ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/347446475

Deep learning approach to Hubble parameter

Article *in* Computer Physics Communications · April 2021

DOI: 10.1016/j.cpc.2020.107809

CITATIONS
11

READS
681

4 authors, including:

Mustafa Salti
Mersin University
99 PUBLICATIONS **867** CITATIONS
SEE PROFILE

Oktay Aydoğdu
Mersin University
90 PUBLICATIONS **1,296** CITATIONS
SEE PROFILE

Evrim Ersin Kangal
Mersin University
295 PUBLICATIONS **26,361** CITATIONS
SEE PROFILE

All content following this page was uploaded by Mustafa Salti on 11 February 2021.

The user has requested enhancement of the downloaded file.

Deep learning approach to Hubble parameter

H. Tilaver*,[1] M. Salti†,[1] O. Aydogdu‡,[1] and E. E. Kangal§[2]

[1]*Department of Physics, Faculty of Arts and Science, Mersin University, Mersin, TR 33343, Turkey*

[2]*Computer Technology and Information Systems, School of Applied Technology and Management of Erdemli, Mersin University, Mersin, TR 33740, Turkey*

Abstract

The main purpose of this work is to show that machine learning algorithms (MLAs) can be used to improve the abilities of cosmological models and to make meaningful astrophysical predictions. As a preliminary step, we construct an expression for the Hubble parameter in the caloric variable Chaplygin gas (cVCG) framework including a particle creation scenario. Then, making use of a set of updated observational data, we obtain the best-fitting values of the auxiliary model parameters. In the main part of the article, we discuss the model from the machine learning (ML) perspective via two different supervised learning-training algorithms: Long Short-Term Memory (LSTM) cells with Dropout and the Fisher Information Matrix (FIM). We see that the constructed theoretical ground yields very successful results when it is used in the MLAs for training and the reliability level of deep learning (DL) analysis is above %93.

PACS numbers: 04.50.+h, 07.05.Mh, 95.35.+d, 98.80.-k

Keywords: Machine learning, deep network, cosmology, dark energy.

* hakantilaver@hotmail.com

† msalti@mersin.edu.tr

‡ oktaydogdu@mersin.edu.tr

§ evrimersin@gmail.com

I. INTRODUCTION

General Relativity (GR) has reshaped our ideas about gravity and space-time structure because of revolutionary innovations. The theory introduced various new issues into physics such as space-time expansion and gravitational waves. Recently, astronomers have come to a brand-new outcome in which the cosmos enters in a speedy expansion phase[1–10]. This significant detection could only be accounted for by the dynamics of a hitherto unknown form of constituents of the cosmos: "Dark Matter" (DM) and "Dark Energy" (DE). Till now, scientists have introduced a variety of ideas to explain this extraordinary behavior. All of these proposals can be broadly classified (check Ref.[11, 12] and references therein) into two major families: (i) changing the geometry part of the Einstein field equations, (ii) redefining the matter-energy part. On this purpose, theoretical physicists were looking for an interesting idea which could simultaneously describe the dynamical mechanism of both the DM and the DE. Consequently, these investigations led to the birth of the Chaplygin Gas (CG) cosmology[13–16], which automatically removes the cosmic coincidence puzzle (*why the late-time enlargement is happening now and why it is speedy?*). In cosmology, the CG model has been taken into account extensively to interpret the recent astrophysical data more accurately in different papers[17–21]. On the other hand, in astronomy, different forms of the CG model have widely been used (check Refs.[22–24] and references therein) to discuss stellar evolution of the compact star, which refers collectively to white dwarfs, black holes, neutron stars, gravastars and exotic stars.

Artificial neural network (ANN) mechanism has become prevalent recently in diverse scientific fields such as cosmology[25–27], medicine[28], economy[29] and meteorology[30]. The main idea behind such use of the ANN is the ML algorithm, which is a subset of artificial intelligence (AI) and includes advanced methods that enable computers to learn from a family of data. In an ML mechanism, computers use a parameterized (or non-parameterized) proposal describing a set of possible rules to reach the best results among the possible ones. Hence, it has been applied to forecast and learn cause-effect relationship between variables. In this paper, we apply two different DL mechanisms to the Hubble parameter with the help of the cVCG cosmology in order to check whether the designed model can be used to introduce meaningful predictions for the future of the universe.

The layout of the paper is as follows. In the next section, we reconstruct the Hubble

parameter with the help of the cVCG model and fit its free parameters by using the most recent astrophysical data. In the third section, we apply two well-known ML approaches to the model and discuss the statistical aspects of obtained data. In the fourth section, we give some closing remarks. Note that all numerical investigations were completed by means of high-level programming environments Mathematica[31] and Pyhton[32]. Throughout this study, the unit $c = 1$ is adopted.

II. RECONSTRUCTING HUBBLE PARAMETER

As a first step, we assume that the cosmos is represented by the flat Friedmann-Robertson-Walker (FRW henceforth) metric. The recent astrophysical data[1–10, 33] have strongly suggested that the geometry of the cosmos is spatially flat at large scales. So, the corresponding line-element including the scale factor $a(t)$ is $ds^2 = dt^2 - a^2(t)\sum_{i=1}^{3}(dx_i)^2$. In the present work, we suppose also that the cosmos is filled with a perfect fluid described by the energy-momentum tensor $T_{\mu\nu} = (\rho_t + p_t)u_\mu u_\nu - g_{\mu\nu}p_t$, where u_μ, $\rho_t = \rho_b + \rho$, $p_t = p_c + p$, $g_{\mu\nu}$, ρ_b and p_c represent the four-velocity vector, total energy density, total pressure, metric tensor, energy density of the baryonic matter (BM) and the particle creation pressure, respectively. Also, note that $\rho = \rho_m + \rho_e$ and $p = p_e$, where the subscripts m and e stand for the DM and the DE, respectively. Here, the continuity equation $T^{\mu\nu}_{;\nu} = 0$ yields the solution $\rho_b = \rho_b^0 a^{-3}$ for the BM.

The VCG[34–36] is defined by the following Equation-of-State (EoS)

$$p = -\frac{Ba^{-n}}{\rho}, \tag{1}$$

where B denotes any positive constant and n is a free parameter. It is known that adiabatic particle creation indicates the entropy produced in space-time with a constant quantity $\delta = \frac{S}{N}$, which represents entropy per particle. The particle creation pressure is described by[37–39]

$$p_c = -(\rho + p)\eta, \tag{2}$$

where η is connected with the particle number density and satisfies the condition $0 \leq \eta \leq 1$.

From thermodynamics, we have

$$p = -\left(\frac{\partial U}{\partial V}\right)_S, \tag{3}$$

where $U=\rho V$, $V=a^3$ and S represent the energy, volume and the entropy of the system, respectively. From this point of view, it can be written that $\left(\frac{\partial U}{\partial V}\right)_S=-(p+p_c)$. After some algebra, we reach the following differential equation

$$\frac{dU}{dV}+(\eta-1)B\frac{V^{3n+1}}{U}-\frac{U}{V}\eta=0. \tag{4}$$

Then, it is found that

$$U(V)=V^n\sqrt{\frac{2B(1-\eta)}{3n+2(1-\eta)}V^{n+2}+c}, \tag{5}$$

where c denotes an integration constant which may depend on the entropy only or be a universal constant. Consequently, we get

$$\rho=a^{3n-3}\sqrt{\frac{2B(1-\eta)}{3n+2(1-\eta)}a^{3n+6}+c}, \tag{6}$$

$$p=-\frac{Ba^{-4n+3}}{\sqrt{\frac{2B(1-\eta)}{3n+2(1-\eta)}a^{3n+6}+c}}, \tag{7}$$

$$p_c=a^{3n-3}\eta\left[\sqrt{\frac{2B(1-\eta)}{3n+2(1-\eta)}a^{3n+6}+c}-\frac{Ba^{-7n+6}}{\sqrt{\frac{2B(1-\eta)}{3n+2(1-\eta)}a^{3n+6}+c}}\right]. \tag{8}$$

Now, we are in a place to construct a theoretical expression for the cosmic Hubble parameter H. We may write the Friedman equation (FE) as[40, 41]

$$H^2=\left(\frac{1}{a}\frac{da}{dt}\right)^2=\frac{8\pi G}{3}(\rho_b+\rho). \tag{9}$$

For the sake of simplicity, we can define the dimensionless fractional densities $\Omega_b=\frac{\rho_b}{\rho_{cri}}$ and $\Omega_d=\frac{\rho}{\rho_{cri}}$, where $\rho_{cri}=\frac{3H_0^2}{8\pi G}$ is known as the critical density and H_0 shows the Hubble constant. According to the Planck-results[9, 10], we have $(\Omega_b^0,\Omega_m^0,\Omega_e^0)=(0.049,0.278,0.673)$; thus, we can write $\rho_b^0=0.042\rho_0$, where $\rho_0=\rho_m^0+\rho_e^0$. In addition to this, it is generally known that the present energy density is related[21] to the cosmological density by $\rho_t^0=1.31\rho_{cos}$, where $\rho_t^0=\rho_b^0+\rho_0$. Hence, one should take $\rho_t^0=1.31$ for $a_0=1$, as a result, it is obtained that $(\rho_b^0,\rho_0)=(0.05,1.26)$. Subsequently, we can write the cosmic Hubble parameter H as a function of the cosmic red shift parameter $z=\frac{1}{a}-1$:

$$H^2(z)=H_0^2\left[\frac{1-\Omega_d^0}{(1+z)^{-3}}+\frac{\Omega_d^0}{\rho_0(1+z)^{3n-3}}\left\{\frac{2B(1-\eta)(1+z)^{-3n-6}}{3n+2(1-\eta)}+c\right\}^{\frac{1}{2}}\right], \tag{10}$$

with

$$c = 1.59 - \frac{2B(1-\eta)}{3n + 2(1-\eta)}. \tag{11}$$

It is significant to emphasize here that the reconstruction of the Hubble parameter has been performed with the help of a well-known method developed in literature[40, 41]. However, the same goal can be reached by following different ways (please check Refs.[42, 43] and references therein for detailed information).

At this point, we need to get best-fitting values of the auxiliary parameters. Markov Chain Monte-Carlo (MCMC) algorithm has found its place in the field of cosmology and has become a significant numerical tool especially in parameter estimation. Now, we consider the fundamental concepts of the MCMC method implemented in python to find out the combination of auxiliary parameters that best describes the universe.

TABLE I: Cosmic chronometer measurements.

z	$H(z)$	Ref.	z	$H(z)$	Ref.	z	$H(z)$	Ref.
0.071	69.00∓19.7	[44]	0.430	86.50∓3.70	[48]	0.730	97.30∓7.00	[50]
0.090	69.00∓12.0	[45]	0.440	82.60∓7.80	[50]	0.781	105.0∓12.0	[47]
0.120	68.60∓26.2	[44]	0.449	92.80∓12.9	[49]	0.875	125.0∓17.0	[47]
0.170	83.00∓8.00	[46]	0.470	89.00∓49.6	[51]	0.880	90.00∓40.0	[52]
0.179	75.00∓4.00	[47]	0.478	80.90∓9.00	[49]	0.900	117.0∓23.0	[46]
0.199	75.00∓5.00	[47]	0.480	97.00∓62.0	[52]	1.037	154.0∓20.0	[47]
0.200	72.90∓29.6	[44]	0.510	90.40∓1.90	[53]	1.300	168.0∓17.0	[46]
0.240	79.69∓2.65	[48]	0.520	94.35∓2.64	[54]	1.363	160.0∓33.6	[56]
0.270	77.00∓14.0	[46]	0.560	93.34∓2.30	[50]	1.430	177.0∓18.0	[46]
0.280	88.80∓36.6	[44]	0.570	96.80∓3.40	[55]	1.530	140.0∓14.0	[46]
0.352	83.00∓13.5	[47]	0.593	104.0∓13.0	[47]	1.750	202.0∓40.0	[46]
0.380	83.00∓17.0	[49]	0.600	87.90∓6.10	[50]	1.965	186.5∓50.4	[56]
0.400	95.00∓17.0	[46]	0.610	97.30∓2.10	[53]	2.300	224.0∓8.00	[57]
0.401	77.00∓10.2	[49]	0.640	98.82∓2.98	[54]	2.340	222.0∓8.50	[58]
0.425	87.10∓11.2	[49]	0.679	92.00∓8.00	[47]	2.360	226.0∓9.30	[59]

In TABLE I, we display the recent $z - H(z)$ data set (the blue values were calculated

by the differential ages technique while the red ones were obtained by the standard ruler method). On the basis of the expression (10) and the TABLE I, we can constrain the cVCG proposal. Now, the main goal of parameter prediction is to maximize the likelihood in order to obtain the most suitable set of model parameters. If we focus on the Gaussian approximation, one can conclude that the likelihood will be maximum if the quantity

$$\chi^2 = \sum_i^{27} \frac{[H_{mea}(z_i) - H_{mod}(z_i)]^2}{\sigma_i^2}, \tag{12}$$

is minimum. Here, H_{mea}, H_{mod} and σ_i describe the measured values of the Hubble parameter, theoretical values of the Hubble parameter and the observational uncertainty, respectively. The quantity χ^2 is connected with the Gaussian likelihood via the relation

$$L = L_0 \exp\left[-\frac{1}{2}\chi^2\right]. \tag{13}$$

Consequently, it can be said that maximizing the Gaussian likelihood is equivalent to minimizing χ^2. As a result, we find the following best-fitting values for the arbitrary parameters:

$$(H_0, \Omega_d^0, \eta, B, n) = (73.2134, 0.9404, 0.6412, 2.4026, 0.2615), \tag{14}$$

with

$$\chi^2_{min} = 19.9022. \tag{15}$$

Once the best-fitting values of the free model parameters are obtained, we would like to discuss the confidence territories, where the values reasonably considered good candidates for our model, with the help of a fast MCMC algorithm. In FIG. 1, we display likelihood distributions for combination of the model parameters.

The Planck spacecraft data[9, 10] have indicated that $\Omega_d^0 = 0.951$, and the most recent astrophysical observations have implied that

$$H_0\left[\frac{\text{km}}{\text{sec Mpc}}\right] = \begin{cases} 70.00^{+12.0}_{-8.00} & 2017, \text{LIGO-VIRGO Collaboration[60]} \\ 67.66 \mp 0.42 & 2018, \text{Planck Spacecraft[10]} \\ 73.52 \mp 1.62 & 2018, \text{Gaia Spacecraft[61]}, \\ 73.3^{+1.70}_{-1.80} & 2019, \text{H0LiCOW Collaboration[62]}, \\ 74.2^{+2.70}_{-3.00} & 2020, \text{STIDES[63]}. \end{cases} \tag{16}$$

We see that best-fitting values of the free parameters H_0 and Ω_d^0 are in good consistency with the most recent observational data presented in equation (16). Moreover, our understanding

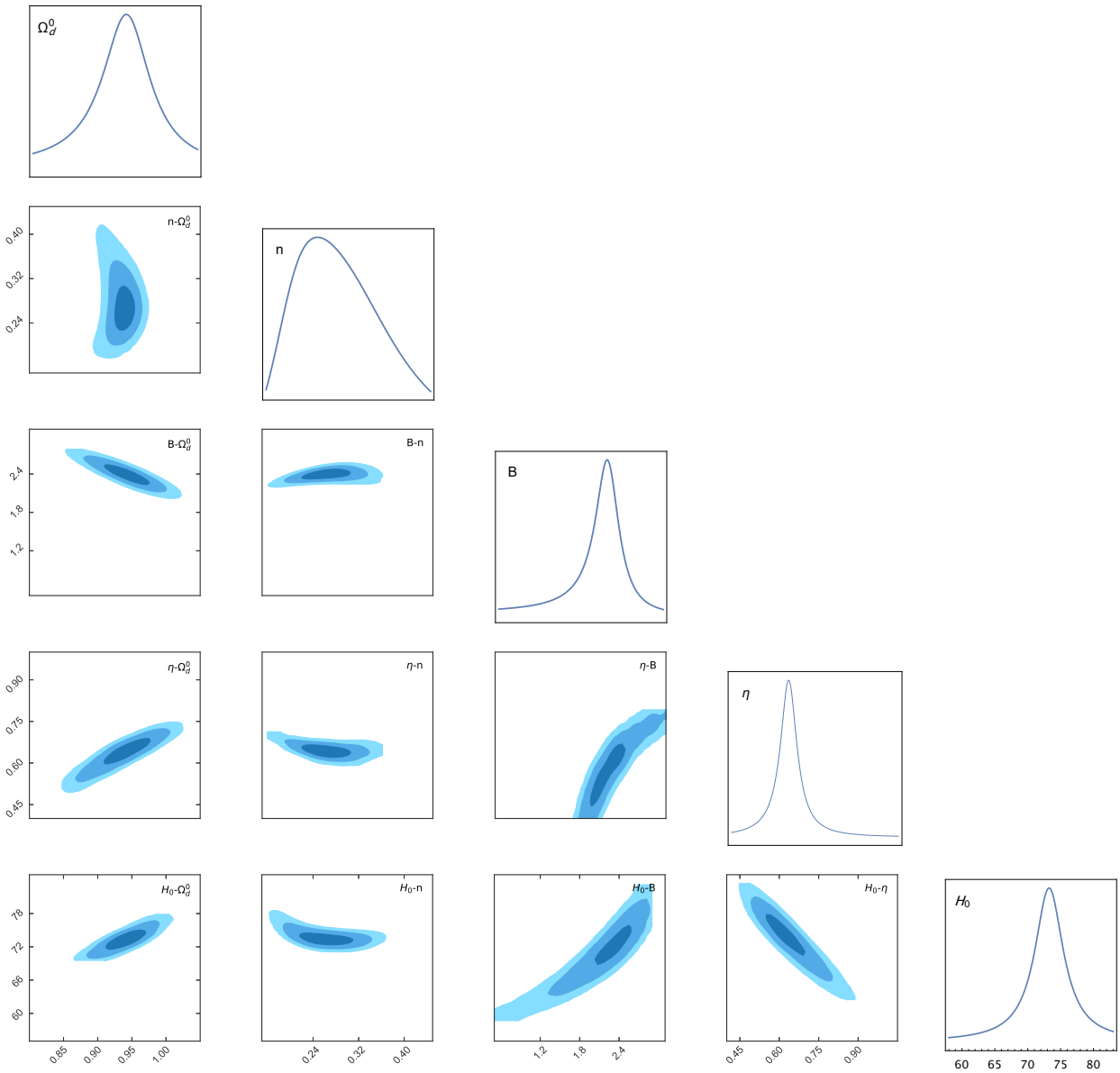

FIG. 1: Two-dimensional likelihood distributions for the combination of free parameters in 1σ (dark blue), 2σ (blue) and 3σ (light blue) confidence regions.

of the cosmic accelerated expansion includes a significant discrepancy between the indirect and direct observations of the present enlargement rate of the cosmos[64–66]. Since the first determination by Hubble[67], astronomers have derived different values for the Hubble constant in the range $50 < H_0 < 100$ km/s/Mpc. Our result agrees with the most recent measurements[61–63] but in tension with Planck-results[9, 10].

In FIG. 2, we depict the evolutionary nature of the cosmic Hubble parameter and illustrate some recent observational values. It is easily seen that our $H(z)$ function is in good

agreement with the recent astrophysical measurements.

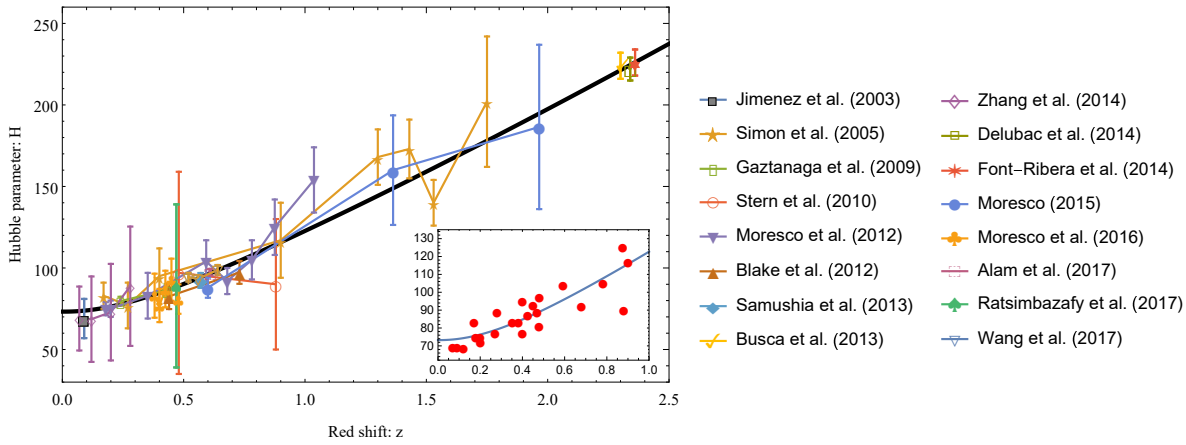

FIG. 2: $H \sim z$ relation according to the cVCG model in the 1σ confidence region and recent measured values.

III. DEEP LEARNING ANALYSIS

Supervised MLAs allow us to produce an output data from previous experiences and to solve various types of real-world computation problems. Here, we want a mechanism that considers a value of the red shift as an argument and returns an estimated value for the Hubble parameter. The first option is to give explicitly the red shift value as a well-defined function of the Hubble parameter, which means that we have to introduce a function f satisfying $H = f(z)$. However, there could be no idea of what the function f is. Then, one can rely on a dataset to construct a rule in the ML way.

The DL algorithms learn from more complex patterns than the Neural Network (NN) mechanism with the help of more layers provided by the Recurrent Neural Network (RNN)[68]. The LSTM cell[69] is a famous RNN approach, however, it suffers from the over-fitting issue. One of the best ways to remove this effect is to use the Dropout regularization like Gaussian processes because a unit in a NN mechanism is temporarily removed from a network. On the other hand, the FIM algorithm is a fundamental mechanism in both statistics and deep network topics and takes a significant role in the parameter space. In this section, we focus on the LSTM-Dropout and the FIM mechanisms for the DL analysis of our investigation.

A. Basics of the LSTM-Dropout architecture

Connections between neuron-like nodes form a directed network along a temporal sequence in the RNN algorithm, thus, this feature allows the mechanism to exhibit a temporal dynamic behavior. The RNN is based on a network of neurons which is organized into successive layers and connected with a one-way path to every other node in the next successive layer. There are principally four main modes to run the RNN: one-to-one, one-to-many, many-to-one and many-to-many[70–72]. Here, we should use a form of the RNN mechanism with one-to-one mode. Thus, while designing an architecture for our problem, we can start with a point where the output layer of the previous step is used to calculate a new cell (see FIG. 3).

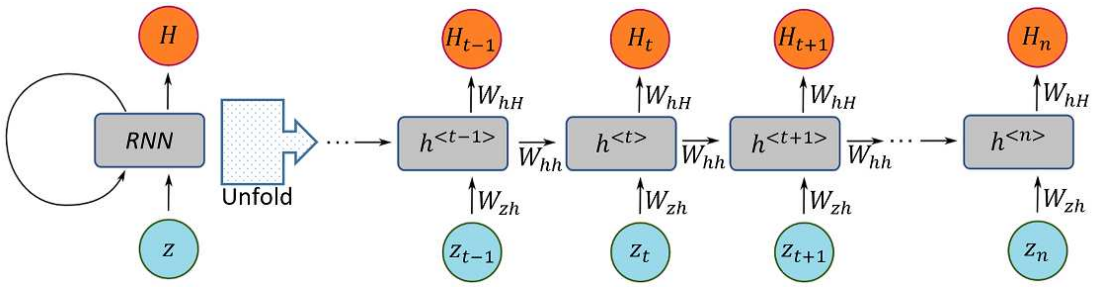

FIG. 3: RNN architecture scanning through the data from left to right.

The Bayesian Neural Networks (BNNs) can give us a mathematical perspective about uncertainty, but these mechanisms have prohibitive computational costs. Recently, Gal and Ghahramani[73] have developed a new tool giving a chance to capture uncertainty. Similar to other regularization approaches, the LSTM-Dropout network yields effectively better results when there is a limited amount of training data and the model is likely to over-fit the training data[74]. In the LSTM-Dropout network, the aim is to reduce the difference between the input data and the estimated output, and this task is done by a function that we need to minimize (it is called the cost function or the loss function)[74]. Besides, optimizers update the weight parameters to minimize the loss function. As a matter of fact, the loss function acts as a guide that allows the optimizer to move in the right direction to reach the global minimum. Therefore, optimizers play a very crucial role while increasing the accuracy of the model. There are several optimizer variants: the Nesterov Accelerated Gradient (NAG)[75], Adagrad[76], AdaDelta[77], Adaptive Moment

(Adam)[78], AdaMax[78], Nesterov-accelerated Adaptive Moment Estimation (Nadam)[79] and the AMSGrad[80]. Adam computes adaptive learning rates for each parameter and is generally selected as the best overall choice.

In the LSTM-Dropout mechanism, the hidden state values and output data can be calculated as

$$h^{<t>} = \Theta_f(W_{hh} \odot h^{<t-1>} + W_{zh} \odot z^{<t>} + b_h), \tag{17}$$

$$H^{<t>} = \Theta_f(W_{hH} \odot h^{<t>} + b_H), \tag{18}$$

where $h^{<t>}$ represents the hidden layer and its $h^{<t-1>}$ value indicates the previous one, Θ_f shows the activation function and $H^{<t>}$ corresponds to the output layer. W_{zh}, W_{hh} and W_{hH} are weights for the connection of the input layer to the hidden layer, the hidden layer to the hidden layer and the hidden layer to the output layer, respectively. Also, the superscript $<>$ defines a vector and the symbol $\odot$ indicates a matrix product. Next, t implies the time (here we take $t = 1$ for the first red shift value) and b denotes the corresponding bias. In this algorithm, for making a prediction at time t, the mechanism uses not only the input $z^{<t>}$ at time t but also information from the previous state at time $t-1$ through the hidden state parameter h and the weight W_{hh} which passes from the previous hidden state to the current hidden state.

Hyper-parameters, which basically determine a network structure and cannot be learned by an NN, are crucially significant parts of an ML model, because they can make an algorithm successful or unsuccessful. Learning rate, momentum, number of epochs, batch size and activation function are some of these parameters. The same kind of ML mechanisms may require different values of these parameters; thus, we need to tune them so that the model can optimally solve the ML problem. An epoch defines one forward pass and one backward pass of all the training while the batch size represents the number of examples that will be propagated through the network. It is advantageous to choose the batch size smaller than the number of all samples. The number of epoch decides how many times weights of the network are changed, and it is traditionally taken large (often hundreds or thousands). Activation functions are mathematical expressions that describe the output of a NN, and we can use different activation functions in a training process. Here, we consider the Rectifier Linear Unit (ReLU)[81], Leaky Rectifier Linear Unit (LReLU)[82], Exponential Linear Unit

(ELU)[83] and Scaled Exponential Linear Unit (SELU)[84] models, which are described by

$$\text{(ReLU, LReLU)} \qquad \Rightarrow \qquad \Theta_f = \begin{cases} \alpha x & \text{if } x \leq 0, \\ x & \text{if } x > 0, \end{cases} \tag{19}$$

$$\text{(ELU, SELU)} \qquad \Rightarrow \qquad \Theta_f = \lambda \begin{cases} \beta[\exp(x) - 1] & \text{if } x \leq 0, \\ x & \text{if } x > 0, \end{cases} \tag{20}$$

where $\alpha = 0$ for the ReLU, $\alpha = 0.01$ for the LReLU, $\beta = \lambda = 1$ for the ELU, and $\beta = 1.6733$ and $\lambda = 1.0507$ for the SELU.

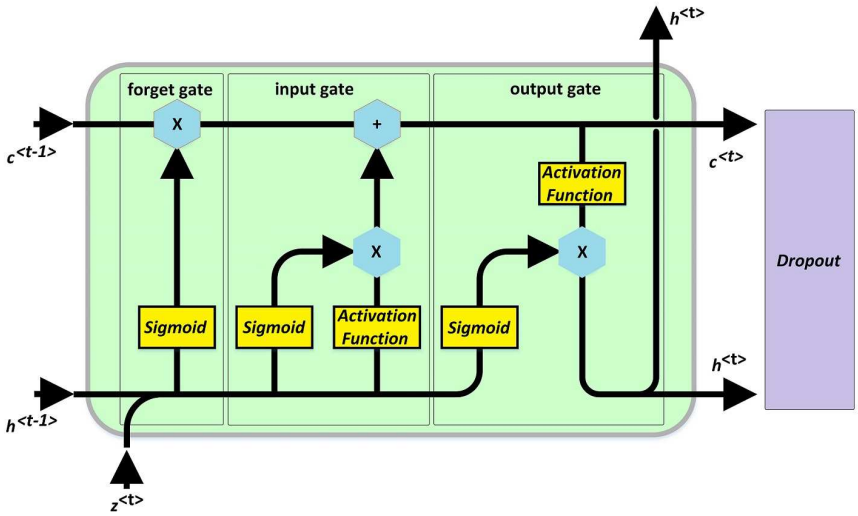

FIG. 4: The LSTM-Dropout architecture.

Furthermore, the LSTM mechanism (see FIG. 4) includes four main units: a cell, an input gate, a forget gate and an output gate[85, 86]. The gates organize the transfer of information into and out of the cell while the cell recalls values over optional time gaps. Here, we have[87]

$$i^{<t>} = \sigma(W_{zi}^T \odot z^{<t>} + W_{hi}^T \odot h^{<t-1>} + b_i), \tag{21}$$

$$f^{<t>} = \sigma(W_{zf}^T \odot z^{<t>} + W_{hf}^T \odot h^{<t-1>} + b_f), \tag{22}$$

$$o^{<t>} = \sigma(W_{zo}^T \odot z^{<t>} + W_{ho}^T \odot h^{<t-1>} + b_o), \tag{23}$$

$$\hat{p}^{<t>} = \Theta_f(W_{z\hat{p}}^T \odot z^{<t>} + W_{h\hat{p}}^T \odot h^{<t-1>} + b_{\hat{p}}), \tag{24}$$

$$c^{<t>} = f^{<t>} \otimes c^{<t-1>} + i^{<t>} \otimes \hat{p}^{<t>}, \tag{25}$$

$$H^{<t>} = h^{<t>} = o^{<t>} \otimes \Theta_f c^{<t>}, \tag{26}$$

where W denotes the weights of each layer, $\otimes$ is the direct product, the superscript T represents the transpose of the quantity where it is indicated and

$$\sigma = \frac{1}{1 + \exp(-x)} \tag{27}$$

is the sigmoid function. Equation (21) defines the input gate $i^{<t>}$, which helps the cell to know which values must be remembered in a long term. The forget gate $f^{<t>}$ reads the equation (22) and lets the cells to neglect particular values. The output gate $o^{<t>}$, which supervises which values will be the output and which not, is given by equation (23). The expression (24) defines a potential new case to recall in a long term. According to the LSTM algorithm, equations (25) and (26) represent the output of the approach. The mechanism will recall nothing about the former case if equation (25) indicates a null vector for the forget gate. Whereas, if it gives 1, the network will remember everything.

B. Applying the LSTM-Dropout network

The approach includes several technical hyper-parameters, and they can lead to a high possibility of over-fitting. To find out appropriate values of the batch size and the number of epochs and select the most suitable activation function, we can use the R-squared analysis that measures the goodness-of-fit of the trend. The R-squared statistics is mainly based on the following relation[88, 89]

$$R^2 = \frac{RSS}{TSS}, \tag{28}$$

where the Regression Sum of Squares is written as

$$RSS = \sum_i (\hat{H}_i - \overline{H})^2 \tag{29}$$

and the Total Sum of Squares is given by

$$TSS = \sum_i (H_i - \overline{H})^2. \tag{30}$$

Here $0 \leq R^2 \leq 1$ and H_i and $\hat{H}_i$ represent the theoretical and the estimated variables with respect to the independent variable z_i, respectively. Here, the case $R^2 = 1$ means there is no deviation between the set of estimated values and the actual theoretical data.

In the first part of the LSTM-Dropout analysis, we want to find best tuned values of hyper-parameters. Thus, we carry out the following steps;

- Construction of the mechanism: We use the Mean Squared Error (MSE) loss with the Adam optimizer and consider an LSTM cell including 100 neurons.

- Activation functions: We use the ReLU, LReLU, ELU and the SELU activation functions at hidden layers of the LSTM-Dropout architecture.

- Dropout probability: The putative commentary of the dropout hyper-parameter is the possibility of training a node selected in a layer, where the case 1.0 indicates no dropout while 0.0 implies that there is no outputs[90]. In order to get successful predictions, the value of dropout in a hidden layer is generally taken in the interval $[0.5, 0.8]$[90]. Hence, we use dropout probability of 0.5.

- Organizing the cosmological data: We produce a set of $H(z)$ values with the help of the relation (10) and order them from lower to higher red shift.

- Tuning the hyper-parameters: We run the network for each activation functions by using different values of the hyper-parameters. Here, we consider the range of $4i$, where i represents integers in $[1, 5]$, for the best batch size value and the range of $100j$, where j denotes integers in $[5, 15]$, to get the appropriate epochs number.

- The appropriate values of the hyper parameters for each activation functions are clearly revealed. Here, we should focus on the highest R-squared values (check the values presented with white color in FIG. 5).

TABLE II: Suitable settings used for producing the forecasts.

Activation function	Batch size	Number of epochs	z_{min}	z_{max}	Highest R-squared
ReLU	8	1000	0.071	2.36	0.94667
LReLU	4	1200	0.071	2.36	0.94734
ELU	8	1000	0.071	2.36	0.94617
SELU	4	1000	0.071	2.36	0.94849

- In TABLE II, we summarize our results.

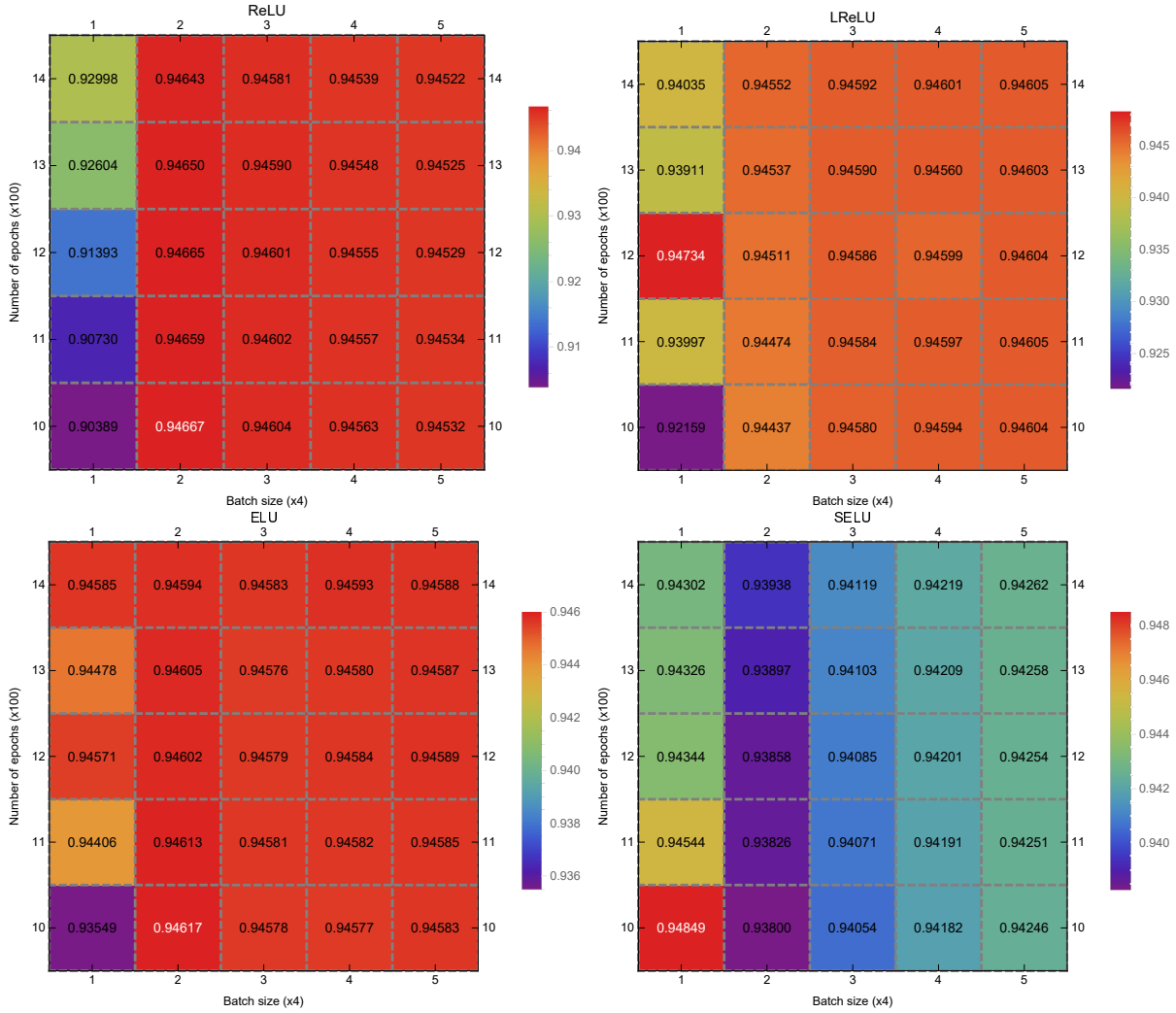

FIG. 5: R-squared analysis of the ReLU, LReLU, ELU and the SELU for different batch sizes and epoch numbers.

Now, we are in a position to obtain estimated values for the Hubble parameter. In the second part of the DL analysis, we follow the subsequent steps;

- Use the appropriate values of hyper-parameters, which were introduced in TABLE II, for the activation functions.
- Take the dropout probability of 0.5.
- Apply the LSTM-Dropout algorithm 500 times to obtain an uncertainty contour.
- Plot the estimated $H(z)$ data in the 1σ confidence region (see FIG. 6).

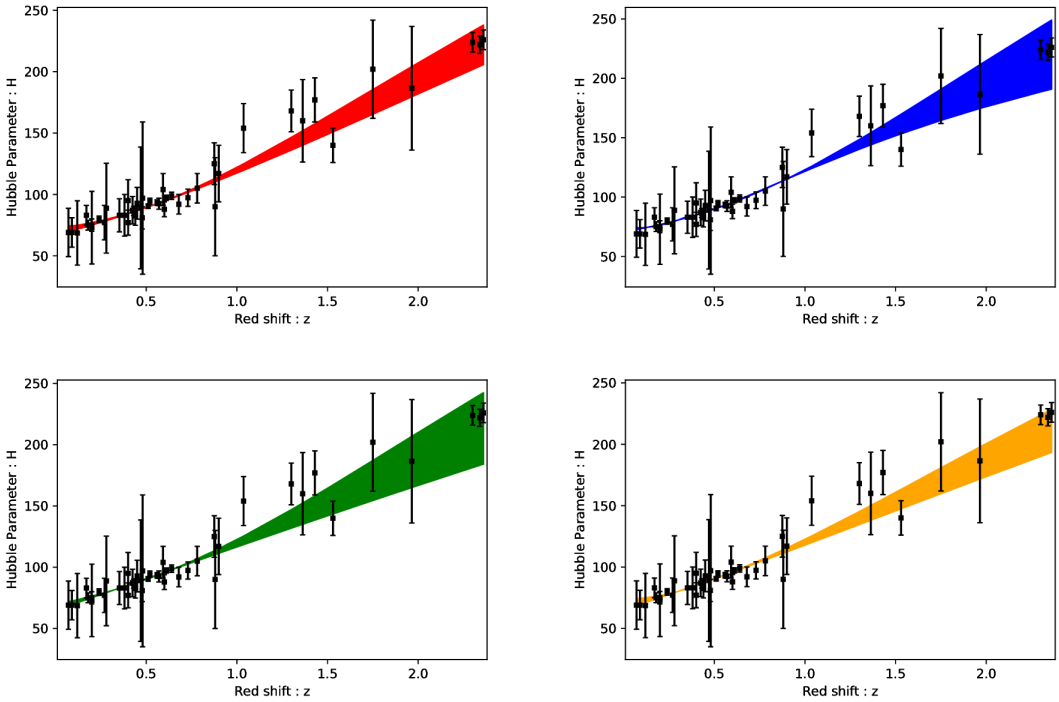

FIG. 6: $H \sim z$ relation for the ReLU (top left), LReLU (top right), ELU (bottom left) and the SELU (bottom right) activation functions for suitable values of the hyper-parameters. Here, black dots with error-bars represent observational data.

C. The FIM approach

The FIM is a fundamental quantity of deep neural networks with random weights and biases and it represents the characteristics of a stochastic model. On the other hand, the maximum likelihood parameter is one of the most significant and widely considered tools in statistics and the main idea behind the maximum likelihood forecast is to obtain the estimated values of the parameter that maximizes the likelihood of the sample data[91].

To study the distribution of a set of parameters θ under the conjecture of selected proposal M and given experimental data D, we need to forecast the posterior possibility[92]

$$\varepsilon(\theta) \equiv \epsilon(\theta|D;M). \tag{31}$$

This quantity can be defined in terms of the likelihood $Ł(\theta) \equiv \epsilon(\theta|D;M)$, the prior $\Pi(\theta) = \epsilon(\theta|M)$ and the Bayesian evidence $\Upsilon \equiv \epsilon(D|M)$ by the Bayes theorem[92] as

$$\varepsilon(\theta) = \frac{Ł(\theta)\Pi(\theta)}{\Upsilon}. \tag{32}$$

The Bayesian evidence Υ becomes a trivial normalising factor when we are focused on parameter forecast for a fixed proposal only. From this point of view, we neglect this quantity in further analysis. As a result, for a posterior probability distribution $\varepsilon(\theta)$, the FIM is written as[93, 94]

$$F_{\mu\nu} = -\frac{\partial^2 \ln \varepsilon(\theta)}{\partial\theta_\mu \partial\theta_\nu}. \tag{33}$$

where the set of parameters is $\theta_\mu = (H_0, \Omega_d^0, \eta, B, n)$. Note that the notions of likelihood and posterior distributions are interchangeable due to broad uniform priors which are taken into account for all parameters here. Furthermore, the curvature of the likelihood calculated at the mean is defined by the inverse of the FIM, and the mean is usually assumed to be in the same class with the maximum likelihood forecaster[92]. Cramér[95] and Rao[96], in 1945 and 1946, independently discussed the problem of finding a simple lower bound to the variance of point predictions. Consequently, in accordance with the Cramér-Rao inequality, the FIM yields an upper limit on the parameter error Δ of the parameter θ_μ[92]

$$\Delta\theta_\mu \leq \sqrt{(F^{-1})_{\mu\mu}}, \tag{34}$$

Features defined by equations (33) and (34) make the FIM a first-order optimistic approximation of the likelihood[92].

Now, we can apply the FIM approach to the cVCG model. In our investigation, the corresponding FIM is written as

$$F_{\mu\nu} = \begin{pmatrix} \frac{\partial^2\chi^2}{\partial H_0^2} & \frac{\partial^2\chi^2}{\partial H_0\partial\Omega_d^0} & \frac{\partial^2\chi^2}{\partial H_0\partial\eta} & \frac{\partial^2\chi^2}{\partial H_0\partial B} & \frac{\partial^2\chi^2}{\partial H_0\partial n} \\ \frac{\partial^2\chi^2}{\partial\Omega_d^0\partial H_0} & \frac{\partial^2\chi^2}{\partial\Omega_d^{0\,2}} & \frac{\partial^2\chi^2}{\partial\Omega_d^0\partial\eta} & \frac{\partial^2\chi^2}{\partial\Omega_d^0\partial B} & \frac{\partial^2\chi^2}{\partial\Omega_d^0\partial n} \\ \frac{\partial^2\chi^2}{\partial\eta\partial H_0} & \frac{\partial^2\chi^2}{\partial\eta\partial\Omega_d^0} & \frac{\partial^2\chi^2}{\partial\eta^2} & \frac{\partial^2\chi^2}{\partial\eta\partial B} & \frac{\partial^2\chi^2}{\partial\eta\partial n} \\ \frac{\partial^2\chi^2}{\partial B\partial H_0} & \frac{\partial^2\chi^2}{\partial B\partial\Omega_d^0} & \frac{\partial^2\chi^2}{\partial B\partial\eta} & \frac{\partial^2\chi^2}{\partial B^2} & \frac{\partial^2\chi^2}{\partial B\partial n} \\ \frac{\partial^2\chi^2}{\partial n\partial H_0} & \frac{\partial^2\chi^2}{\partial n\partial\Omega_d^0} & \frac{\partial^2\chi^2}{\partial n\partial\eta} & \frac{\partial^2\chi^2}{\partial n\partial B} & \frac{\partial^2\chi^2}{\partial n^2} \end{pmatrix}, \tag{35}$$

thus, we find the following bounds for the auxiliary parameters of the model,

$$\theta_\mu \Rightarrow \begin{cases} H_0 = 73.2134 \mp 8.13 \\ \Omega_d^0 = 0.9404 \mp 0.102 \\ \eta = 0.6412 \mp 0.33 \\ B = 2.4026 \mp 0.24 \\ n = 0.2615 \mp 0.22. \end{cases} \tag{36}$$

Furthermore, in FIG. 7, we display estimated $H(z)$ data obtained via the FIM mechanism. Additionally, it is found from the R-squared analysis of the FIM approach that $R^2 = 0.93432$.

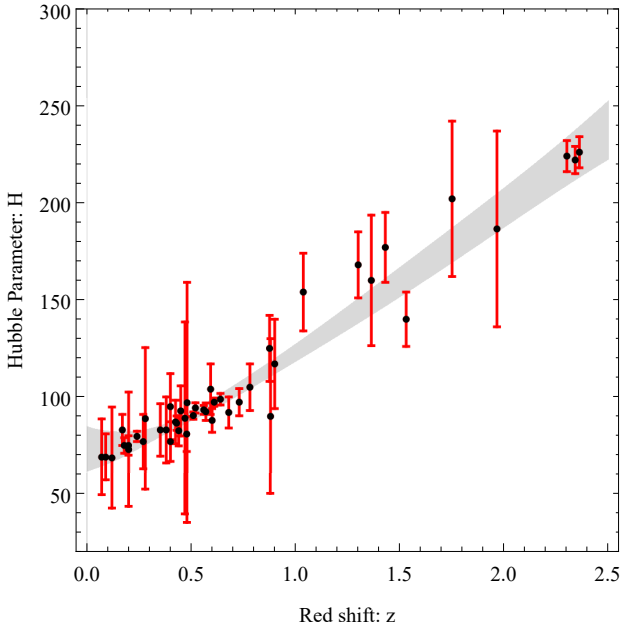

FIG. 7: Estimated $H \sim z$ data obtained via the FIM approach. Here, black dots with red error-bars denote astrophysical measurements.

IV. CLOSING REMARKS

It can be seen that scientists across extremely different fields are making use of the ANN to make their investigations much easier. The default concept behind such use of the ANN is the MLA, which is a subset of the ANN and enables the present-day computers to learn from a set of data.

In the present study, we have mainly constructed an exact expression for the cosmic Hubble parameter to explain the recent astrophysical data at the first step and showed that the DL algorithm can be used to make successful predictions for the cosmic chronometer observations. Here, we have obtained several interesting conclusions:

- $H(z)$ *function*: The constructed relation of the Hubble parameter with the best-fitting values of the arbitrary parameters is in very good consistency with the most recent astrophysical measurements.

- *Batch size and number of the epochs in the LSTM-Dropout approach*: Increasing the values of hyper-parameters does not always lead to more successful results. As a matter of fact, increasing these values may help with training the network faster, but the accuracy values in comparable cases are not similar. This conclusion agrees with literature[97].

- *Activation function in the LSTM-Dropout algorithm*: There are different types of activation function formulations in the RNN mechanism. Thanks to the $H \sim z$ relation, we have seen that linear activation functions produce successfully estimated values for the cosmic Hubble parameter.

- *The FIM mechanism*: This approach is relatively simpler than the LSTM-Dropout algorithm, but this method also has produced successful results.

The reliability level of the designed theoretical ground has been found to be nearly %95 for the LSTM-Dropout algorithm and %93 for the FIM approach, which means that reliable predictions can be made for the future of the universe by using the cVCG model. Furthermore, our method can be employed in future investigations to different CG and dark energy formulations.

Acknowledgements

We would like to thank all anonymous reviewers for giving such constructive comments which substantially helped improving the quality of the paper. We are also grateful to Gulkan Doner from Mersin University School of Foreign Languages for carefully reading of this paper.

---

[1] S. Perlmutter, et al., Nature 391 (1998) 51; Astrophys. J. 517 (1999) 565; Astrophys. J. 598 (2003) 102.

[2] P. de Bernardis, et al., Nature 404 (2000) 955.

[3] A.D. Miller, et al., Astrophys. J. Lett. 524 (1999) L1.

[4] N.A. Bahcall, J.P. Ostriker, S. Perlmutter and P.J. Steinhardt, Science 284 (1999) 1481.

[5] C.L. Bennett, et al., Astrophys. J. Suppl. 148 (2003) 1.

[6] S.L. Brile, O. Lahav, J.P. Ostriker and P.J. Steinhardt, Science 299 (2003) 1532.

[7] D.N. Spergel, et al., ApJs 148 (2003) 175.

[8] M. Tegmark, et al., Phys. Rev. D 69 (2004) 103501.

[9] P.A.R. Ade, et al., A&A 571 (2014) A16; A&A 594 (2016) A13.

[10] N. Aghanim, et al., A&A 641 (2020) A6.

[11] Y.-F. Cai, S. Capozziello, M. De Laurentis and E.N. Saridakis, Rep. Progr. Phys. 79 (2016) 106901.

[12] S. Capozziello, R. D'Agostino and O. Luongo, Int. J. Mod. Phys. D 28 (2019) 1930016.

[13] A. Yu. Kamenshchik, U. Moschella and V. Pasquier, Phys. Lett. B 511 (2001) 265.

[14] M. C. Bento, O. Bertolami and A. A. Sen, Phys. Rev. D 66 (2002) 043507.

[15] S. Chattopadhyay and S. Karmakar, Int. J. Geo. Met. Mod. Phys. 16 (2019) 1950101.

[16] H.B. Benaoum, O. Luongo and H. Quevedo, Eur. Phys. J. C 79 (2019) 577

[17] J. Lu, Phys. Lett. B 680 (2009) 404.

[18] V. Gorini, A.Yu. Kamenshchik and U. Moschella, Phys. Rev. D 67 (2003) 063509.

[19] D. Panigrahi and S. Chatterjee, JCAP 05 (2016) 052.

[20] D. Panigrahi and S. Chatterjee, Gen. Rel. Grav. 49 (2017) 35.

[21] E.O. Kahya, M. Khurshudyan, B. Pourhassan, R. Myrzakulov and A. Pasqua, Eur. Phys. J. C 75 (2015) 43.

[22] G. Abbas, et al., Astrophys. Space Sci. 357 (2015) 158.

[23] P. Bhar, M. Govender and R. Sharma, Prama J. Phys. 90 (2018) 5.

[24] P. Saha and U. Debnath, Eur. Phys. J. C 79 (2019) 919.

[25] O. Luongo and M. Muccino, e-Print: 2011.13590 [astro-ph.CO] (2020).

[26] E.E. Kangal, M. Salti and O. Aydogdu, Physics of the Dark Universe 26 (2019) 100369.

[27] C. Escamilla-Rivera, M.A.C.Quintero and S. Capozziello, JCAP 03 (2020)008.

[28] E.I. Zacharaki, et al., Magn. Reson. Med. 62 (2009) 1609.

[29] S. Mullainathan and J. Spiess, JEP 31 (2017) 87.

[30] P.M. Taga and J.E. Peak, Journal of Applied Meteorology 35 (1996) 714.

[31] Wolfram Research Inc., Mathematica 9.0, 2012.

[32] Python Software Foundation, Pyton 3.7.4, 2019.

[33] S.W. Allen, et al., Mon. Not. Roy. Astron. Soc. 353 (2004) 457.

[34] L.P. Chimento and A.S. Jakubi, Int. J. Mod. Phys. D 5 (1996) 71.

[35] A. Gonzales, T. Matos and I. Quiros, Phys. Rev. D 71 (2005) 084029.

[36] Z.K. Guo and Y.Z. Zhang. Phys. Lett. B 645 (2007) 326.

[37] J.A.S. Lima, M.O. Calvao and I. Waga, Cosmology: Thermodynamics and matter creation, Frontier Physics, Essays in Honor of Jayme Timno, World Scientific, Singapore (1990).

[38] M.O. Calvao, J.A.S. Lima and I. Waga, Phys. Lett. A 162 (1992) 223.

[39] N. Hulke, G. P. Singh, Binaya K. Bishi and A. Singh, New Astronomy 77 (2020) 101357.

[40] P.J.E. Peebles, Principle of physical cosmology, Princeton Univ. Press, Princeton (USA), (1993).

[41] J. Peacock, Cosmological physics, Cambridge University Press, Cambridge (UK), (1999).

[42] A. Aviles, C. Gruber, O. Luongo and H. Quevedo, Phys. Rev. D 86 (2012) 123516.

[43] P.K.S. Dunsby and O. Luongo, Int. J. Geo. Med. Mod. Phys. 13 (2016) 1630002.

[44] C. Zhang, et al., RAA 14 (2014) 1221.

[45] R. Jimenez, L. Verde, T. Treu and D. Stern, Ap. J. 593 (2003) 622.

[46] J. Simon, L. Verde and R. Jimenez, Phys. Rev. D 71 (2005) 123001.

[47] M. Moresco, L. Verde, L. Pozzetti, R. Jimenez and A. Cimatti, JCAP 7 (2012) 053.

[48] E. Gaztanaga, A. Cabre and L. Hui, MNRAS 399 (2009) 1663.

[49] M. Moresco, et al., J. Cosmol. Astropart. Phys. 5 (2016) 014.

[50] C. Blake, et al., MNRAS 425 (2012) 405.

[51] A.L. Ratsimbazafy, et al., MNRAS 467 (2017) 3254.

[52] D. Stern, R. Jimenez, L. Verde, M. Kamionkowski and S.A. Stanford, JCAP 02 (2010) 008.

[53] S. Alam, et al., MNRAS 470 (2017) 2617.

[54] Y. Wang, et al., MNRAS 469 (2017) 3762.

[55] L. Samushia, et al., MNRAS 429 (2013) 1514.

[56] M. Moresco, MNRAS 450 (2015) L16.

[57] N.G. Busca, et al., A&A 552 (2013) 18.

[58] T. Delubac, et al., A&A 574 (2015) A59.

[59] A. Font-Ribera, et al., JCAP 1405 (2014) 027.

[60] LIGO and VIRGO Collaborations, et al., Nature 551 (2017) 85.

[61] A.G. Riess, et al., ApJ 861 (2018) 126.

[62] S. Birrer, et al., MNRAS 484 (2018) 4726.

[63] A. J. Shajib, et al., MNRAS 494 (2020) 6072.

[64] C. Gruber and O. Luongo, Phys. Rev. D 89 (2014) 103506.

[65] V.V. Lukovic, R. D'Agostino and N. Vittorio, Astron. Astrophys. 595 (2016) A109.

[66] S. Capozziello, R. D'Agostino and O. Luongo, Int. J. Mod. Phys. D 28 (2019) 1930016.

[67] E. Hubble, Proc. Nat. Acad. Sci. 15 (1929) 168.

[68] M. Ntampaka et al., The Role of Machine Learning in the Next Decade of Cosmology, e-Print: 1902.10159 [astro-ph.IM].

[69] A. Mathuriya et al., "CosmoFlow: Using Deep Learning to Learn the Universe at Scale," SC18: International Conference for High Performance Computing, Networking, Storage and Analysis, Dallas, TX, USA, 2018, pp. 819-829. e-Print: 1808.04728 [astro-ph.CO].

[70] A. Almahairi, et al., Augmented CycleGAN: Learning Many-to-Many Mappings from Unpaired Data, Proceedings of the 35th International Conference on Machine Learning, PMLR 80 (2018) 195.

[71] H. Hoffmann and R. Moller, Unsupervised Learning of a Kinematic Arm Model. In: Artificial Neural Networks and Neural Information Processing - Lecture Notes in Computer Science, vol 2714, Springer, Berlin, Heidelberg (2003).

[72] J. Ebrahimi and D. Dou, Chain based RNN for Relation Classification, Human Language Technologies: The 2015 Annual Conference of the North American Chapter of the ACL, pages 1244–1249, Denver, Colorado, May 31-June 5, (2015).

[73] Y. Gal and Z. Ghahramani, Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning, Proceedings of the 33rd International Conference on Machine Learning, New York-USA, JMLR: W&CP Vol.48, (2016).

[74] I. Goodfellow, Y.Bengio and A. Courville, "Deep Learning (Adaptive Computation and Machine Learning series)", MIT Press, (2016).

[75] Y. Nesterov, Doklady ANSSSR (Soviet. Math. Docl.) 269 (1983) 543.

[76] J. Duchi, E. Hazan and Y. Singer, J. Mach. Learn. Res. 12 (2011) 2121.

[77] M.D. Zeiler, "ADADELTA: An Adaptive Learning Rate Method", e-Print: 1212.5701 [cs.LG].

[78] D.P. Kingma and J. Ba, "Adam: a Method for Stochastic Optimization", 3rd International Conference on Learning Representations (ICLR), San Diego-CA, USA, May 7-9 (2015), e-Print: 1412.6980 [cs.LG].

[79] T. Dozat, "Incorporating Nesterov Momentum into Adam". 4th International Conference on

Learning Representations, San Juan, Puerto Rico, May 2-4 (2016).

[80] S.J. Reddi, S. Kale and K. Sanjiv, "On the Convergence of Adam and Beyond". 6th International Conference on Learning Representations, Vancouver, Canada, Apr 30-May 3 (2018), e-Print: 1904.09237 [cs.LG].

[81] V. Nair and G.E. Hinton, Rectified linear units improve restricted boltzmann machines, Proceedings of the 27th International Conference on International Conference on Machine Learning, Haifa-Israel on June 21-24 (2010).

[82] A.L. Maas, A.Y. Hannun and A.Y. Ng, Rectifier nonlinearities improve neural network acoustic models, Proceedings of the 30th International Conference on Machine Learning, Atlanta-USA on June 16-June 21 (2013).

[83] D.-A. Clevert, T. Unterthiner and S. Hochreiter, Fast and accurate deep network learning by exponential linear units (ELUs), 4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, (2016), e-Print: 1511.07289 [cs.LG].

[84] G. Klambauer, T. Unterthiner, A. Mayr and S. Hochreiter, Self-Normalizing Neural Networks, 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA. e-Print: 1706.02515 [cs.LG].

[85] A. Graves, et al., IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (2009) 855.

[86] H. Sak, A. Senior and F. Beaufays, "Long Short-Term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling", 15th Annual Conference of the International Speech Communication Association, Singapore, September 14-18 (2014).

[87] W. Zaremba and I. Sutskever, "Reinforcement Learning Neural Turing Machines-Revised", e-Print: 1505.00521 [cs.LG].

[88] A.C. Camerona and F.A.G. Windmeijer, Journal of Econometrics 77 (1997) 329.

[89] R. Fonticella, "The Usefulness of the R2 Statistic", ACAS (1998) 55.

[90] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, J. Mach. Learn. Res. 15 (2014) 1929.

[91] S. Saxena and S. Zarrin, Int. J. Prob. Stat. 2 (2013) 1.

[92] L. Wolz, M. Kilbinger, J. Wellerb and T. Giannantonio, JCAP 09 (2012) 009.

[93] M.G. Kendall and A. Stuart, The Advanced Theory of Statistics. Vol. II Griffin, London U.K. (1969).

[94] M. Tegmark, A. Taylor and A. Heavens, Astrophys. J. 480 (1997) 22.

[95] H. Cramér, Skandinavisk Aktuarietidskrift 29 (1946) 85.

[96] C.R. Rao, Bull. Calcutta Math. Soc. 37 (1945) 81.

[97] S.L. Smith, P.-J. Kindermans, C. Ying and Q.V. Le, Don't Decay the Learning Rate, Increase the Batch Size, 6th International Conference on Learning Representatio, Vancouver Convention Center, Vancouver, BC, Canada (2018). e-Print: 1711.00489 [cs.LG].

View publication stats