IMPROVING PHOTOMETRIC REDSHIFTS BY VARYING ACTIVATION FUNCTIONS IN ARTIFICIAL NEURAL NETWORKS

IMDAD BINTI MAHMUD PATHI

UNIVERSITI SAINS MALAYSIA

2024

IMPROVING PHOTOMETRIC REDSHIFTS BY VARYING ACTIVATION FUNCTIONS IN ARTIFICIAL NEURAL NETWORKS

by

IMDAD BINTI MAHMUD PATHI

Thesis submitted in fulfilment of the requirements
for the degree of
Master of Science

December 2024

ACKNOWLEDGEMENT

I am deeply grateful to God, the Most Gracious and Merciful, for the opportunity to explore the universe through science. This journey has humbled me, reminding me of my own limitations and the vastness of creation. I owe immense thanks to Dr. John Soo Yue Han for his unwavering support and guidance throughout this research. His patience, wisdom, and integrity have been invaluable. You taught me every single thing. Your enthusiasm has inspired me to become the best version of myself. I will always remember your kindness and empathy during challenging times, particularly when my mom was sick until she recovered. I am extremely grateful for your support. You are a good role model for me.

I also extend gratitude to Prof. Enrique Gaztanaga for introducing me to the PAUS research group and to Prof. Carlton Baugh and Dr. Giorgio Manzoni for allowing me access to PAUS-GALFORM data and providing feedback as well as Assoc. Prof. Yoon Tiem Leong. Financial support from the Universiti Sains Malaysia through the Short Term Grant and Gra-Assist 2020 scheme, along with assistance from GREP MARA, has been crucial.

I am grateful to my academic mentors Dr Ramzun Maizan, Lance Orion Ong, Assoc. Prof. Iskandar Shahrim, Assoc. Prof. Mohd Zamir, Assoc. Prof. Azhar and Dr Nur Azwin for their guidance. My peers and friends have provided invaluable camaraderie and support. At the heart of my gratitude are my parents, Mahmud Pathi and Umi Kalthum, for their financial and emotional support, teaching me the essence of being a devout servant of God. My siblings and their spouses have been a source of endless love and encouragement. Thank you to everyone who has been part of this journey. Your collective support, kindness, and inspiration have been indispensable in reaching this milestone.

Sincerely, Imdad Al-Jarumi.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF SYMBOLS

α	Alpha
Å	Angstrom
θ	Angle in radians
β	Beta
Δ δ	Delta
$^\circ$	Degree
ε	Epsilon
η	Eta
$\dot{a}$	First derivative
∞	Infinity
lim	Limit
Λ λ	Lambda
μ	Mu
ν	Nu
Ω	Omega
Φ ϕ	Phi
π	Pi
pc	Parsec
σ	Sigma
$\ddot{a}$	Second derivative
c	Speed of light

LIST OF ABBREVIATIONS

ΛCDM	Λ Cold Dark Matter
4MOST	4-meter Multi-Object Spectroscopic Telescope
APO	Apache Point Observatory
ANN	Artificial Neural Network
BDT	Boosted Decision Tree
CEN	Creatio Ex-Nihilo
CMB	Cosmic Microwave Background
CVNN	Complex-Valued Neural Network
CAS	Catalogue Archive Server
CasJobs	Catalog Archive Server Jobs System
CFHT	Canada-France-Hawaii Telescope
DES	Dark Energy Survey
DESI	Dark Energy Spectroscopic Instrument
DR	Data Release
GALFORM	Galaxy Formation Model
GP	Gaussian Process
GA	Genetic Algorithm
HSC	Hyper Suprime-Cam
JWST	James Webb Space Telescop
KNN	K-Nearest Neighbor
KiDS	Kilo-Degree Survey
LRG	Luminous Red Galaxy

LSST	Legacy Survey of Space and Time
ML	Machine Learning
MLP	Multilayer Perceptron
MGS	Main Galaxy Sample
ORM	Roque de los Muchachos Observatory
PAUCam	PAU Camera
PFS	Subaru Prime Focus Spectrograph
PAUS	Physics of the Accelerating Universe Survey
PDF	Probability Density Function
QNN	Quantised Neural network
RF	Random Forest
SDSS	Sloan Digital Sky Survey
SVM	Support Vector Machine
SAS	Science Archive Server
SED	Spectral Energy Distribution
WHT	William Herschel Telescope
WFIRST	Wide-Field Infrared Survey Telescope

LIST OF APPENDICES

PENAMBAHBAIKAN ANGGARAN ANJAKAN MERAH FOTOMETRI DENGAN MENGUBAH FUNGSI PENGAKTIFAN DALAM RANGKAIAN NEURAL BUATAN

ABSTRAK

Mutakhir ini, kemajuan teknologi teleskop telah menyebabkan komuniti astronomi berhadapan dengan sejumlah data yang banyak sehingga mencapai kapasiti eksabait. Sehubungan dengan itu, kewujudan teknik pembelajaran mesin telah memudahkan tugas ahli astronomi menganalisis data-data tersebut. Salah satu daripada aplikasi pembelajaran mesin dan kaedah statistik yang relevan dalam bidang kosmologi ialah anggaran anjakan merah fotometri atau foto-z. Disebabkan kepentingannya yang jitu dalam bidang kosmologi, permintaan terhadap ketepatan foto-z semakin meningkat oleh kumpulan tinjauan langit seperti misi *Euclid* dan LSST untuk projek yang akan datang. Lantaran itu, ketepatan dan prestasi algoritma foto-z telah dipertingkatkan dengan mengadaptasi dan mengubah suai elemen-elemen penting yang terdapat dalam pembelajaran mesin. Algoritma "Artificial Neural Network Redshift" (ANNZ) adalah penganggar anjakan merah fotometri pembelajaran mesin yang pantas dan mudah digunakan. Salah satu elemen penting dalam rangkaian neural buatan (ANN) ialah fungsi pengaktifan. Ia digunakan sebagai unit untuk membuat keputusan dengan memperkenalkan ketidaklinearan di dalam model ANN agar keupayaan membezakan sesuatu adalah lebih baik serta berpotensi meningkatkan prestasi ANN jika dioptimumkan. Kami menguji prestasi ANNZ dengan memanipulasi fungsi pengaktifan iaitu menggantikan fungsi sigmoid logistik dengan fungsi tanh, Softplus, SiLU, ReLU, Leaky ReLU dan Mish. ANNZ dilatih menggunakan sampel "Luminous Red Galaxy" (LRG) dan Stripe-82 dari tinjauan "Sloan Digital Sky Survey" (SDSS) serta sampel dari tinjauan "Physics of the Accelerating Universe Survey" (PAUS). Kami juga menguji prestasi fungsi pengaktifan ini dengan mengubah kedalaman dan kelebaran seni bina ANN. Kajian kami menunjukkan bahawa tanh dan Leaky ReLU adalah fungsi pengaktifan yang

lebih berdaya saing serta stabil merentas seni bina dalam dan lebar dengan peningkatan >1 peratus dalam ralat σ_{RMS} dan ralat σ_{68} apabila diuji pada set data SDSS berbanding dengan tetapan asal yang digunakan dalam ANNZ. Dengan mengoptimumkan seni bina yang sesuai, kami mencapai peningkatan sebanyak 11 peratus dalam σ_{RMS} dan 6 peratus dalam σ_{68} apabila menggunakan tanh pada set data PAUS; peningkatan ketara sehingga 44 peratus, ANNZ memperoleh prestasi lebih baik daripada ANNZ2 jika kami menggunakan tanh dan seni bina yang lebih mendalam. Bukti ini menunjukkan bahawa algoritma ANNZ yang berusia 20 tahun ini perlu dikemaskini untuk membolehkannya kekal berdaya maju dan berdaya saing pada masa ini. Kod yang dikemaskini, dinamakan ANNZ+, boleh didapati secara umum di `https://github.com/imdadmpt/ANNzPlus`.

IMPROVING PHOTOMETRIC REDSHIFTS BY VARYING ACTIVATION FUNCTIONS IN ARTIFICIAL NEURAL NETWORKS

ABSTRACT

In recent years, the astronomical community has faced a data deluge of up to exabytes due to the advancement of telescope technology. In fact, the application of machine learning techniques has simplified the task of analysing data for astronomers. Photometric redshift estimation or photo-z is one of the relevant applications of machine learning and statistical methods in cosmology. Due to its significant contribution to cosmology, there is an increasing demand for precise photo-z for forthcoming astronomical surveys, including the *Euclid* mission and LSST. The accuracy and performance of the photo-z algorithm have been improved by adopting and modifying machine learning hyperparameters. The Artificial Neural Network Redshift (ANNZ) algorithm is a fast and simple machine learning photometric redshift estimator. One of the hyperparameters in the artificial neural network (ANN) is the activation function. It acts as a decision-making unit, which introduces non-linearity into the model, leading to better differentiation capabilities and potentially boosting the ANN's performance if optimally tuned. We test the performance of ANNZ by varying the activation functions, replacing the original logistic sigmoid with tanh, Softplus, SiLU, ReLU, Leaky ReLU and Mish. The training is demonstrated on the Luminous Red Galaxy (LRG) sample of the Sloan Digital Sky Survey (SDSS), Stripe-82 survey, and Physics of the Accelerating Universe Survey (PAUS). We also tested the performances of these activation functions by varying the depth and width of the ANN architectures. Our results show that tanh and Leaky ReLU are more competitive and stable across deeper and broader architectures with >1 per cent improvement of the root-mean-square error ($\sigma_{\rm RMS}$) and 68th percentile error (σ_{68}) when tested on the SDSS data sets as compared to the original setting used in ANNZ. With proper optimisation of the architecture, we achieved an improvement of 11 per cent in $\sigma_{\rm RMS}$ and 6 per cent in σ_{68} with tanh when we applied to the PAUS dataset; the performance is also shown to be up to 44 per cent better than ANNZ2 if we used tanh and

deeper architecture. This evidence gives good reason for an update to the 20-year-old algorithm ANNZ, allowing it to remain viable and competitive today. The updated code, named ANNZ+ is publicly available at `https://github.com/imdadmpt/ANNzPlus`.

CHAPTER 1

INTRODUCTION

In daily life, the human brain tends to compare what it sees in nature with something else in order to get an idea and categorise it into patterns of similarities or differences. During ancient times, the wanderers noticed the shiny and blinking objects in the night sky's darkness. The comparison between the dark and the light makes us wonder, what is it? How far is the object? Where are we? Are we the only intellectual creature in the world? Those questions, in light of the curiosities of ancient Greeks, came up with the idea to study the universe as a whole, cosmology (kosmos, "universe" and -logia, "study"). It is natural for humans to think the universe has an edge, as we can compare it with rooms that have walls, countries that have borders, and the atmosphere as the edge of a sphere surrounding the Earth. However, the biggest question in cosmology is whether the universe has boundaries remain unknown.

Seeds and Backman (2012) clearly describe the difference between the universe and the observable universe. The universe is everything that exists, and it could be infinite, while the observable universe, in contrast, is the small part that you can see from Earth using the most powerful telescopes. The observable universe is finite, with a limitation due to the abilities of humankind. It is estimated to be around 14 billion years old, which we can determine by looking for the oldest star. Here is where the study of the distance in the universe becomes crucial. The distance from us to the point estimate of the shiny astronomical objects are we called as redshifts. Now, let's dive into the world of redshifts.

1.1 Background of the Study

In recent years, the astronomical community has faced a data deluge of up to exabytes due to the enormous research to comprehend the cosmos (Rodríguez et al., 2022). Numerous planned and ongoing surveys such as Sloan Digital Sky Survey

(SDSS; York et al., 2000), Dark Energy Survey (DES; The Dark Energy Survey Collaboration., 2005), Euclid (Laureijs et al., 2011), Kilo-Degree Survey (KiDS; de Jong et al., 2013), Wide-Field Infrared Survey Telescope (WFIRST; Spergel et al., 2013), Dark Energy Spectroscopic Instrument (DESI; Levi et al., 2013), Subaru Prime Focus Spectrograph (PFS; Takada et al., 2014), Hyper Suprime-Cam (HSC; Aihara et al., 2017), Legacy Survey of Space and Time (LSST; Ivezić et al., 2019), and 4-meter Multi-Object Spectroscopic Telescope (4MOST; de Jong et al., 2019) are expected to observe a significant number of galaxies in both wide and deep areas. These surveys will be able to provide highly accurate measurements of the large-scale structure of the universe, a better understanding of the formation and evolution of the galaxies changed over time and the nature of dark energy and dark matter. Additionally, brand-new resources like the James Webb Space Telescop (JWST; Yan et al., 2023) allow the first batch of galaxies with $11 < z < 20$ to be observed. The latest discovery of early galaxies, using JWST medium band filters (Withers et al., 2023), highlights the crucial role of photometric redshifts in the study of the universe's structure. These surveys require good quality photometric redshifts (photo-zs) to attain their research target.

Spectroscopic redshifts, or spec-zs refer to the displacement of spectral lines measured by a spectrograph. These redshifts are generally determined by analysing the galaxy's spectrum, but the process requires a lot of telescope time and uses expensive facilities. This is where photo-zs come in: photo-z is a technique that allows us to estimate the distance to a galaxy based on its observed photometry. It is called "photometric" because it involves measuring the amount of light emitted by a galaxy at different wavelengths. After all, the observed wavelengths of light are typically redshifted, or shifted to longer wavelengths, as the light travels through expanding space (Koo, 1985; Loh & Spillar, 1986). Photo-zs are usually estimated by using broad-band magnitudes and there are two main methods which are spectral template fitting and empirical/machine learning techniques. The first approach commonly employs a likelihood utilising a χ^2 fit including codes like Le Phare (Arnouts et al., 1999), Bpz (Benítez, 2000), Hyperz (Bolzonella et al., 2000), Zebra (Feldmann et al.,

2006), EAZY (Brammer et al., 2008), GAZELLE (Kotulla & Fritze, 2009) and DELIGHT (Leistedt et al., 2023). The second approach employs machine learning methodologies, like self-organizing maps (Buchs et al., 2019), deep convolutional neural networks (Pasquet et al., 2019), neural networks (Lee & Shin, 2021), galaxy morphoto-z with neural networks (GAZNETS; Li et al., 2022), interpretable deep capsule networks (Dey et al., 2022), multi-task learning networks (Cabayol et al., 2023) and bayesian neural networks (Jones et al., 2024).

Artificial Neural Network Redshift (ANNZ[1]) introduced by Collister and Lahav (2004) was among the first machine learning photometric redshift algorithms that existed and were widely used by many; it was used for photo-z estimation (Bonfield et al., 2010; Collister et al., 2007) and furthermore concurrently with other photo-z codes (Abdalla et al., 2011; Bundy et al., 2015; Hildebrandt et al., 2010; Reis et al., 2012; Sánchez et al., 2014; Soo et al., 2018). It has also been employed for morphological classification of galaxies (Ball et al., 2004; Banerji et al., 2008; Lahav et al., 1996) and star-galaxy separation (Bertin & Arnouts, 1996; Kim & Brunner, 2016; Soumagnac et al., 2015). In a recent work, ANNZ is being used by Euclid Collaboration et al. (2023) in preparations for the Euclid project. ANNZ has been replaced by its successor ANNZ2[2] (Sadeh et al., 2016) which includes some quantification of epistemic uncertainties through an ensemble of randomised estimator techniques reminiscent of modern deep ensembles.

Despite that, ANNZ is fast, simple and lightweight compared to many other algorithms. The ANNZ algorithm has several parameters that can be adjusted to optimise its performance, such as its network architecture and random seed. However, Soo et al. (2021) demonstrated that artificial neural networks can yield poor results when used with high-dimensional narrowband inputs without comprehensive optimisation. Considering the significant advancements in the technology and theory of artificial neural networks, it raises the question: is ANNZ still competitive 20 years after its debut?

[1] `http://www.homepages.ucl.ac.uk/~ucapola/annz.html`
[2] `https://github.com/IftachSadeh/ANNZ`

Activation functions are mathematical functions that determine the output layer of a neuron in an artificial neural network (ANN) and transform the weighted input of a neuron into more linearly separable abstract features using a hierarchy of layers, enhancing the effectiveness of the ANN by creating non-linear relationships between variables. The most popular activation functions are the logistic sigmoid (Grossberg, 1982), hyperbolic tangent (tanh; LeCun et al., 1989), softplus (Dugas et al., 2001), rectified linear unit (ReLU; Nair and Hinton, 2010), leaky rectified linear unit (Leaky ReLU; Maas et al., 2013), exponential linear unit (ELU; Clevert et al., 2015), parametric rectified linear unit (PReLU; He et al., 2015), S-shaped rectified linear unit (SReLU; Jin et al., 2016), sigmoid linear unit (SiLU; Hendrycks and Gimpel, 2016), and mish (Misra, 2019). Dubey et al. (2022) and Jagtap and Karniadakis (2023) categorised these activation functions into three groups, which are fixed, adaptive and nonstandard activation functions. Despite the trial of many different activation functions, Santos-Olmsted et al. (2023) did not provide a conclusion on optimising the performance of the neural network for estimating photometric redshifts.

1.2 Problem Statement

In this era of precision cosmology, the spectroscopic method using advanced telescopes is the most reliable way to estimate redshifts and obtain high-quality stellar imaging. However, the acquisition of redshifts spectroscopically is a significantly expensive and time-consuming endeavour. Due to this problem, various photometric methods have been developed from extensive available photometric datasets to obtain redshifts, which we call photometric redshifts (photo-zs). Although spectroscopic redshifts (spec-zs) are known as truth redshifts, photometric redshifts are widely used in various cosmological research fields, including weak lensing, galaxy clustering, and large-scale structures.

Recently, an artificial neural network photometric redshift estimation showed poor results when dealing with high-dimensional inputs compared to other methods. Therefore, in this study, our interest is to enhance the precision and accuracy of photometric

redshift estimation using the artificial neural network method.

1.3 Research Objective

This research was conducted with the following objectives:

1. To improve the accuracy of photometric redshifts estimation by choosing the most stable and consistent activation function that can optimise the artificial neural network.

2. To gain valuable insights on how the different activation functions respond to the complexity of high dimensional input.

3. To assess the relevancy and competitiveness of ANNZ after varying the activation functions with other works.

1.4 The Scope of Study

Firstly, our primary focus is not on obtaining the absolute global minimum value of performance metrics because the most common approach to achieve the absolute global minimum is through trial and error, since specific parameters may not work for every sample. Instead, we are more interested in ensuring the consistency and stability of ANNZ+ under any condition and making them suitable for widespread use. We have tested various activation functions in ANNZ to observe their behaviour on different samples, various network architectures, random seed variations, and the effectiveness of each activation function. We study the impact of the logistic sigmoid, tanh, Softplus, ReLU, Leaky ReLU, SiLU and Mish activation functions on the photometric redshift performance in Luminous Red Galaxy (LRG), the Stripe-82 and Physics of the Accelerating Universe Survey (PAUS) - Galaxy Formation Model (GALFORM) samples. The PAUS-GALFORM sample is used to test the effectiveness of the activation functions on high dimensional input training, and we gained valuable insights on how the different activation functions respond to the complexity of high dimensional input.

1.5 Thesis Contribution

Since ANNZ debuted way before many of these activation functions were introduced, we were interested in testing the viability and competitiveness of ANNZ by "updating" it with the choice of these more up-to-date activation functions while in the process analyse their performances as well. ANNZ is a machine learning algorithm coded from scratch using C++, therefore it gives us the flexibility to insert and test any known activation function in the literature, not limited to those made available in public machine learning packages. The study of activation functions can significantly impact the performance and improve the accuracy of a model that can be used for regression or classification tasks. Hence, different activation functions are well suited to different types of problems and network architectures. The modified code, named ANNZ+ produces competitive results when compared with the latest photo-z in the literature, which can be useful for the photo-z community.

1.6 Outline of the Thesis

This thesis consists of six chapters as follows. Chapter 1 is the introduction of the study, which comprises the background of the study, the problem statement, the research objectives, the scope of the study and the contribution of this research. In Chapter 2, we explore the literature review related to the thesis. We will start by briefly discussing the big picture of cosmology and the various theories related to it, followed by an explanation of the distance ladder, redshifts, galaxy photometry and spectroscopy. Next, with all the theories needed to understand this study, we proceed to discuss our main topic: photometric redshifts. We discuss photo-z's in detail, as well as the various efforts to improve photo-z estimations. In the following Chapter 3, we are introducing a photo-z algorithm called ANNZ and describing its algorithms. We list down the use of ANNZ over the years and the development of ANNs. We also introduce the important components of ANNs, which are activation functions and their characteristic. Besides, we describe the metrics used to evaluate our photo-z performances. In Chapter 4, we describe the sources of data sets used and any selection cuts applied to these samples.We

present answers and discussions for objectives (1), (2) and (3) in Chapter 5. Our work concludes in Chapter 6 which also includes suggestions for future work.

CHAPTER 2

LITERATURE REVIEW

2.1 Cosmology

Cosmology is the study of the universe as a whole that includes the origin, formation and evolution of the universe over time (Coles, 2001). The concept of the universe traditionally addressed by various religions, involves the idea of the universe originating from nothing, also known as *Creatio Ex-Nihilo* (CEN). This idea is central to the religious understanding of the creation of the universe, where it asserts that God created the universe from nothing (Soskice, 2010). Many religious traditions and theological doctrines have delved into the philosophical and metaphysical implications of this concept, shaping their perspectives on the nature of existence and the divine.

In the early 20th century, the present era of modern cosmology has made significant advancements from the development of the theory of relativity and observational instruments. These instruments have played a crucial role in gathering and analysing data in order to validate and refine the theoretical framework of cosmology. By collecting observational data from distant celestial objects and phenomena, we have been able to test and support various cosmological theories, leading to a deeper understanding of the universe and its evolution. In this work, we will put aside the theological perspectives since human brain capabilities are limited, and we focus more on using mathematics in fundamental physics to describe the universe. For derivations, we follow the approach taken by Ryden (2003).

2.1.1 Cosmological Principle

The cosmological principle states that on a large-scale structure involving scales of 100 million light-years ($\geq$ 100 Mpc), the distribution of matter and energy in the universe is homogenous and isotropic. Homogeneity means that the distribution of matter is the same in any region of the universe, indicating no special locations, while

isotropy means that the universe looks the same in any direction on a large scale, signifying no special directions in our universe.

2.1.2 FLRW Metric

Einstein's field equations are the differential equations in his theory of general relativity that describe the fundamental interaction of gravitation that causes spacetime to be curved by matter and energy. In cosmology, these equations are crucial because they provide the theoretical framework for understanding the large-scale structure and dynamics of the universe.

Relativity teaches us that space and time together make up a four-dimensional spacetime. This information contained in a *metric tensor*. In the context of a Euclidean universe, the *Minkowski metric* (Minkowski, 1909) is used to represent the simplest form of a metric tensor:

$$ds^2 = -c^2dt^2 + dx^2 + dy^2 + dz^2, \tag{2.1}$$

which becomes more convenient when written in spherical coordinates. (r, θ, ϕ):

$$ds^2 = -c^2dt^2 + dr^2 + r^2\left[d\theta^2 + \sin^2\theta d\phi^2\right], \tag{2.2}$$

and even more compact if we define the angular part as $d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2$:

$$ds^2 = -c^2dt^2 + dr^2 + r^2d\Omega^2, \tag{2.3}$$

where $-c\,dt$ represents the temporal contribution with c denoting the speed of light, and dx, dy, dz indicates the spatial contribution. Here, r, θ, ϕ signify the radial distance, polar angle, and azimuthal angle of spherical coordinates. The Minkowski metric applies to a static universe with the 4-dimensional Euclidean spacetime. However, it does not explain our universe.

The appropriate metric that describes a homogeneous, isotropic and expanding universe involves the incorporation of spatial curvature and the expansion of the universe

into our metric. The spatial curvature can be described by including a factor $S_k(r)$ in the angular part, $d\Omega$ of Equation 2.3. This convention is called the hyperspherical coordinates. This factor assumes three types of spatial curvature: closed (spherical shape, $k = 1$), flat ($k = 0$), and open (hyperbolic or saddle shape, $k = -1$).

$$S_k(r) = \begin{cases} R\sin(r/R), & k = +1 \\ r, & k = 0 \\ R\sinh(r/R), & k = -1, \end{cases} \tag{2.4}$$

where k is the curvature constant and R is the curvature radius ($R \rightarrow \infty$ for flat universe). The measurement of Cosmic Microwave Background (CMB) radiation, discovered by Penzias and Wilson (1965), proved that our universe is flat followed by the latest results from Tristram et al. (2024) which showed our universe appears consistently to be spatially flat to a 1σ accuracy of 0.16%. In this work, we assume the type of spatial curvature is a flat universe, setting k = 0.

The universe has been proven to be expanding, as demonstrated by Hubble (1929), and accelerating, as shown by Riess et al. (1998) and the Supernova Cosmology Project (SCP; Perlmutter et al., 1999). This is represented by the condition that the derivative of the scale factor with respect to time, an expanding universe has $\frac{d}{dt}a(t) = \dot{a}(t) > 0$, while an accelerating universe has $\frac{d^2}{dt^2}a(t) = \ddot{a}(t) > 0$. In this work, we apply the expansion of the spatial components to the metric, which involves introducing a scale factor $a(t)$. We set this parameter to be $a(t_0) = 1$ at the present moment.

With the addition of curvature and expansion, the Minkowski metric (Equation 2.3) becomes the *Friedmann-Lemaître-Robertson-Walker* metric (Friedman, 1922; Lemaître, 1931; Robertson, 1935; Walker, 1937):

$$ds^2 = -c^2dt^2 + a(t)^2\left[dr^2 + S_k(r)^2 d\Omega^2\right], \tag{2.5}$$

where the distance has been expressed as the product of a time-dependent scale factor, denoted as $a(t)$, and a time-dependent comoving coordinate, r.

2.1.3 Einstein Field Equation of General Relativity

Later on, this metric will evolves according to general relativity. Einstein's field equations (Einstein, 1916) calculate the geometry of spacetime produced by a given distribution of mass and energy:

$$G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}, \tag{2.6}$$

where the left-hand side describes the geometry of space and time. $G_{\mu\nu}$ is the Einstein tensor, describing the curvature of spacetime. The indices μ, ν run over the time coordinate and the three spatial coordinates, while $g_{\mu\nu}$ is the metric tensor. The right hand side includes the matter-energy terms, like $T_{\mu\nu}$ is the stress-energy tensor that calculates the effect of a given distribution of mass and energy on the curvature of spacetime while G is the gravitational constant, $G = 6.67x10^{-11}\mathrm{m}^3\mathrm{kg}^{-1}\mathrm{s}^{-2}$. The Einstein tensor $G_{\mu\nu}$ is defined as:

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R, \tag{2.7}$$

where $R_{\mu\nu}$ is the Ricci tensor, which measures how volume changes on the surface when we move along its curvature and R, the Ricci scalar, compares the volume deviation between an Euclidean space and the said spacetime.

2.1.4 Friedman Equations

Applying the Einstein field equation of general relativity to Equation 2.5 resulting the first *Friedman* equation (Friedman, 1922) which describe the expansion of the universe:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3c^2}\varepsilon(t) - \frac{kc^2}{(R_0)^2 a(t)^2}, \tag{2.8}$$

where ε denotes the energy density encompassing all components of the universe, namely matter, radiation, and the cosmological constant which can be expressed as $\varepsilon = \varepsilon_{\mathrm{mat}} + \varepsilon_{\mathrm{rad}} + \varepsilon_{\Lambda}$. R_0 is the curvature radius. If we rearrange the equation by normalising the energy densities to the critical energy density for which the Universe

would be flat today, and making explicit their dependence on the scale factor, we modify the Friedmann equation (2.8) to take the form

$$\frac{H^2(t)}{{H_0}^2} = \frac{\Omega_{\text{rad},0}}{a^4(t)} + \frac{\Omega_{\text{mat},0}}{a^3(t)} + \Omega_\Lambda + \frac{1 - \Omega_{\text{tot},0}}{a^2(t)}, \tag{2.9}$$

where the Hubble parameter, $H(t) = \frac{\dot{a}(t)}{a(t)}$, which defines the expansion rate of the universe. $H_0 = H(t_0)$ is the Hubble parameter in the current era, also known as the *Hubble* constant (Hubble, 1929) which the value of $H_0 = (67.64 \pm 0.52)\ \text{kms}^{-1}\text{Mpc}^{-1}$, it is consistent with recent Planck results (Tristram et al., 2024). The density parameter $\Omega(t) = \frac{\varepsilon(t)}{\varepsilon_c(t)}$ are the energy densities normalised to the critical energy density for a flat universe today, i.e. $\Omega_{\text{rad},0} = \frac{\varepsilon_{\text{rad},0}}{\varepsilon_{c,0}}$, $\Omega_{\text{mat},0} = \frac{\varepsilon_{\text{mat},0}}{\varepsilon_{c,0}}$ and $\Omega_{\Lambda,0} = \frac{\varepsilon_\Lambda}{\varepsilon_{c,0}}$ with $\varepsilon_{c,0} = \frac{3c^2{H_0}^2}{8\pi G}$. $\Omega_{\text{tot},0}$ is the total density parameter today and is made up of the sum of the individual component density parameters: $\Omega_{\text{tot},0} = \Omega_{\text{rad},0} + \Omega_{\text{mat},0} + \Omega_\Lambda$. The last term of Equation 2.9 is related to the curvature. We should keep in mind the possibility that the curvature might be nonzero but with the consistent result in Planck Collaboration et al. (2020), we assume the flat universe, which $\Omega_{\text{tot},0} = 1$:

$$\frac{H^2(t)}{{H_0}^2} = \frac{\Omega_{\text{rad},0}}{a^4(t)} + \frac{\Omega_{\text{mat},0}}{a^3(t)} + \Omega_\Lambda. \tag{2.10}$$

Equation 2.10 describes the evolution of the different components in the universe, each with the appropriate power of the scale factor. As expected, we can see that the densities of matter and radiation decrease in an expanding universe at different rates. We have the three spatial dimensions for matter density, which dilute the energy density as the universe expands. When considering radiation, it's important to take into account the additional impact of photons and neutrinos, which serve to increase the scale factor. Consequently, the expansion factor seems to behave as the fourth power, leading to further dilution. The unexpected behaviour comes from the energy density of the cosmological constant, Λ which does not show any dependence on the scale factor and does not dilute. Since the density of the cosmological constant does not change to keep the density constant, the energy needs to increase as a^3, to counteract the increase in volume.

Equation 2.9 implies the expansion of the universe has been influenced by different components (known as cosmological parameters) during various eras. In the initial phases of the universe, radiation was dominant in the universe, a period known as the radiation-dominated era. Following this, the dynamics of the universe shifted as matter began to overcome radiation, leading to the matter-dominated era. As the energy density of the cosmological constant does not diminish, the ultimate era of cosmic expansion is one dominated by the cosmological constant - an era in which we presently exist.

The Λ Cold Dark Matter (ΛCDM) model proposes the existence of two key components in the universe. The cosmological constant Λ represents dark energy, a mysterious force responsible for the accelerating expansion of the universe. On the other hand, CDM stands for cold dark matter, a slow-moving and weakly interacting form of matter that exerts gravitational influence but does not interact electromagnetically. Originally, in Einstein's field equation, Λ was used to explain a static universe, but now Λ denotes dark energy. The model assumes that the majority of the matter in the universe is in the form of dark matter and this dark matter is made of non-relativistic particles, i.e. with velocities considerably lower than the speed of light.

The cosmological parameters measured by Planck Collaboration et al. (2020) are $\Omega_{m,0} = 0.32 \pm 0.01, \Omega_{\Lambda,0} = 0.68 \pm 0.01$ and $\Omega_{k,0} = 0.001 \pm 0.002$. The matter density parameter $\Omega_{m,0}$ is the total of the baryon density $\Omega_{b,0}$ and cold dark matter density $\Omega_{c,0}$, which separately yield $\Omega_{b,0} = 0.0224 \pm 0.0001$ and $\Omega_{c,0} = 0.120 \pm 0.001$. The value of $\Omega_{\Lambda,0}$ seems to have decreased from the previous measurement (Planck Collaboration et al., 2016), where the previous value is $\Omega_{\Lambda,0} = 0.69 \pm 0.01$. Λ is known to be constant. However, according to the recent discovery by the DESI Collaboration et al. (2024), there are potentially interesting differences that could indicate that dark energy is evolving with time. What exactly is dark energy? The answer remains unknown.

To sum up, Einstein's field equations connect the geometry of the universe, expressed through the curvature of spacetime and can be applied to model the universe's evolution over time, including the current observed expansion. They are used to derive

the Friedmann equations, which describe the expansion of the universe in homogeneous and isotropic models. These models are foundational to modern cosmological theories, including the standard ΛCDM model that describes the universe as composed of dark energy, dark matter, and ordinary matter. This relationship is central to cosmology, as it allows us to understand how the distribution of matter and energy affects the past, present, and future structure as well as the expansion of the universe.

2.1.5 The Cosmological Redshift

Galaxies, supernovae, and quasars are essential celestial entities that provide valuable insights into the evolution and dynamics of the universe. In this work, our focus will be on measuring the distances of galaxies. The cosmological redshift, z is used to determine the distance of distant galaxies. It is defined as a phenomenon in which the wavelength of electromagnetic radiation from an object is stretched as the object moves away from us. This effect is a result of the expansion of the universe, causing the light from distant objects to appear more redshifted as they move farther away. On the other hand, the Doppler redshift occurs when there is relative motion between the source of light and the observer, causing a shift in the frequency of the light waves.

The cosmological redshift differs from the Doppler redshift. This concept is a key component of understanding the large-scale structure and expansion of the universe. We can define mathematically the cosmological redshift as:

$$z = \frac{\lambda_{\mathrm{o}} - \lambda_{\mathrm{e}}}{\lambda_{\mathrm{e}}}, \tag{2.11}$$

where λ_{o} is the wavelength of light received by the observer and λ_{e} is the wavelength of light emitted by the source. Any mention of 'redshift' from this point onwards will refer to the cosmological redshift. When examining a galaxy spectrum, it is typical to see emission or absorption lines of particular elements shifted from their expected wavelengths. This shift can be easily measured when the galaxy's spectrum is available.

To relate the redshift, z and the scale factor $a(t)$, we consider the light travelled along a null geodesic with $ds = 0$ and $d\Omega = 0$. From the FLRW metric in Equation 2.5, we find:

$$c\frac{dt}{a(t)} = dr. \tag{2.12}$$

Considering the light from a distant galaxy at the time t_e with wavelength λ_e, and reaches us at t_0 with wavelength λ_0, we integrate this expression to find:

$$\frac{\lambda_{\mathrm{e}}}{a(t_{\mathrm{e}})} = \frac{\lambda_{\mathrm{o}}}{a(t_0)}. \tag{2.13}$$

We compare the definition of redshift in Equation 2.11 with Equation 2.13 and rearrange to obtain the scale factor in terms of redshift:

$$a(z) = \frac{1}{1+z}. \tag{2.14}$$

Equation 2.14 suggests that redshift can be used as an alternative way of quantifying the relative size of the universe. As an example, at $z = 2$, Equation 2.14 shows that $a = \frac{1}{3}$, so the universe would have been one-third its current size. The scale factor, which is linked to time by definition, allows us to use the redshift to determine when specific events took place.

2.1.6 Magnitudes and Fluxes

When measuring the redshift of an object, astronomers rely on important data such as magnitudes and fluxes. Magnitude is defined as the measure of the brightness of an object. The concept of magnitudes originates from ancient Greece, representing the effort to classify the brightness of stars. The magnitude system is a method of measuring the brightness of celestial objects using a logarithmic scale where lower numbers indicate brighter objects which mirrors the sensitivity of the human eye which is also logarithmic.

To measure the light from a distant galaxy, it's more practical to measure the amount of photons on a linear scale rather than a logarithmic scale. We can use charge-coupled

devices (CCDs) to measure the photon count emitted by the source per unit area and time, also known as the flux:

$$f = \frac{L}{4\pi d_L^2}, \tag{2.15}$$

where L is the luminosity of the source and d_L is the luminosity distance to the light source. As mentioned earlier, the brightness (luminosity) scale of magnitudes is logarithmic, meaning that a change in 1 magnitude corresponds to the change in brightness by 2.51 times (or a change in 5 magnitudes corresponds to the change in brightness by 100 times). The magnitude system can be divided into two types: *apparent magnitude*, m, and *absolute magnitude*, M:

$$m - m_{\text{ref}} = -2.5 \log_{10}\left(\frac{f}{f_{\text{ref}}}\right) \qquad M - M_{\text{ref}} = -2.5 \log_{10}\left(\frac{L}{L_{\text{ref}}}\right), \tag{2.16}$$

where f_{ref} is the reference flux and L_{ref} is the reference luminosity. The relation between an object's apparent magnitude and its absolute magnitude can be expressed in the form

$$M = m - 5\log\left(\frac{d_L}{10\text{pc}}\right). \tag{2.17}$$

It is calculated as the difference between the apparent magnitude m and the absolute magnitude M. This method provides a more convenient way to express distances, especially when dealing with objects that are relatively close to Earth.

2.1.7 Spectroscopy and Photometry

The most accurate way to determine redshift is through spectroscopy, which is defined as the technique of splitting light (which is a form of electromagnetic radiation) into its constituent wavelengths, creating what is known as a spectrum. This technique allows us to analyse the shifts in wavelength and accurately measure the redshift of an objects. This technique involves using a spectrograph to disperse an object's light into a spectrum. Ground-based surveys typically measure light wavelengths in nanometers or Angstroms. For example, Sloan Digital Sky Survey (SDSS) distinguishes between stars, galaxies, and quasars based on their spectra (Stoughton et al., 2002). Spectroscopic

observations are crucial for differentiating between celestial objects in images and for determining an object's distance from Earth, as well as its physical properties such as mass, age, composition, and velocity. The information is referred to as spectroscopic data.

The other technique to estimate the redshift is called photometry. Photometry is the technique that measures flux (the amount of electromagnetic radiation energy) of an object that goes through a passband or filter that blocks light except for a wavelength region. The filters larger than 300 Å are known as broadband filters, and those smaller as narrowband filters. These data are referred to as photometric data ("photo" = light; "metric" = measurement).

2.2 Photometric Redshift

The process of obtaining spectroscopic data requires a lot of telescope time and uses expensive facilities. Even with the advent of spectrographs with a huge multiplex factor, like Dark Energy Spectroscopic Instrument (DESI) which can target 5 000 objects in one pointing, deeper observations can reach much higher surface densities of galaxies than this, requiring many passes of the same area to achieve high spectroscopic completeness. This is where photometric redshift (photo-z) estimation comes in: photo-z is a technique that allows us to estimate the distance to a galaxy based on its observed photometry.

It is called "photometric" because it involves measuring the amount of light emitted by a galaxy at different wavelengths. The filters used to image the galaxies can be seen as sparse spectra, which can be used to obtain photo-z. After all, the observed wavelengths of light are typically redshifted, or shifted to longer wavelengths, as the light travels through expanding space (Koo, 1985; Loh & Spillar, 1986). Photo-z values are typically estimated using broadband magnitudes. In this study, we also incorporated the use of narrowband magnitudes. There are two main methods to obtain photo-z: spectral template fitting and machine learning techniques.

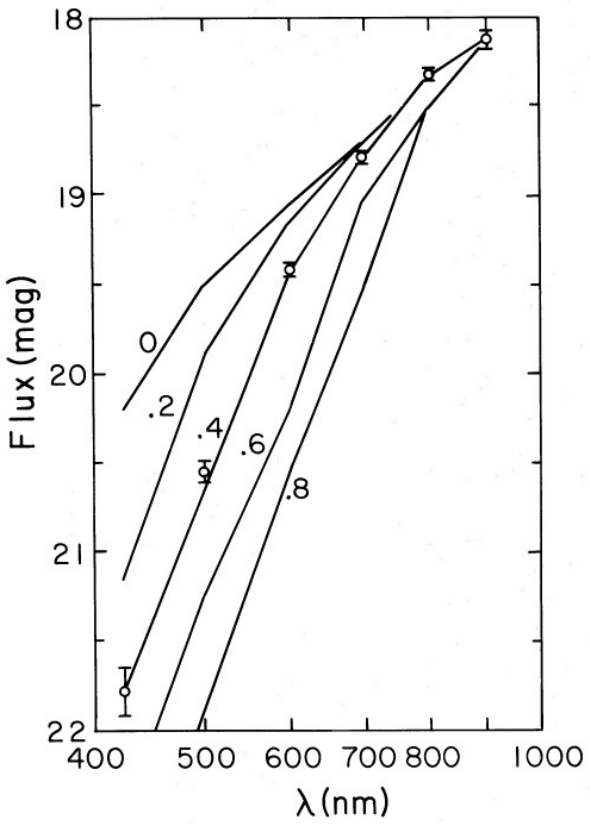

Figure 2.1: The original figure from Loh and Spillar (1986) illustrates the SED template of an elliptical galaxy highlighting how the flux is expected to change at different labeled redshifts. The measured flux of the observed galaxy is shown as open circles. By aligning it with the template, particularly near the line $z = 0.4$ line, one can obtain the photo-z estimate.

2.2.1 Template Fitting Method

Template fitting is the method of utilising a collection of predefined typical theoretical galaxy spectra, known as templates, to match the photometry of the galaxy. The photo-z of the galaxy is obtained using the best-matched template. It was first implemented by Puschell et al. (1982), who obtained photo-z estimates of faint radio galaxies using Spectral Energy Distribution (SED) templates. Later on, the method was improved by Loh and Spillar (1986) who were able to obtain photo-z estimates for thousands of galaxies and adapted the method into the template fitting we used today.

The method can be visualised using Figure 2.1, taken from Loh and Spillar (1986). It shows the SED template of an elliptical galaxy template and how the flux of the galaxy is expected to change at different redshifts. The measured flux of the observed galaxy

is shown as open circles and acts as an approximate measure of the underlying SED. As they are seen to lie close to the line at $z = 0.4$, it resulted in a photo-z estimate of $z = 0.398 \pm 0.018$ for this galaxy (which had a spectroscopic redshift of $z_{\rm spec} = 0.390$).

Examples of template fitting codes are LE PHARE (Arnouts et al., 1999), HYPERZ (Bolzonella et al., 2000), BPZ (Benítez, 2000), ZEBRA (Feldmann et al., 2006), EAZY (Brammer et al., 2008), GAZELLE (Kotulla & Fritze, 2009) and DELIGHT (Leistedt et al., 2023).

2.2.2 Machine Learning Method

The second approach employs Machine Learning (ML) methodologies. The original definition of machine learning by Samuel (1959) stated that it is a field of study that gives computers the ability to learn without being explicitly programmed. The latest definition by Jordan and Mitchell (2015) states that a computer program is said to learn from experience E with respect to some task T and performance measure P, if its performance on T, as measured by P, improves with experience E.

The use of traditional methods has evolved to incorporate machine learning techniques for the photo-zs. Firth et al. (2003) is widely regarded as the trailblazer in the development of machine learning-based photo-z methods. They employed an Artificial Neural Network (ANN) to predict the photo-z of 20,000 galaxies from SDSS, achieving a root-mean-square error of $\sigma_{\rm RMS} \sim 0.021$ for galaxies with redshift $z < 0.35$. Their pioneering work inspired the creation of the first publicly accessible machine learning photo-z algorithm, known as Artificial Neural Network Redshifts (ANNZ; Collister and Lahav, 2004).

This code has been selected for use throughout the entire project. Since that time, many machine learning algorithms, including Boosted Decision Trees (BDTs), Gaussian Processes (GPs), K-Nearest Neighbors (KNNs), Support Vector Machines (SVMs), Random Forests (RFs) and Genetic Algorithms (GAs) have been developed, in addition to ANNs. Certain algorithms have been integrated with Probability Density

Functions (PDFs), and some algorithms have the capability to incorporate a few or all machine learning methods in a single algorithm. Table 2.1 provides a list of machine learning algorithms organised by year.

2.2.3 Artificial Neural Network Redshift

Artificial Neural Network Redshift (ANNZ[1]) introduced by Collister and Lahav (2004) was among the first machine learning photometric redshift algorithms produced and widely used; it was used for photo-z estimation (Bonfield et al., 2010; Koukoufilippas et al., 2020) concurrently with other photo-z codes (Abdalla et al., 2011; Bundy et al., 2015; Euclid Collaboration et al., 2020; Hildebrandt et al., 2010; Reis et al., 2012; Sánchez et al., 2014; Soo et al., 2018) and produced photo-z catalogues for various different surveys (Bilicki et al., 2016; Collister et al., 2007; Koukoufilippas et al., 2020). It has also been employed for morphological classification of galaxies (Ball et al., 2004; Banerji et al., 2008; Lahav et al., 1996) and star-galaxy separation (Bertin & Arnouts, 1996; Kim & Brunner, 2016; Soumagnac et al., 2015).

ANNZ has been replaced by its successor ANNZ2[2] (Sadeh et al., 2016) which includes some quantification of epistemic uncertainties through an ensemble of randomised estimator techniques reminiscent of modern deep ensembles. Despite the latest updates, ANNZ is fast, simple and lightweight compared to many other algorithms. The ANNZ algorithm has several parameters that can be adjusted to optimise its performance, such as its network architecture and random seed.

In recent work, ANNZ was used in Euclid Collaboration et al. (2020) as one of the 13 photometric redshift code to assess the strengths and weaknesses of current photo-zs methods for *Euclid* mission focusing particularly on the $0.2 - 2.6$ redshift range that the *Euclid* mission will probe in order to provide precise and accurate photo-z measurements. The results showed ANNZ was able to provide reliable single value estimates for redshifts but they faced challenges in producing useful PDF for

[1] `http://www.homepages.ucl.ac.uk/~ucapola/annz.html`
[2] `https://github.com/IftachSadeh/ANNZ`

Table 2.1: The compilation of the machine learning techniques used for estimating photo-z, arranged chronologically by year, along with main references for accessing the algorithms.

Introduced By	Machine Learning Techniques
Firth et al. (2003)	Artificial neural networks
Collister and Lahav (2004)	Artificial neural networks
Wadadekar (2005)	Support vector machines
Way and Srivastava (2006)	Gaussian process regression
Ball et al. (2007)	K-nearest neighbours
Oyaizu et al. (2008)	Artificial neural networks
Freeman et al. (2009)	Spectral connectivity analysis
Carliles Carliles et al. (2010)	Random forests
Gerdes et al. (2010)	Boosted decision trees
Laurino et al. (2011)	Weak gated experts
Carrasco Kind and Brunner (2013)	Prediction tress, random forests with PDF
Krone-Martins et al. (2014)	Symbolic regression
Graff et al. (2014)	Generic neural networks
Bonnett (2015)	Classification neural networks
Hoyle et al. (2015)	Artificial neural networks, BDT & ADABOOST
Elliott et al. (2015)	Generalized linear models
Hogan et al. (2015)	Genetic algorithms
Rau et al. (2015)	Ordinal class PDF estimation
Almosallam et al. (2016)	Sparse gaussian processes
Sadeh et al. (2016)	Integrated ML methods with PDF
De Vicente et al. (2016)	Directional Neighbourhood Fitting with PDF
Hoyle (2016)	Deep neural networks
Cavuoti et al. (2017)	Integrated ML methods with PDF
Samui and Samui Pal (2017)	Clustering back propagation neural networks
Buchs et al. (2019)	Self organizing maps
Pasquet et al. (2019)	Deep convolutional neural networks
Beck et al. (2020)	Artificial neural networks
Zhou et al. (2021)	Convolutional neural networks
de Diego et al. (2021)	Artificial neural networks
Razim et al. (2021)	Artificial neural networks
Lee and Shin (2021)	Integrated ML methods
Li et al. (2022)	Artificial neural networks
Dey et al. (2022)	Deep capsule networks
Cabayol et al. (2023)	Multi-task learning networks
Jones et al. (2024)	Bayesian neural networks

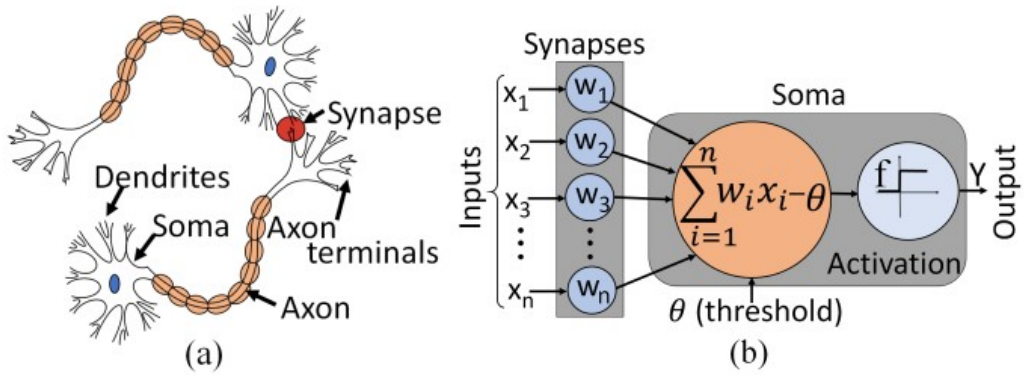

Figure 2.2: The original figure from Maheshwari et al. (2022) illustrates the similarities between biological neurons and artificial neurons. (a) Biological neurons connected by synapses and the neuron body are called Soma. (b) The basic artificial neurons block diagram where synapses are modelled as programmable weights and the soma represents the summation and the activation function.

the validation sample sources. ANNZ also experienced sharp drops in performance metrics above redshift $z \approx 1.3$ which it struggled with maintaining performance at higher redshifts but ANNZ showed good results and performed generally better than the template-fitting ones in the lower redshift. Despite various machine learning algorithms, after two decades, the ANNZ is still in use today.

However, Soo et al. (2021) demonstrated that artificial neural networks can yield poor results when used with high-dimensional narrowband inputs without comprehensive optimisation. Euclid Collaboration et al. (2022) mentioned they have not fully explored all the possible combinations of the number of neurons and hidden layers using DLNN and CNN. This is the opportunity of this work to provide insight since ANNZ has the ability to do it. The significant recent advances in the technology and theory of artificial neural networks raise the question: is ANNZ still competitive 20 years after its debut? Does ANNZ need to be updated?

2.3 The Activation Functions in the Artificial Neural Network

In this work, we choose to study the activation function, which is one of the hyperparameters in the ANN. Basically, the ANN is a machine learning algorithm that imitates how neurons in a human's brain learn. McCulloch and Pitts (1943) was

the pioneer in developing the first computational model for neural networks, while Rosenblatt (1958) invented the first ANN known as a perceptron. Biologically, in the neuron system, the electrical signal which carries information is transmitted from the dendrites to the axon terminals through the axon body. The electrical signals continue transmitting across the synapse from one neuron to another. A similar mechanism is applied to the ANN, where the neural network nodes receive input data, process it, and pass it to the next node until the end node predicts outcomes. The interpreted area in the neuron cell is called soma, known as the cell body (refer Figure 2.2). The soma receives all the information from the dendrites and sends a signal to the axon if it is strong enough. The soma mimics the activation functions in ANNs (Maheshwari et al., 2022).

Activation functions are the mathematical equations that act as decision-making units that determine whether the summation of the equation should be activated or not. It introduces non-linearity into the model, leading to better differentiation capabilities to determine whether the node inputs are relevant or not for the network. The activation functions are attached to each node in the ANN. Analogously, in a big family, individuals are often presented with varying opinions from their family members, and it is crucial to note that not all opinions are applicable to their circumstances. To avoid making decisions based on invalid or irrelevant information, individuals need to filter their opinions and assess their relevance to their requirements before taking appropriate action. Relative to the analogy, we need the human brain's decision-making unit in the ANN.

2.3.1 Desired Characteristics of the Activation Functions

Determining the most suitable activation function for a neural network involves a complex set of considerations rather than a universal rule. The choice of activation function varies based on the specific nature of the problem at hand, including factors such as the type of data being processed, the network architecture, and the desired output. The following are the essential characteristics that any activation function

should possess to effectively contribute to the neural network's performance.

1. Non-linearity: The non-linear nature of the activation function enables the model to learn the complex relationship between the network's input and output, allowing the model to capture the patterns within the dataset. This characteristic has significantly improved the learning capability of neural networks. Cybenko (1989) and Hornik et al. (1989) suggested that the activation function must be bounded, non-constant, monotonic and continuous to ensure the network's ability to approximate any continuous function to any desired degree of accuracy. Morita (1993, 1996) found that neural networks utilising non-monotonic activation functions exhibit improved memory capacity and retrieval ability. Sonoda and Murata (2017) demonstrated that neural networks with unbounded but non-polynomial activation functions (e.g., ReLU) can act as universal approximations. If the activation function is linear, everything will collapsed into one even though we have many hidden layers.

2. Computational efficiency: The activation function used in neural networks should be designed to be computationally efficient based on the time taken to compute the model, the computational cost, and energy consumption.

3. The vanishing and exploding gradient problems: These are vital problems of activation functions. The variation in the inputs and outputs of certain activation functions, such as the logistic sigmoid function (see Section 2.3.3(a)), is extremely large. Sigmoid reduces and converts a larger input space into a smaller output space range between 0 and 1. As a result, the back-propagation algorithm has almost no gradients to propagate backwards in the network, and any remaining gradients that do exist continue to diminish as the algorithm progresses down through the top layers. Due to this, the initial hidden layers are left with no information about the gradients. For hyperbolic tangent and sigmoid activation functions, it has been observed that the saturation region for large input (both positive and negative) is the major reason behind the vanishing of the gradient. To