 1.5M_\odot$. All test sets were generated according to Set 1 properties (see Table I). The gray regions represent the NLL of each NF model on the test sets with $\Delta R = 0$ (purely hadronic solution), and the vertical black lines correspond to the respective threshold value defined in that region. A comparison between $N = 5$ (dotted) and $N = 15$ (solid) NF models for $\Delta R = 2.4$ km (purple) is also shown (right), where the vertical purple lines mark the first value of the histogram.

five stars. This is evident by analyzing the distance between the model's thresholds and the first histogram's values (represented as purple vertical lines): higher NLL values are assigned to samples in which an anomaly (two-branch solution) is most likely present.

For each threshold within one of the three possible critical masses, we determined how many true positives and false negatives the model detects by counting the number of values above or below the threshold across all different values of ΔR . This allows us to determine the true positive rate (TPR),² also known as recall—a metric that quantifies the ability of the model to correctly identify positive instances—for each ΔR . The TPR for Set 1 is illustrated in Fig. 4 for the three critical masses and in Fig. 5 for two fixed critical masses but across all sets.

The observed increasing trend in Fig. 4, which presents results exclusively for Set 1, aligns with expectations—it is easier to distinguish a two-branch solution as ΔR increases. As expected, a larger size for the target vector (N) is more informative and yields larger TPR values. Comparing the different critical masses, the model shows inferior results for $1.8M_\odot$ because the phase transition occurs at a higher mass value, leading the model to statistically encounter a fewer number of hybrid stars than hadronic stars, increasing the difficulty in detecting any potential phase transition. In the $M_c = 1.2M_\odot$ and $1.5M_\odot$ cases, we observe distinct behaviors for the two different values of N . For $N = 5$, the TPR is higher for $1.5M_\odot$ as the critical mass is positioned in the middle of the curve, making it more susceptible to

variations compared to $1.2M_\odot$, which tends to have only one star below the critical mass for the majority of instances. However, the behavior is different for $N = 15$. This can be attributed to the availability of more stars, and the model is exposed to a greater number of instances below $1.2M_\odot$. The model's increased sensitivity to detect a shifted curve, owing to its training with more information, leads to similar results for the two critical masses.

Figure 5 shows consistently inferior results for TPR at $1.8M_\odot$, in agreement with earlier findings. Furthermore, TPR values are consistently higher for $N = 15$. Examining the three distinct sets, a clear pattern emerges: Set 0 exhibits

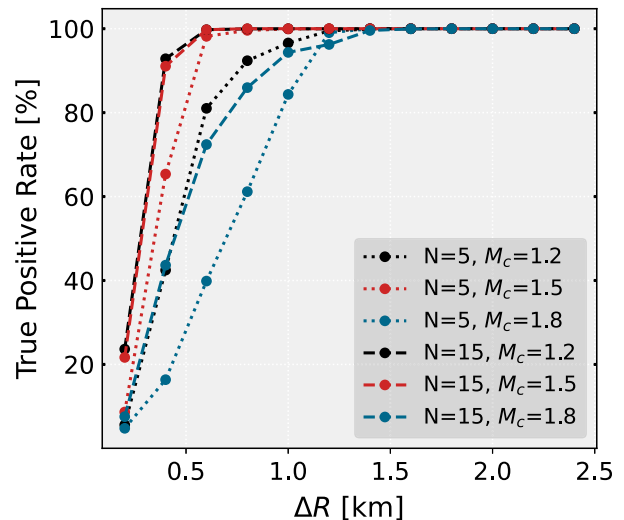


FIG. 4. True positive rate (TPR), phase transition detection rate, as a function of ΔR for different M_c/M_\odot values: 1.2 (black), 1.5 (red), and 1.8 (blue) for Set 1. The use of $N = 5$ or 15 stars is represented by dotted and dashed lines, respectively.

²The TPR is determined as $TPR = TP/(TP + FN)$ where TP is the true positives (instances correctly predicted as positives) and FN is the false negatives (instances incorrectly predicted as negatives).

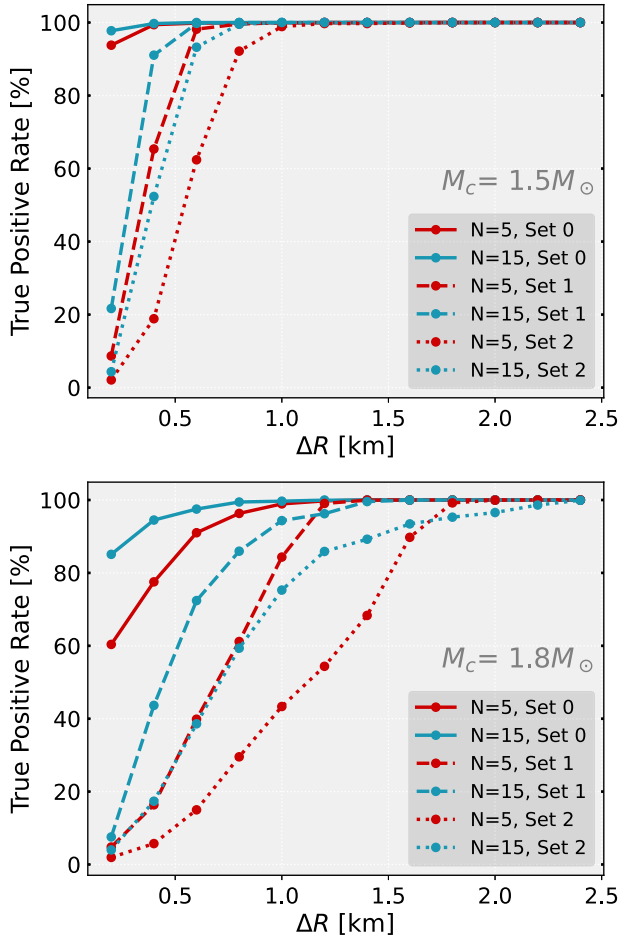


FIG. 5. TPR as a function of ΔR for $M_c/M_\odot = 1.5$ (upper plot) and 1.8 (bottom plot). The different datasets, which represent different (M, R) observation uncertainties (see Table I), are distinguished by line type: Set 0 (solid), Set 1 (dashed), and Set 2 (dotted). The use of $N = 5$ or 15 stars is represented by red and blue lines, respectively.

higher TPR than Set 1, which, in turn, outperforms Set 2. This hierarchy is expected, as introducing more noise to the model's input makes it progressively challenging for the model to effectively detect phase transitions. The results indicate that, given 15 NS observations and $M_c = 1.5M_\odot$, the trained NF models are capable of detecting a phase transition with 100% accuracy if $\Delta R > 0.4$ km (Set 0), $\Delta R > 0.6$ km (Set 1), and $\Delta R > 0.8$ km (Set 2). The accuracy gets considerably lower for $M_c = 1.8M_\odot$, requiring a much larger ΔR to reach 100% accuracy: $\Delta R > 1$ km (Set 0), $\Delta R > 1.6$ km (Set 1), and $\Delta R > 2.4$ km (Set 2).

To evaluate our NF models using an alternative metric, we employed the receiver operating characteristic (ROC) curve [30]. Defined in the (TPR, FPR) space, where both (0, 0) and (1, 1) represent extreme classifier models, this curve provides valuable insights into the model's performance across different threshold settings. The point (0, 0) is achieved by increasing the threshold to classify everything

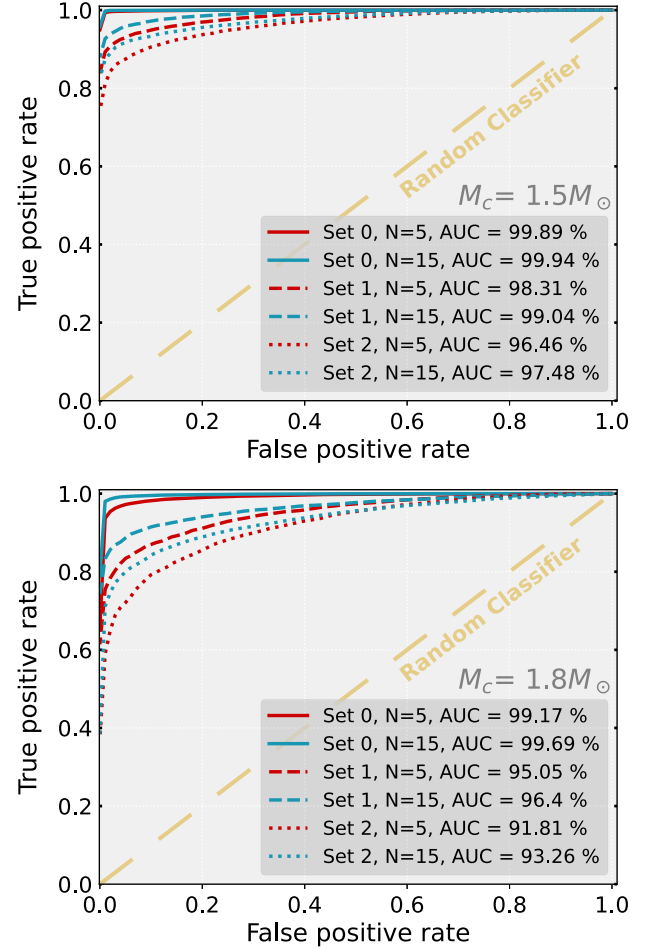


FIG. 6. ROC curves illustrating the performance of the six different trained models, along with the corresponding AUC values, for the critical masses of $M_c = 1.5M_\odot$ (top plot) and $M_c = 1.8M_\odot$ (bottom plot).

as negative for the anomaly presence, resulting in both the TPR and FPR being zero. The opposite occurs at (1, 1), where the threshold classifies everything as an anomaly, leading to both TPR and FPR being 1. The ideal point is (0, 1), signifying a TPR of 1 and an FPR of 0. The ROC curve is then calculated as (TPR, FPR) for a continuous change of the threshold. An important aspect of this metric is that it provides an overview of the model's quality without the need to select a specific threshold. The area under the curve (AUC) [31] provided by the ROC curve offers a straightforward classification of our models, ranging from 0 to 1 (or as a percentage). An AUC value of 1 indicates a perfect classifier, achieving a TPR of 1 with no false positives, visually represented by an ROC curve reaching the top-left corner. In contrast, a diagonal line corresponds to an AUC of 0.5, indicating a random classifier. By examining how the ROC curves approximate toward the top-left corner, we can estimate the AUC value. We use ROC curves to classify the differences between various scenarios constructed within our different sets. In

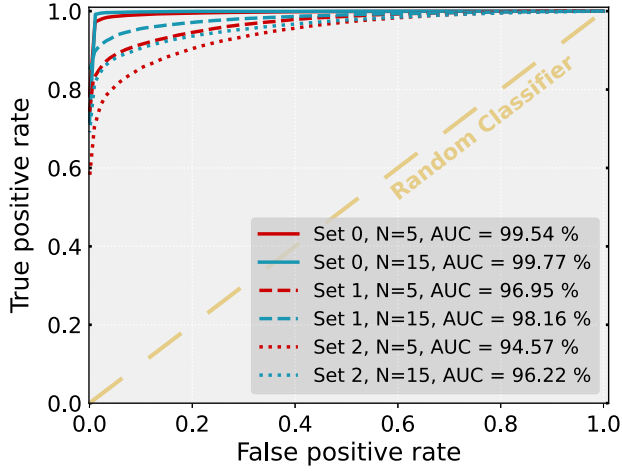


FIG. 7. ROC curves illustrating the performance of the six different trained models, along with the corresponding AUC values, considering all the hypotheses for the critical masses.

Fig. 6, we present the ROC curves and the corresponding AUC values across the six distinct trained models for $M_c = 1.5M_\odot$ (top plot) and $M_c = 1.8M_\odot$ (bottom plot). The AUC values remain larger for $M_c = 1.5M_\odot$, indicating that the model adeptly discerns the presence of a phase transition or anomaly with heightened precision. This pattern persists across all sets, with AUC values ranking in the order Set0 > Set1 > Set2, further emphasizing that higher noise levels in the input hinder the model’s ability to detect phase transitions, and larger input sizes ($N = 15$) consistently yield better performance.

Figure 7 averages each of the six distinct trained models over the M_c values. Therefore, the ROC curves and the corresponding AUCs can be interpreted as an overall measure of the model’s performance. The same pattern of model performance is reinforced: increasing the set size increases the performance, and noisier $M(R)$ samples decrease the anomaly detection rate. This robust performance provides strong evidence supporting the model’s capability to effectively detect phase transitions.

Lastly, we tested the NF model trained with Set 1 and $N = 15$ to a two-branch $M(R)$ curve obtained from microscopic models [4], based on the DD2 nuclear model [32] and the Nambu–Jona-Lasinio quark model [33]. This evaluation involved analyzing the behavior of the histogram of the NLL in relation to the predefined threshold. To conduct this test, we adapted the microscopic model set, generating the same number of samples as in our test set with the specifications of Set 1 and $N = 15$. Additionally, for a more comprehensive evaluation, we created a specific test set with $(M_c, \Delta R) = (1.91M_\odot, 0.77 \text{ km})$, matching the critical mass and ΔR shift of the microscopic model [4]. The results in Fig. 8 show that the NF model detects the two-branch solution for the microscopic model (yellow) with higher confidence than our test set (green), producing

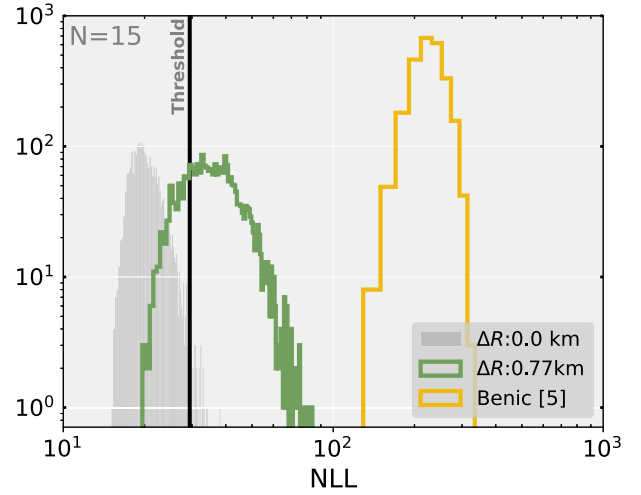


FIG. 8. NLL histogram comparing a test set generated for Set 1 and $N = 15$ with $(M_c, \Delta R) = (1.91M_\odot, 0.77 \text{ km})$ in green and the two-branch $M(R)$ curves derived from microscopic models [4] in yellow. The gray region represents the NLL of the NF model on the test set with $\Delta R = 0$ (purely hadronic solution), and the vertical black line corresponds to the respective threshold value.

NLL values that are noticeably farther from the defined threshold.

Our two-branch solutions can be seen as a lower bound on the model performance, and any other two-branch solution construction becomes easily detectable. This is the case because we are attributing the same properties at the two branches, and it is hard for the NF model to detect a phase transition from a noise sample for small ΔR , which may be easily and wrongly considered from a single one-branch hadronic solution. In the case of the [4] results, the considerable scattering of radii values for hybrid stars and their small mass range make them easier to detect.

V. CONCLUSIONS

We employed NF to identify the presence of a second and disconnected branch of hybrid stars in the NS mass-radius diagram. NF are generative machine learning models capable of generating new samples and estimating the probability density of new data points. Our approach involved training NF models initially on continuous and unique hadronic branch solutions. The input samples incorporate three levels of noise (see Table I) with two possible sample sizes (five stars and 15 stars), resulting in the training of a total of six NF models focused on hadronic solutions. After the training phase, we evaluated each model’s capability to detect possible phase transitions, simulating diverse two-branch solutions by parametrizing them using the star mass at the transition and the radius gap, respectively $(M_c, \Delta R)$. There is no quark model underlying the hybrid branch because our focus was solely on the general $M(R)$ structure. The onset of the hybrid branch is

marked by the critical mass value M_c , with the radius gap between the heaviest hadronic star and the lightest hybrid star represented by ΔR . To quantify the model's ability to detect phase transitions, we approached it as a binary classification problem, defining a threshold condition based on 1% FPR for the NLL space.

The key findings of our investigation highlight the model's enhanced phase transition detection accuracy. These outcomes include the model's adept response to the increase in ΔR across all scenarios, consistently superior precision with an input size of 15 stars as opposed to five stars, and the varied outcomes arising from distinct critical masses—highlighting a lower susceptibility to anomaly detection for $M_c \gtrsim 1.8M_\odot$. Furthermore, another anticipated observation was the reduction in model precision with the increase of the input noise. Furthermore, we evaluated the reliance of our NF models using two-branch $M(R)$ curves obtained from microscopic models [4]. As anticipated, our trained NF models demonstrated enhanced detection capabilities for these two-branch solutions when compared to our test sets. This observation aligns with the straightforward phase transition method we implemented in our study. In summary, our model demonstrated highly expected outcomes, aligning with the anticipated behavior we were expecting for each evaluation. The consistently high AUC values further underline the reliability and robustness of our approach.

In this study, we have not applied the NF models to real observations due to the considerable uncertainty of the present observational data, which could potentially yield inconclusive results, as shown in [14]. A natural next step for

future work would be to expand the data input by gravitational-wave observations, by including the measurements of tidal deformabilities (the mass-weighted tidal deformability of a binary system $\tilde{\Lambda}$, and individual components' tidal deformabilities), in order to systematically investigate regions of phase-transition EOS parameter space which may be degenerate in this type of data [34,35], and directly infer microscopic parameters, e.g., the size of the density discontinuity from ΔR .

A noteworthy aspect not addressed in this study is the quantification of uncertainty, a subject extensively explored in our prior two works [22,23]. This aspect gains increased importance for these models, as they involve multiple neural network architectures within a single forward pass, accentuating the importance of uncertainty quantification. While acknowledging that implementing Bayesian neural networks within the framework of NF would undoubtedly pose computational challenges, it holds potential for future investigations, as exemplified in the work by [36].

ACKNOWLEDGMENTS

This work was supported by FCT—Fundação para a Ciência e Tecnologia, I. P. through the Projects No. UIDB/04564/2020 and No. UIDP/04564/2020, with DOI identifiers 10.54499/UIDB/04564/2020 and 10.54499/UIDP/04564/2020, respectively, and Project No. 2022.06460.PTDC with the associated DOI identifier 10.54499/2022.06460.PTDC, as well as the Polish National Science Centre Grant No. 2021/43/B/ST9/01714.

-
- [1] M. G. Alford, S. Han, and M. Prakash, *Phys. Rev. D* **88**, 083013 (2013).
 - [2] M. G. Alford and S. Han, *Eur. Phys. J. A* **52**, 62 (2016).
 - [3] K. Schertler, C. Greiner, J. Schaffner-Bielich, and M. H. Thoma, *Nucl. Phys. A* **677**, 463 (2000).
 - [4] S. Benic, D. Blaschke, D. E. Alvarez-Castillo, T. Fischer, and S. Typel, *Astron. Astrophys.* **577**, A40 (2015).
 - [5] D. E. Alvarez-Castillo, D. B. Blaschke, A. G. Grunfeld, and V. P. Pagura, *Phys. Rev. D* **99**, 063010 (2019).
 - [6] D. Blaschke, A. Ayriyan, D. E. Alvarez-Castillo, and H. Grigorian, *Universe* **6**, 81 (2020).
 - [7] K. Zhou, L. Wang, L.-G. Pang, and S. Shi, *Prog. Part. Nucl. Phys.* **135**, 104084 (2024).
 - [8] V. Belis, P. Odagiu, and T. K. Aarrestad, *Rev. Phys.* **12**, 100091 (2024).
 - [9] D. J. Rezende and S. Mohamed, Variational inference with normalizing flows, [arXiv:1505.05770](https://arxiv.org/abs/1505.05770).
 - [10] I. Kobyzev, S. J. Prince, and M. A. Brubaker, *IEEE Trans. Pattern Anal. Mach. Intell.* **43**, 3964 (2021).
 - [11] S. R. Green, C. Simpson, and J. Gair, *Phys. Rev. D* **102**, 104057 (2020).
 - [12] J. Wildberger, M. Dax, S. R. Green, J. Gair, M. Pürner, J. H. Macke, A. Buonanno, and B. Schölkopf, *Phys. Rev. D* **107**, 084046 (2023).
 - [13] L. Brandes, C. Modi, A. Ghosh, D. Farrell, L. Lindblom, L. Heinrich, A. W. Steiner, F. Weber, and D. Whiteson, [arXiv:2403.00287](https://arxiv.org/abs/2403.00287).
 - [14] F. Morawski and M. Bejger, *Phys. Rev. C* **106**, 065802 (2022).
 - [15] G. Papamakarios, E. T. Nalisnick, D. J. Rezende, S. Mohamed, and B. Lakshminarayanan, *J. Machine Learning Res.* **22**, 1 (2021).
 - [16] C. Durkan, A. Bekasov, I. Murray, and G. Papamakarios, *Neural Spline Flows* (Neural Information Processing Systems Foundation, Inc, New York, 2019), pp. 7511–7522.
 - [17] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, in *Advances in Neural Information Processing*

- Systems* 32 (Curran Associates, Inc., New York, 2019), pp. 8024–8035.
- [18] C. Durkan, A. Bekasov, I. Murray, and G. Papamakarios, *NFLOWS: Normalizing flows in PyTorch*, Zenodo, [10.5281/zenodo.4296287](https://doi.org/10.5281/zenodo.4296287) (2020).
- [19] N. Yasutake, R. Lastowiecki, S. Benic, D. Blaschke, T. Maruyama, and T. Tatsumi, *Phys. Rev. C* **89**, 065803 (2014).
- [20] T. Malik, M. Ferreira, M. B. Albino, and C. Providência, *Phys. Rev. D* **107**, 103018 (2023).
- [21] T. Malik, M. Ferreira, M. Bastos Albino, and C. Providência, The equation of state for neutron star matter has been obtained through Bayesian inference utilizing a relativistic mean field model with a nonlinear mesonic interaction, Zenodo, [10.5281/zenodo.7854111](https://doi.org/10.5281/zenodo.7854111) (2022).
- [22] V. Carvalho, M. Ferreira, T. Malik, and C. Providência, *Phys. Rev. D* **108**, 043031 (2023).
- [23] V. Carvalho, M. Ferreira, and C. Providência, [arXiv:2401.05770](https://arxiv.org/abs/2401.05770).
- [24] M. Branchesi *et al.*, *J. Cosmol. Astropart. Phys.* **07** (2023) 068.
- [25] M. Evans *et al.*, [arXiv:2109.09882](https://arxiv.org/abs/2109.09882).
- [26] Y. Fujimoto, K. Fukushima, S. Kamata, and K. Murase, [arXiv:2401.12688](https://arxiv.org/abs/2401.12688).
- [27] D. P. Kingma and P. Dhariwal, [arXiv:1807.03039](https://arxiv.org/abs/1807.03039).
- [28] D. P. Kingma and J. Ba, [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
- [29] S. J. Reddi, S. Kale, and S. Kumar, [arXiv:1904.09237](https://arxiv.org/abs/1904.09237).
- [30] M. Majnik and Z. Bosnić, *Intell. Data Anal.* **17**, 531 (2013).
- [31] A. P. Bradley, *Pattern Recognit.* **30**, 1145 (1997).
- [32] S. Typel and H. H. Wolter, *Nucl. Phys. A* **656**, 331 (1999).
- [33] T. Hatsuda and T. Kunihiro, *Phys. Rep.* **247**, 221 (1994).
- [34] S. Soma, H. Stöcker, and K. Zhou, *J. Cosmol. Astropart. Phys.* **01** (2024) 009.
- [35] C. A. Raithel and E. R. Most, *Phys. Rev. Lett.* **130**, 201403 (2023).
- [36] H. J. Hortúa, L. Malagò, and R. Volpi, *Mach. Learn.* **1**, 035014 (2020).