Iranian Journal of Astronomy and Astrophysics
Vol. 11, No. 2, Spring 2024, 81–90
©Available online at http://ijaa.du.ac.ir
DOI: 10.22128/ijaa.2024.809.1180
Online ISSN: 2383–403X

Iranian Journal of Astronomy and Astrophysics

Research Paper

Luminosity Distance Reconstruction using an Ensemble of Neural Networks

Mohammad Hadi Mohammadi[1] · Ahmad Mehrabi*[2]

[1] Department of Physics, Bu-Ali Sina University, Hamedan 65178, 016016, Iran; email: hadi.symbol@gmail.com

[2] Department of Physics, Bu-Ali Sina University, Hamedan 65178, 016016, Iran; *email: mehrabi@ipm.ir

Received: 6 March 2024; **Accepted:** 18 July 2024; **Published:** 28 July 2024

Abstract. Exploring the consistency of a dataset with the ΛCDM model across low and high redshifts stands as a compelling topic in cosmology. Given the capability of neural networks to reconstruct an unknown function, we employed an ensemble of neural networks to reconstruct the luminosity distance based on the Pantheon+ dataset. Each network in the ensemble consists of various numbers of layers and neurons. Since, the neural network can easily provide a reconstruction with a small value of $\chi 2$, it is possible to find a reconstruction with χ^2 smaller than the standard ΛCDM. We selectively choose those reconstructions with a χ^2 value smaller than the best-fit ΛCDM model. Our findings reveal that all reconstructions yield a smaller luminosity distance at high redshifts compared to the best ΛCDM. Assuming a flat universe, we transformed the reconstructions into the Hubble parameter as a function of redshifts and compared the results with predictions of the ΛCDM model.

Keywords: Model-independent reconstruction, ΛCDM tension

1 Introduction

The most well-known model in cosmology is the ΛCDM which describes most cosmological data very well. The Λ stands for the cosmological constant and CDM for the cold dark matter. The model generally is very good at describing the cosmological data including SNIa [1–3], baryon acoustic oscillation (BAO) [4–8], cosmic microwave background (CMB) [9–11] and large scale structure (LSS) [12–14]. However, there are some problems not only from theoretical points of view [15–17] but also from observation at small scale [18]. There are a lot of efforts to understand consistency of the cosmological data with the model and also all of its drawbacks. In addition, recent works show a $> 4\sigma$ deviation from the ΛCDM at high redshift when considering quasars as high redshift standard candles [19–21]. In this direction, [22] also investigated the tension by considering a combination of SNIa, quasars and gamma-ray bursts in a model-independent approach. The results also confirm a deviation at high redshifts. Furthermore, the authors of [23,24] studied how MCMC marginalization and data binning can affect the final results and the potential contributions of these factors to the ΛCDM tensions.

* *Corresponding author*

On the other hand, as it is emphasized in [25,26], a model-independent method can help to understand any possible bias from a model. There are various scenarios for investigating of a dataset in a model-independent manner. Gaussian process (GP) is a collection of normal random variables such that every finite set of them has a multivariate normal distribution [27]. The method has been widely used in cosmology to investigate cosmological data [25, 28–31]. On the other hand, the genetic algorithm (GA) is inspired by the process of natural selection, and given some base functions and a dataset, it is able to reconstruct an unknown function (see [32–34] for some works in cosmology). Moreover, in [35,36], it has been shown that neural networks can be utilized to reconstruct an underlying function and estimate cosmological parameters. A neural network (NN) consists of an input layer, some hidden layers, and an output layer. Each layer has some neurons, and the network is able to reconstruct any unknown function. An NN has been used to investigate cosmological dataset and estimate cosmological parameters model-independently in [37,38]. In these works, the NN not only estimates the cosmological parameters but also provides an estimation of their uncertainty. In addition, in [39], an NN is combined with linear Gaussian model to introduce a new approach to investigate cosmological data.

In current work, we consider an NN for a different purpose. Given the SNIa data, the NN has been used to reconstruct the luminosity distance as a function of redshift. In such a scenario, we are able to find functions with smaller value of χ^2 compared to the best ΛCDM. Finally, it is straightforward to compute the Hubble parameter as a function of redshift from those reconstructions.

The structure of the paper is as follows. In Section 2, details of a typical NN have been given and the roles of hyper-parameters in an NN have been discussed. Additionally, Section 3 covers in-depth details about the dataset and outlines our methodological approach. Subsequently, in Section 4, we comprehensively present and discuss our results. Finally, in Section 5, we summarize our findings and discuss the results.

2 Neural Networks

An NN is a tool in machine learning and has been widely used in regression and classification problems. The network is able to estimate an unknown function and is sometimes called a "universal approximator" [40]. An NN consists of an input layer, some hidden layers, and an output layer. A network with several hidden layers is usually called a deep NN. With advancements in computer hardware, an NN with a large number of hidden layers can now be trained with large amounts of data. In recent years, methods based on NNs have demonstrated exceptional performance in solving cosmological problems with high accuracy and efficiency. For instance, NNs have shown excellent performance in analyzing gravitational waves[41], estimating parameters of the 21 cm signal [42], estimating cosmological parameters [43] and classifying the large-scale structure of the universe [44].

The NNs are designed to identify underlying relationships between input and output data and are purely data-driven methods. Based on this property, we can study a cosmological dataset in a model-independent manner. In this study, we use an NN to reconstruct the luminosity distance of the SNIa pantheon+ data and then find the Hubble parameter from the reconstructions. To implement an NN, we use the PyTorch library in Python and develop a simple code to reconstruct the luminosity distance. In such a scenario, the input data is propagated through the hidden layers, undergoes linear transformations and nonlinear activations, and finally produces an output. Using the given output and targets, one can define a cost function and consider an optimization method to decrease it and adjust the weights and biases of the network.

In each layer, a vector containing neurons is fed as input, which undergoes a linear transformation and then a nonlinear activation. The resulting output is then passed on to the next layer. Formally, this process can be represented in a vectorized style.

$$z_{i+1} = x_i W_{i+1} + b_{i+1}, \tag{1}$$

$$x_{i+1} = f(z_{i+1}). \tag{2}$$

In this context, x_i represents the input row vector of the ith layer. W_{i+1} and b_{i+1} correspond to the linear weights and biases that should be adjusted during the learning process. z_{i+1} denotes the intermediate vector obtained after applying the linear transformation, while f represents the element-wise nonlinear activation function. Notably, the output layer solely performs linear transformations. In practical implementations, when there are n inputs of x with a shape of $1 \times n$ and m neurons, the matrix W has a shape of $n \times m$, and b has a shape of $1 \times m$. Consequently, the intermediate vector z has a shape of $1 \times m$. In Figure 1, a simple NN with two hidden layers has been shown. The network takes a vector of x_i as input and provides two outputs $\hat{y}_1$ and $\hat{y}_2$.

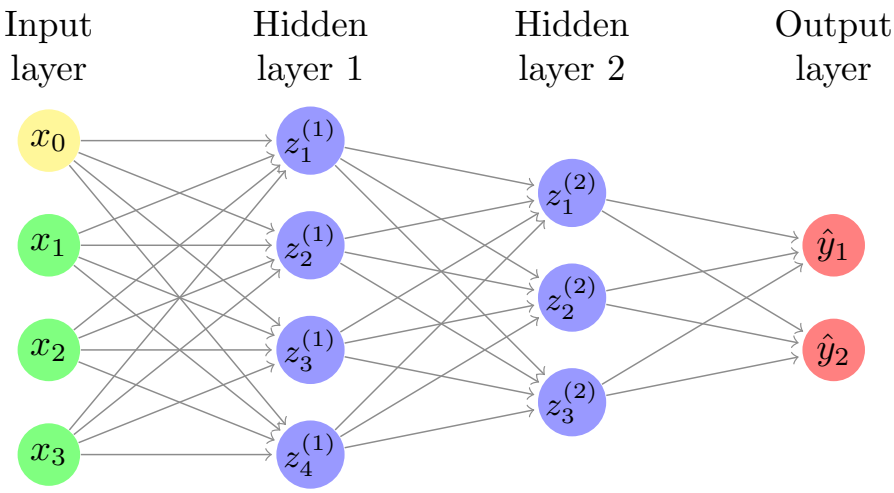

Figure 1: A schematic of an NN with two hidden layers: The network takes a 4D vector and produces two outputs.

An NN can be understood as a function $f_{W,b}$ that takes input X. In supervised learning tasks, each input data is paired with a corresponding ground-truth target, Y. The objective of training a network is to minimize the difference between the predicted result $\hat{Y} = f_{W,b}(X)$ and the ground truth. This difference is measured quantitatively using a function called loss function L(W,b). The network's parameters W and b are then optimized to minimize the loss function. In our procedure, we use the χ^2 as the loss function, and the network tries to find the minimum of the χ^2 during the training phase.

In summary, given a dataset, NNs have proven to be powerful tools for reconstructing a function that describes the data very well. In the current study, we introduce a new nonparametric method based on NNs to reconstruct the luminosity distance from the SNIa data. These reconstructions can then be utilized to obtain the Hubble parameter as a function of redshift.

3 Dataset and method

In the current study, we use the Pantheon+, the latest SNIa sample. The dataset consists of 1701 SNIa light curves observed from 1550 distinct SN. The redshifts of these SN range from a low redshift of $z = 0.001$ up to a maximum redshift of $z_{max} = 2.26$ [3]. Compared to the Pantheon sample, the new sample includes more than 700 additional SN in the range $z < 0.8$, but there is also a significant deletion in the range $0.8 < z < 1$. We consider the full covariance as $C = C_{stat} + C_{syst}$, where C_{stat} mainly includes the statistical uncertainty, such as the full distance error and measurement noise, and C_{syst} includes the systematic uncertainty as described in [45]. Since, those supernovae with $z \leq 0.01$ are highly affected by peculiar velocity, we exclude them from our analysis (The dataset and the full covariance matrix can be obtained from this link).

First of all, we perform an MCMC analysis to find the best value of the ΛCDM parameters namely Ω_m and H_0. We found a minimum $\chi^2 = 1394.5$ and $\Omega_m = 0.332 \pm 0.018$ and $H_0 = 73.84 \pm 0.25$ considering the Pantheon+ dataset.

Now, we will describe the details of our approach to find the Hubble parameter. The module distance is converted to the luminosity distance by

$$D_l(z) = 10^{(\mu-25)/5}, \tag{3}$$

the dataset is subsequently normalized by a constant to achieve a luminosity distance of order ~ 1 (this normalization is necessary for the neural network). The normalized data is then input into an NN to reconstruct the luminosity distance. The loss function employed by the NN is defined as

$$\chi^2 = \Delta\mu C^{-1} \Delta\mu^T, \tag{4}$$

where $\Delta\mu = \mu_{obs} - \mu_{NN}$. μ_{NN} is the module distance at observational points provided by the NN. The input layer is the redshift and the output layer provides the module distance. As for the hidden layers, we have the flexibility to determine their quantity and the number of neurons within each. Upon establishing an NN, the network's weights and biases are initialized randomly, initially resulting in a high χ^2. Through subsequent epochs, the network minimizes this value, ultimately achieving a smaller χ^2. Our investigation into the impact of hyper-parameters (number of layers, number of neurons, activation function and ...) reveals that the final outcomes remain consistent regardless of these parameters, mainly affecting the training speed.

It becomes evident that an NN with more layers and neurons can give a reconstruction with a smaller χ^2 at a faster rate compared to a smaller NN. To enhance the reliability of our approach, we opt for an ensemble of NNs. We employ a random selection method for determining the number of hidden layers and neurons. Each NN's hidden layer count is randomly chosen from 2, 3, or 4, and within each layer, the number of neurons is randomly selected within the range of 4 to 12. Since, each NN has a large number of free parameters, the over-fitting might happen easily. To prevent over-fitting, we adopt a similar procedure like frequntist model selection. In this scenario, we estimate the number of degree of freedom for each NN and exclude all reconstructions with $\chi^2 < \chi^2_{min}$, where χ^2_{min} is given by the χ^2 distribution setting p-value $p = 0.05$.

We generate a set of 50 NNs using the aforementioned criteria and proceed to train them utilizing the Pantheon+ dataset. During this process, reconstructions yielding a χ^2 smaller than the best-fit ΛCDM model are selectively retained, forming a sample set of reconstructions. Finally, we compute the Hubble parameter from

$$H(z) = \frac{c}{(D_l/(1+z))'}, \tag{5}$$

for each reconstruction in the sample. Given a sample of reconstructions, it is straightforward to compute the mean and standard deviation at each point. Similar to the GP method, the standard deviation provides an estimation of uncertainty at each point.

4 Results

In this section, we present and discuss the main results of our analysis. In order to make it as transparent as possible, we present the procedure again.

- The module distances of the Pantheon+ data are converted to the luminosity distances as a function of redshift
- The full covariance matrix including all systematics has been used to compute the χ^2
- The luminosity distances are normalized to have number of order ~ 1 (This is required for feeding data to an NN).
- In the context of the NN approach, we obtain reconstructions from an ensemble of NNs, ensuring independence from the specific network architecture. Furthermore, we select all reconstructions that exhibit a smaller χ^2 than the ΛCDM.
- The reconstructed luminosity distance is converted to the Hubble parameter as a function of redshift

4.1 Neural Networks and ΛCDM

In order to validate our method, we first apply it to a sample of mock data. The dataset is consisted of 500 SN that are simulated in the redshift range of $0.001 - 2.5$ using the ΛCDM model. The uncertainty of each SN is given by a normal distribution with 5% error. For this dataset, we perform above mentioned steps to find the Hubble reconstructions. The simulated data and the luminosity distance reconstructions are shown in Figure 2. The NN provides reconstructions very close to the true one (the reconstructions are very close to each other and provide very narrow area). The Hubble reconstructions and the best fit ΛCDM are presented in Figure 3. The green (red) area shows 95% CI for the ΛCDM (NN reconstructions sample). The results indicate that the NN sample reconstructions are in agreement with the true one.

Considering the NN approach, we follow the aforementioned steps to find the reconstruction of the luminosity distance. The results are shown in Figure 4. The green area (green solid line) represents the 95% confidence interval (CI) (the best fit) for the ΛCDM model. The red one shows the same for the reconstructions from the NN approach. The black points and their bars depict the Pantheon+ data and their uncertainties. As it is clear, since we have a high cadence of data up to redshift ~ 1, both results are almost the same and no significant difference is observed. For higher redshifts, the NN results yield a smaller value for the luminosity distance and the deviation from ΛCDM is significant at redshifts $z > 2$. It is worth noting that in the NN scenario, we select all reconstructions with a χ^2 smaller than the best ΛCDM. Based on this result, all reconstructions that are better than ΛCDM provide a smaller luminosity distance at high redshifts.

Furthermore, we compute the Hubble parameter for all reconstructions and depict the results in Figure 5. The findings indicate a strong agreement between the NN and ΛCDM models up to redshift $z \sim 1$ but at higher redshifts, a significant difference is observed. The NN exhibits a higher expansion rate at high redshifts, which directly stems from the smaller

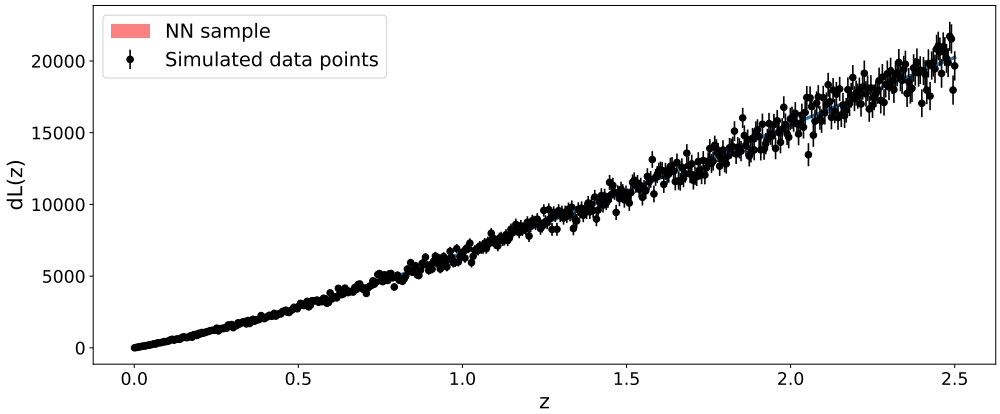

Figure 2: Simulated data and the NN reconstructions sample. NN provides reconstructions very close to the true one.

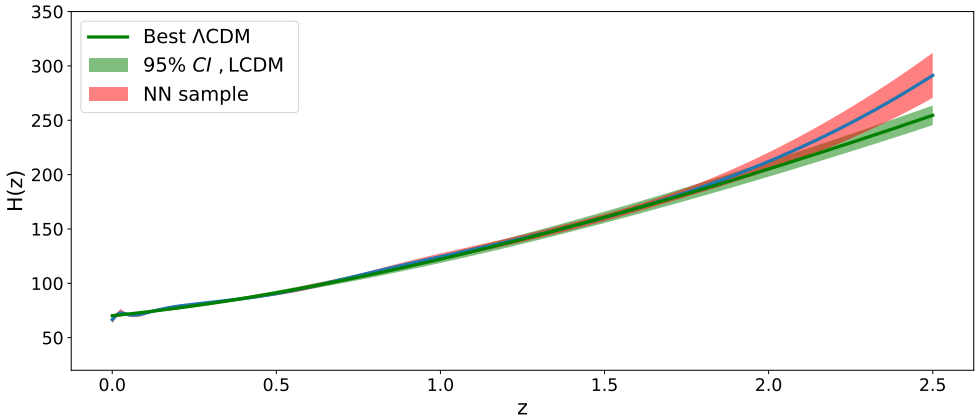

Figure 3: The Hubble reconstruction from the simulated data. The 95% CI for ΛCDM is shown by the green area.

luminosity distance observed at such distances. Additionally, we can readily determine the current expansion rate denoted as H_0, from these reconstructions. The NN provides $H_0 = 75.16 \pm 0.69$ whereas the best fit for ΛCDM is $H_0 = 73.84 \pm 0.25$.

5 Conclusion

Some recent works have demonstrated a deviation from the ΛCDM model to some extent. In this line of research, we investigate the Pantheon+ dataset using two model-independent methods. The distance module is converted into the luminosity distance, which is then inputted into the algorithm to reconstruct the luminosity distance as a function of redshift. Finally, the Hubble parameter is derived from the derivative of the luminosity distance.

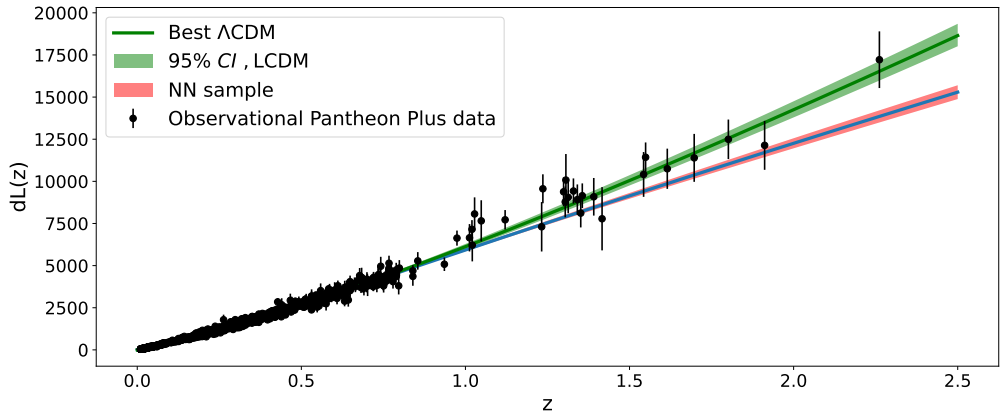

Figure 4: The luminosity distance reconstruction through the NN approach. The red interval shows the 95% CI. The black points show the Pantheon+ data and their uncertainties. The green area indicates the 95% CI for the ΛCDM model.

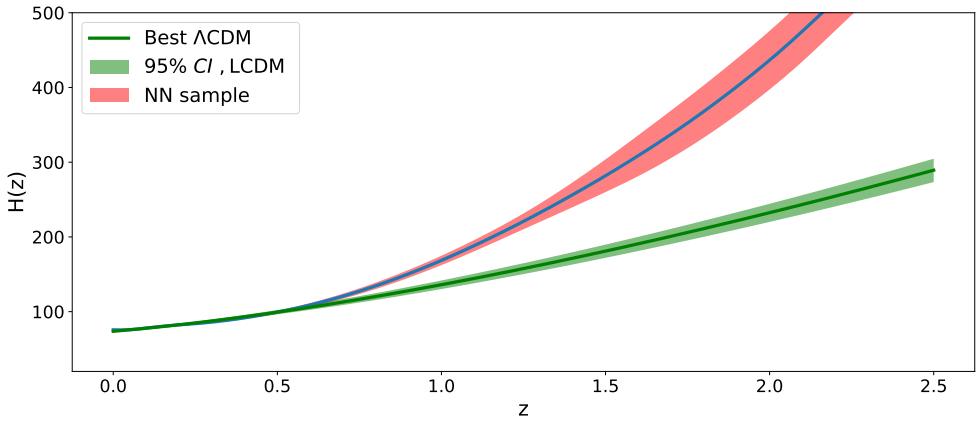

Figure 5: The reconstructed Hubble parameter as a function of redshift in the NN scenario. The red (green) area shows the 95% CI for the NN (ΛCDM).

In the first approach, we utilize an ensemble of NNs to generate the reconstructions. Each network consists of several hidden layers with multiple neurons. The loss function employed is the χ^2 of the dataset, encompassing all systematic uncertainties. We select reconstructions with a χ^2 smaller than that of ΛCDM model. Indeed, the NN approach offers a unique and straightforward method to identify reconstructions with χ^2 values smaller than a specified threshold. Our findings indicate that all reconstructions demonstrate smaller luminosity distance at higher redshifts. Assuming a flat universe, it is simple to convert each reconstruction into the Hubble parameter. From the NN reconstructions, we derived a current expansion rate of $H_0 = 75.16 \pm 0.69$, which is consistent with ΛCDM ($H_0 = 73.84 \pm 0.25$). While the Hubble parameter derived from the NN approach aligns well with ΛCDM up to a redshift of $z \sim 1$, a notable deviation becomes evident at higher

redshifts. This suggests that all reconstructions with smaller χ^2 values provide a larger Hubble parameter compared to ΛCDM at higher redshifts.

Our analysis also indicates that NNs are able to easily give reconstruction with smaller χ^2 value compere to the best ΛCDM. An ensemble of such reconstructions can be used to investigate possible deviation from the ΛCDM.

Authors' Contributions

All authors have the same contribution.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare no potential conflicts of interest.

Ethical Considerations

The authors have diligently addressed ethical concerns, such as informed consent, plagiarism, data fabrication, misconduct, falsification, double publication, redundancy, submission, and other related matters.

Funding

This research did not receive any grant from funding agencies in the public, commercial, or nonprofit sectors.

References

[1] Betoule, M., et al. 2014, A&A, 568, A22.

[2] Scolnic, D. M., et al. 2018, ApJ, 859, 101.

[3] Scolnic, D., et al. 2022, ApJ, 938, 113.

[4] Seo, H.-J., & Eisenstein, D. J. 2005, ApJ, 633, 575.

[5] Percival, W. J., Reid, B. A., Eisenstein, D. J., & et al. 2010, MNRAS, 401, 2148.

[6] Blake, C., et al. 2011, MNRAS, 415, 2876.

[7] Reid, B. A., et al. 2012, MNRAS, 426, 2719.

[8] Abbott, T. M. C., et al. 2019, MNRAS, 483, 4866.

[9] Hinshaw, G., et al. 2013, ApJS, 208, 19.

[10] Ade, P. A. R., et al. 2016, A&A, 594, A13.

[11] Aghanim, N., et al. 2020, A&A, 641, A6, [Erratum: A&A, 652, C4 (2021)].

[12] Tegmark, M., et al. 2004, Phys. Rev. D, 69, 103501.

[13] Beutler, F., et al. 2011, MNRAS, 416, 3017.

[14] Alam, S., et al. 2017, MNRAS, 470, 2617.

[15] Weinberg, S. 1989, Rev. Mod. Phys., 61, 1.

[16] Astashenok, A. V., & Del Popolo, A. 2012, Classical and Quantum Gravity, 29, 085014.

[17] Martin, J. 2012, Comptes Rendus Physique, 13, 566.

[18] Del Popolo, A., & Le Delliou, M. 2017, Galaxies, 5, 17.

[19] Risaliti, G., & Lusso, E. 2019, Nature Astron., 3, 272.

[20] Lusso, E., et al. 2019, A&A, 628, L4.

[21] Risaliti, G., et al. 2023, Astron. Nachr., 344, e230054.

[22] Mehrabi, A., & Basilakos, S. 2020, Eur. Phys. J. C, 80, 632.

[23] Colgáin, E. O., Pourojaghi, S., Sheikh-Jabbari, M. M., & Sherwin, D. 2023, [arXiv:2307.16349v2].

[24] Colgáin, E. O., Sheikh-Jabbari, M. M., & Solomon, R. 2023, Phys. Dark Univ., 40, 101216.

[25] Mehrabi, A., & Rezaei, M. 2021, ApJ, 923, 274.

[26] Vazirnia, M., & Mehrabi, A. 2021, Phys. Rev. D, 104, 123530.

[27] Rasmussen, C. E., & Williams, C. K. I. 2005, Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning), The MIT Press.

[28] Shafieloo, A., Kim, A. G., & Linder, E. V. 2012, Phys. Rev. D, 85.

[29] Liao, K., Shafieloo, A., Keeley, R. E., & Linder, E. V. 2019, ApJ, 886, L23.

[30] Gómez-Valent, A., & Amendola, L. 2018, JCAP, 2018, 051.

[31] Briffa, R., et al. 2020, Class. Quant. Grav., 38, 055007.

[32] Bogdanos, C., & Nesseris, S. 2009, JCAP, 2009, 006.

[33] Nesseris, S., & García-Bellido, J. 2012, JCAP, 2012, 033.

[34] Nesseris, S., & Shafieloo, A. 2010, MNRAS, 408, 1879.

[35] Wang, G.-J., Ma, X.-J., Li, S.-Y., & Xia, J.-Q. 2020, ApJS, 246, 13.

[36] Dialektopoulos, K., et al. 2022, JCAP, 02, 023.

[37] Dialektopoulos, K. F., Mukherjee, P., Levi Said, J., & Mifsud, J. 2023, Eur. Phys. J. C, 83, 956.

[38] Pal, S., & Saha, R. 2023, [arXiv:2309.15179].

[39] Mehrabi, A. 2023, European Physical Journal Plus, 138, 714.

[40] Hornik, K. 1991, Neural Networks, 4, 251.

[41] George, D., & Huerta, E. A. 2018, Phys. Rev. D, 97, 044039.

[42] Schmit, C. J., & Pritchard, J. R. 2017, MNRAS, 475, 1213.

[43] Ribli, D., et al. 2019, MNRAS, 490, 1843.

[44] Aragon-Calvo, M. A. 2019, MNRAS, 484, 5771.

[45] Brout, D., et al. 2022, ApJ, 938, 110.