BIBECHANA
Vol. 23, No. 1, January 2026, 40-47

ISSN 2091-0762 (Print), 2382-5340 (Online)
Journal homepage: http://nepjol.info/index.php/BIBECHANA
Publisher:Dept. of Phys., Mahendra Morang A. M. Campus (Tribhuvan University)Biratnagar

A study of cosmic void properties within redshift $z \leq 0.114$ through analytical shape descriptors and deep latent representations

Santosh Poudel[1], Daya Nidhi Chhatkuli[2,*], Binil Aryal[1]

[1]*Central Department of Physics, Tribhuvan University, Kathmandu, Nepal*
[2]*Department of Physics, Tri-Chandra Multiple Campus, Ghantaghar, Kathmandu, Nepal*
*Corresponding author. Email: chhatkulidn@gmail.com

Abstract

In this paper we examined the distribution and morphology of 518 cosmic voids from the SDSS DR7 REVOLVER catalog using a combination of deep learning and conventional statistics. Using ellipticity, prolateness, and radial alignment metrics, we describe the shapes of voids and find that larger voids have increasing anisotropy but decreasing ellipticity (r=-0.526). There is no discernible change in void size or alignment with redshift according to statistical tests. We processed six void properties by a deep autoencoder, which retained 86.62% of the variance. Our combined method offers fresh insights into the dynamics of the cosmic web by confirming that void shapes are shaped by boundary confinement and tidal forces.

Keywords: cosmic voids, void morphology, large-scale structure, autoencoders, statistical cosmology, Sloan Digital Sky Survey.

Article information
Manuscript received: July 7, 2025; Revised: September 30, 2025; Accepted: October 9, 2025
DOI https://doi.org/10.3126/bibechana.v23i1.81147

1 Introduction

Cosmic voids are vast under-dense regions of the universe surrounded by filaments and clusters that form the cosmic web [1–3]. Unlike galaxies and clusters that form through gravitational collapse, voids evolve through the expansion of their boundaries. When a photon of light traverses through cosmic voids, due to the expansion of the universe, its wavelength gets stretched, which reduces their energy as they exit the void. This phenomenon, known as the Integrated Sachs-Wolfe effect, helps to probe dark energy [4]. Furthermore, voids' relative isolation from non-linear gravitational effects allows them to act as pristine laboratories in cosmological studies. These unique aspects make cosmic voids valuable for cosmological research.

Statistics plays a crucial role in characterizing the properties and understanding the role of cosmic voids in the cosmic web. With the advent of machine learning, this field is being revolutionized by providing innovative methods for data analysis, classification, and prediction [5]. Machine learning algorithms excel at processing huge amounts of data, which is beneficial because data from galaxy surveys are incredibly large and intricate. The algorithms can identify subtle features of the data that traditional statistical tools might overlook.

Moreover, machine learning techniques can extract information from void features for cosmological parameter inference [6]. Integrating machine learning with traditional statistical approaches,

combining data from voids with information from galaxy clusters, is yielding tighter constraints on cosmological parameters than either method alone [5].

In this paper, we examine and characterize void shapes and distribution in our catalogue by analyzing their morphology both by using traditional statistics and deep autoencoder architecture. We begin in section 2 by detailing the dataset and methodology, section 3 presents morphological properties and latent space representations of voids, and finally in section 4 we discuss the cosmological implications of our integrated approach.

2 Data and Methods

2.1 Data

In this paper, we used the updated void catalogs derived from the Sloan Digital Sky Survey Data Release 7 (SDSS DR7) main galaxy sample [7]. These catalogs include voids identified using three void-finding algorithms–VoidFinder, VIDE, and REVOLVER–applied to a volume-limited subsample within redshift $z \leq 0.114$. The catalogs provide effective radii, centers, and other void properties explained in Table 2. For our analysis, we focus on the catalog generated using the REVOLVER algorithm under the Planck 2018 cosmological model.

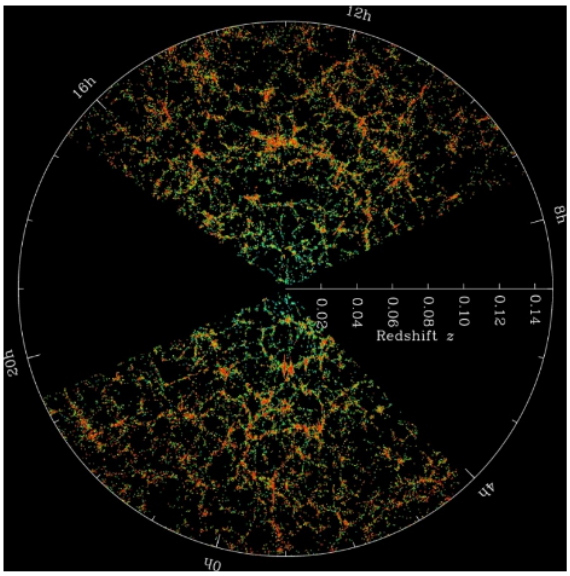

Figure 1: The cosmic web as shown by Sloan Digital Sky Survey [8].

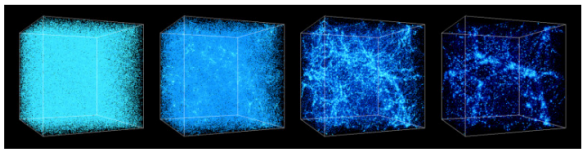

Figure 2: Formation of the cosmic web in an expanding universe as shown by computer simulations [9].

The variables and data structure of our void catalog are shown in Table 2. The catalog contains 518 cosmic voids with properties including spatial coordinates, effective radii, ellipsoid axes, and surface areas.

A sample of the data structure is shown in Table 1.

2.2 Shape Characterization Methods

We quantify void morphology of our catalogue using four principal metrics derived from the ellipsoidal decomposition:

Firstly we do Ellipsoid Axes Computation, here we compute Axis lengths from the ellipsoid components:

$$L_i = \sqrt{x_i^2 + y_i^2 + z_i^2} \quad (i = 1, 2, 3) \tag{1}$$

We also sort axes as $a \geq b \geq c$ (major, intermediate, minor)

Secondly, we do the calculation of shape metrics which includes voids ellipticity, prolateness and its radial alignment. The ellipticity is computed by using the following formula [10, 11].

$$\epsilon = 1 - \frac{c}{a} \tag{2}$$

Ellipticity quantifies deviation from spherical symmetry. We compute Prolateness using the following formula [12, 13].

$$P = 0.25 \frac{(b^2 - a^2) + (b^2 - c^2)}{a^2 + b^2 + c^2} \tag{3}$$

Prolateness basically measures elongation along major axis of cosmic voids present in our catalogue. We used Prolateness as a feature in Latent representation.

Finally we compute the radial alignment using the following formula.

$$\cos\theta = \frac{\mathbf{a} \cdot \mathbf{r}}{||\mathbf{a}|| \cdot ||\mathbf{r}||} \tag{4}$$

Where $\mathbf{r}$ is line-of-sight vector. Radial alignments metric is generally used for the tests for redshift-space distortions.

For the statistical analysis of our computed metrics, we used Pearson correlations and t-tests to assess relationships between: Size ($\log R$) vs. ellipticity (ϵ), Redshift (z) vs. radial alignment ($\cos\theta$), Size ($\log R$) vs. anisotropy ($\Delta\lambda$)

Table 1: Data Table (Partial View)

x	y	z	redshift	ra	dec	x1	x2
-193.038	-150.676	186.001	0.105	217.974	37.220	-9.478	-33.720
-242.970	-20.971	189.490	0.105	184.933	37.847	-12.221	-27.321
-147.355	70.726	98.225	0.065	154.360	-14.883	30.099	...
-257.322	78.918	126.873	0.102	197.050	25.238	-16.105	-23.424
-168.552	91.000	229.828	0.102	151.635	50.191	-9.067	36.506
... (other columns omitted for brevity)							

Table 2: Void Catalogue Data Structure

Column	Description	Unit
x, y, z	Cartesian coordinates of void's weighted center	h^{-1} Mpc
redshift	Redshift of void's weighted center	–
ra	Right Ascension of void's weighted center	degrees
dec	Declination of void's weighted center	degrees
radius	Effective radius of the void	h^{-1} Mpc
x1, y1, z1	Cartesian components of void's first ellipsoid axis	h^{-1} Mpc
x2, y2, z2	Cartesian components of void's second ellipsoid axis	h^{-1} Mpc
x3, y3, z3	Cartesian components of void's third ellipsoid axis	h^{-1} Mpc
area	Total surface area of the void	$(h^{-1}\,\mathrm{Mpc})^2$
edge	Surface area of void adjacent to boundary cells	$(h^{-1}\,\mathrm{Mpc})^2$

2.3 Machine Learning Algorithms

2.3.1 Working of an autoencoder

An autoencoder is a type of artificial neural network used to learn efficient representations for a set of data through unsupervised learning [14]. To put it more simply, it basically does the job of describing large and complex data with the help of fewer features. The architecture of an autoencoder comprises three components: the encoder, the bottleneck or latent space, and the decoder. The encoder maps the input complex data to the latent space (or code). The bottleneck layer that holds the latent space is of lower dimension compared to the input data. Finally, the decoder reconstructs the information received from the latent space. On a basic level, an autoencoder duplicates the input data but with the help of fewer features.

2.3.2 Latent Representation Learning

Now, to extract meaningful patterns from multidimensional void properties, we implemented a deep autoencoder architecture which as explained above learns compressed representations of void characteristics. To collectively characterize the size, shape, orientation and environmental context of voids, we selected six key void properties namely logarithmic radius ($\log R$), ellipticity (ϵ), prolateness (P), radial alignment ($\cos\theta$), boundary fraction (f_b), and shape anisotropy ($\Delta\lambda$) for our model to process. These features were minimum-maximum scaled to normalize their ranges before the processing.

As we know symmetric encoder and decoder pathways make up the autoencoder architecture. Our specific architecture (64-32-8) was selected through an empirical process of testing various configurations (e.g., 128-64-32, 32-16-8). This chosen setup offered the optimal balance between model capacity and performance on our validation set. Using rectified linear unit (ReLU) activations, the encoder converts the input characteristics through two fully-connected layers with 64 and 32 neurons, respectively. These two layers provided sufficient non-linear representational power to learn from our 6 input features. As a result, the fundamental structure of the void attributes is captured in an 8-dimensional latent space representation. The 8-dimensional bottleneck was specifically chosen as it effectively compressed the data while being low-dimensional enough to prevent overfitting and facilitate clear visualization and interpretation of the latent space. This design is mirrored in the decoder pathway, which uses 32 and 64-neuron layers to rebuild the original inputs from the latent representation, with linear activation in the output layer.

The model was trained using the Adam optimizer with a learning rate of 0.001 and mean squared error as the reconstruction loss function. We implemented early stopping with a patience of 10 epochs based on validation loss to prevent overfitting. The training utilized an 80/20 random split

of the void catalog, with batch processing of 64 samples per iteration.

To validate the advantage of our non-linear autoencoder approach, we also performed a comparative analysis with Principal Component Analysis (PCA), with results presented in Section 3.4.

2.3.3 Principal Component Analysis

Principal Component Analysis (PCA) is an unsupervised machine learning algorithm used to reduce higher-dimensional data to fewer dimensions while retaining as much variance as possible. PCA projects data onto principal components, which are new axes oriented along directions of maximum variance. This technique simplifies the dataset, making patterns easy to analyze while preserving essential information. In our analysis, the PCA usage is strictly limited to visualization of the autoencoder's latent space when its dimensionality exceed two [15, 16].

3 Results and Discussion

3.1 Void Size Distribution and Correlation Analysis

The void size distribution and radius versus redshift relationship are shown in Figure 3. We detect no significant size evolution across cosmic time: voids at low-z ($z < z_{\text{median}}$) show statistically indistinguishable size distributions from high-z counterparts ($p = 0.8653$, two-sample t-test).

3.2 Shape Analysis Results

3.2.1 Shape Characterization Results

The morphological analysis as explained above in methods section is showing systematic patterns in void geometry. As shown in Fig. 4a, the ellipticity distribution (1 - c/a) peaks prominently between 0.60-0.75, which tells us that moderate asphericity is most prevalent in our catalogue. The radial alignment distribution (Fig. 4b) is symmetric about $\cos\theta = 0$, which tells us that there is no preferential orientation with respect to the line-of-sight.

A significant correlation emerges between void size and morphology: voids with ($\log R > 0.08$ Mpc/h) are seen exhibiting higher ellipticities Fig. 4d) but as they grow larger their ellipticity is decreasing. We got the correlation value of $r = -0.526$ between size and ellipticity and the p value for this correlation was extremely low. Concurrently, shape anisotropy ($e_1 - e_3$) namely the difference of larger and smaller eigen values follows a unimodal distribution centered at ∼21 (Fig. 4c). The trend in the bottom right plot shows that as voids get larger, their anisotropy also tends to increase. Basically, the plot tells us that smaller voids are more uniform in shape, while larger voids becomes more elongated.

But, Notably, radial alignment is showing no evolution across redshifts($p = 0.76$) (Fig. 4e), which is directing us to the conclusion that orientation properties may remain consistent over cosmic time but however since our void catalogue redshift is upto only 0.114, we cannot definitively confirm this conclusion.

3.3 Latent Representation

The Figure 5 shows the training history of our autoencoder model, with the training and validation loss plotted over each epoch. The x-axis, which represent epochs shows the number of times the model had gone through the entire training data set. Reconstruction loss is plotted in y-axis which basically tells us how much error there is in the model's predictions. As we can see, the loss is decreasing steeply at the beginning i.e. at the first few epochs and then it flattens out and reaches a plateau, which means that the model is learning quickly at first and then stabilizing as it improves.

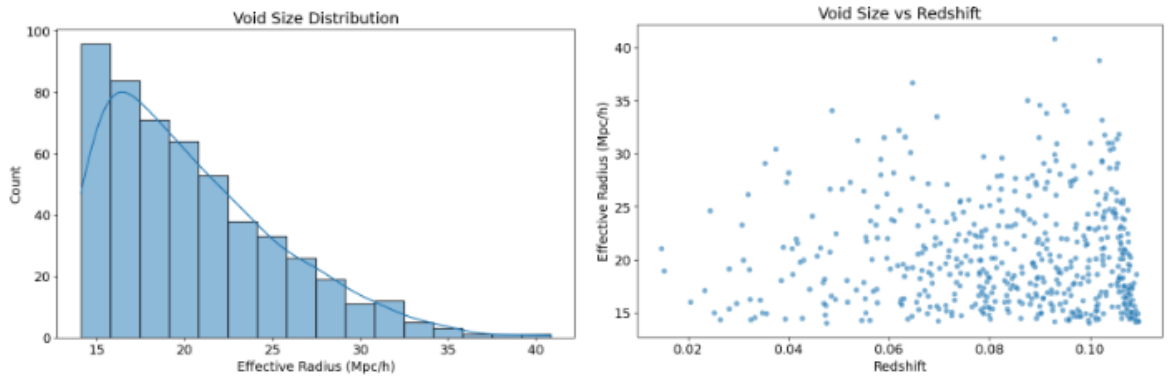

Figure 3: Void properties analysis: (left) void size distribution showing the prevalence of moderate-sized voids, (right) void radius versus redshift with weak correlation.

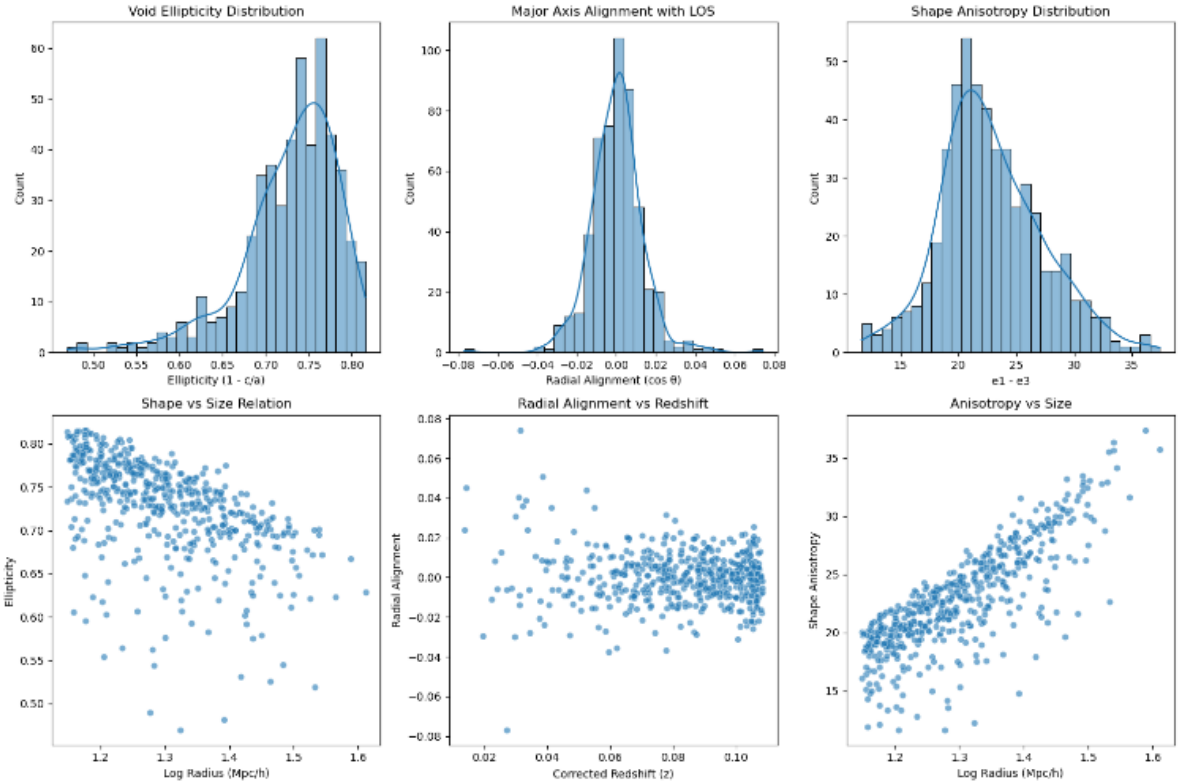

Figure 4: Morphological properties: (a) Ellipticity distribution, (b) Line-of-sight alignment, (c) Shape anisotropy, (d) Size-ellipticity relation, (e) Alignment evolution with redshift, (f) Anisotropy versus size.

The latent space visualization of the features learned by our autoencoder model is shown in Figure 6. We obtained a very high informative latent space, with 86.62% variance. This result is improvement over Principal Component Analysis (PCA) because PCA only captured 63.79% of the variance when applied to the same six void properties. Now lets discuss the subplots of Figure 6 one by one.

The size coded subplot shows a distinctive spatial organization where larger voids represented by darker colours corresponding to higher log radius values, is seen having a trend to cluster in specific regions of the latent space. The observed clustering pattern here tells us that the autoencoder has learned to associate void size with other attributes and hence is able to create a natural segregation between small and large void populations. The smooth gradient observed in the colour transitions shows that size variations are continuous rather than discrete. This result hints us the hierarchical formation scenario where voids grow through merging process.

The ellipticity-coded visualization reveals compelling patterns in void shape distribution within the latent space. The most stretched out i.e. highly elliptical voids, shown in darker regions with $\epsilon > 0.7$, form distinct clusters. This tells us that aspherical voids share common characteristics beyond just their elongated geometry. The intermediate ellipticity regions i.e between very stretched and more rounded voids; is showing gradual transitions, supporting our earlier findings that void asphericity follows a continuous distribution rather than exhibiting bimodal characteristics. The spatial separation of different ellipticity regimes in latent space provides evidence that shape is a fundamental organizing principle in void property space.

The radial alignment subplot demonstrates a remarkably uniform distribution across the latent space, with $\cos\theta$ values showing no preferential clustering or spatial organization. This plain, uniform spread matches our statistical finding that voids point in completely random directions. It also confirms there's no significant distortion in how we measure their positions due to cosmic expansion. The lack of distinct clustering patterns in this feature indicates that radial alignment is largely independent of other void properties such as size and shape, consistent with theoretical predictions that void orientations should be randomly distributed in an isotropic universe.

Finally in the bottom right plot an understanding of the temporal evolution of void properties throughout the surveyed volume can be gained from the redshift-coded visualization. Within our survey range of $z \leq 0.114$, the relatively uniform distri-

bution of various redshift values across the latent space implies that fundamental void characteristics do not change over cosmic time. Subtle patterns do, however, appear, with some regions displaying mild preferences for particular redshift ranges, which may point to mild evolutionary trends in void properties. While acknowledging that more comprehensive surveys covering larger redshift ranges may reveal stronger evolutionary signatures, the lack of strong redshift-dependent clustering validates our strategy of treating the catalog as a single sample.

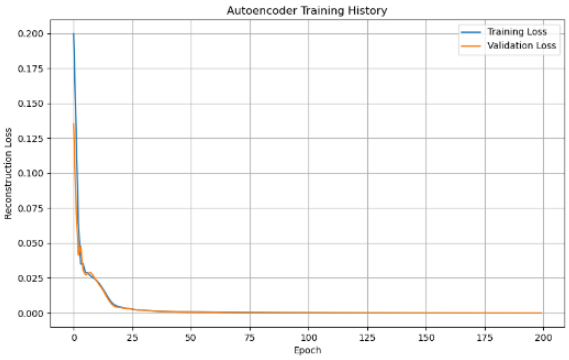

Figure 5: Auto encoder training

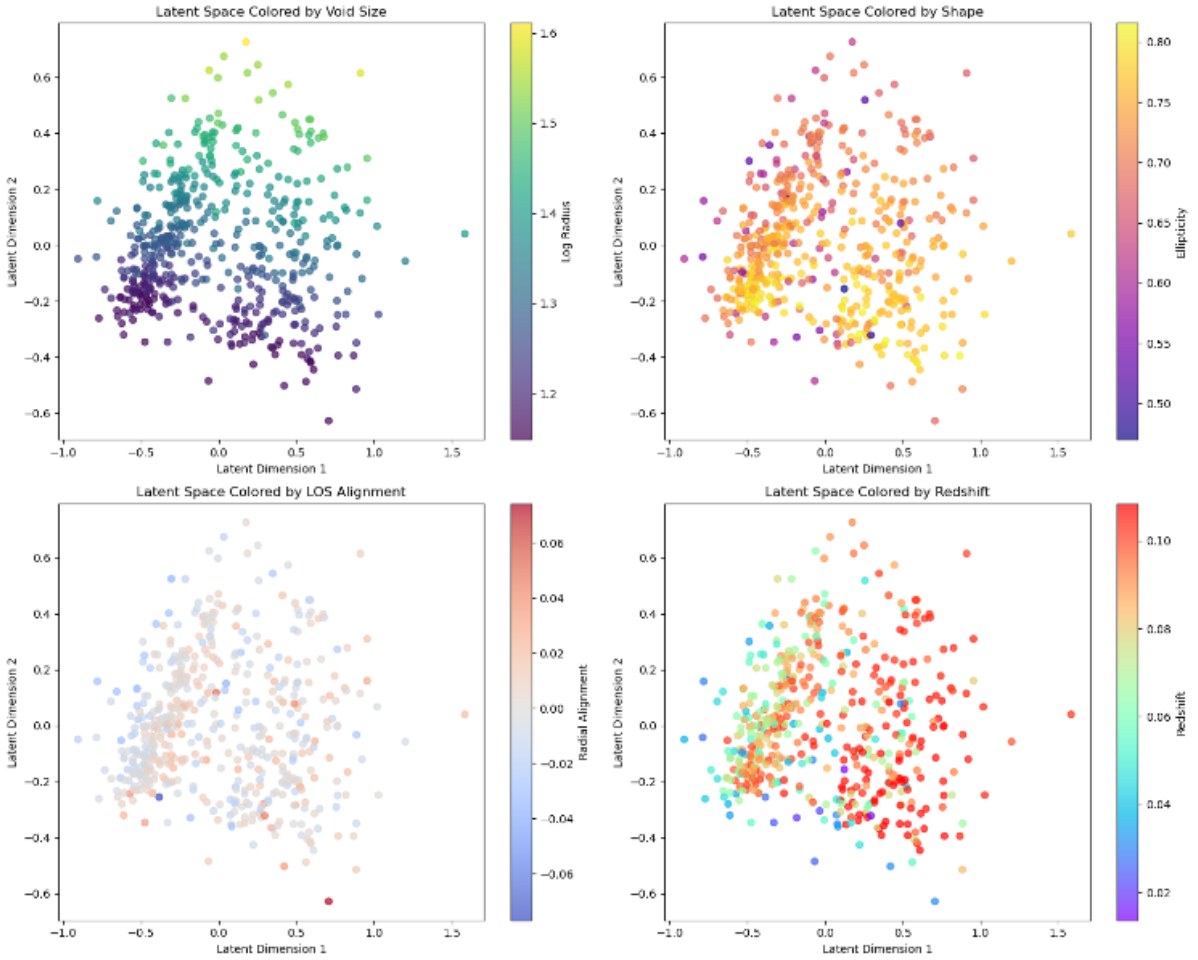

Figure 6: Latent space visualization of features: The 2D projection visualizes the latent space learned by our autoencoder with continuous colour gradients representing void characteristics across the manifold. Panel (a) shows void size (log R), where upper darker hues means larger voids. Panel (b) displays ellipticity with upper darker regions belonging to highly aspherical voids. Panel(c) maps radial alignment. Panel (d) tracks redshift evolutions.

To sum up, Our map of the latent space organization shows a clear difference between large and small voids. Large voids appear scattered across many different areas of the map. This spread suggests their growth is messy and complex, likely influenced by the varied environments they expand into. In contrast, small voids cluster together tightly in specific regions. This tight grouping hints

that small voids hold onto simpler, more "pristine" characteristics, closer to how they first formed.

We see the latent space supporting the idea that voids grow step-by-step (hierarchical evolution). The fact that large and small voids clearly occupy different parts of the map tells us that different physical rules probably govern how voids develop depending on their size.

Overall, this mapping technique or a visualization approach serves as a powerful diagnostic tool for understanding the physics underlying cosmic void formation and evolution. Seeing these clear patterns in the map validates using these advanced deep learning methods to study the structure of our universe.

4 Conclusions

We have presented that the combination of machine learning and statistics can shed light on the secrets of cosmic voids. In the ΛCDM framework, voids form from depressed densities in the primordial Gaussian field of density fluctuations and evolve hierarchially through bottom-up merging of smaller underdensities. They are shaped by two competing process: on one hand, internal expansion due to underdensity results in an effective peculiar gravitational influence, causing them to expand faster than the Hubble flow and driving toward sphericity; on the other hand, external tidal forces introduces anisotropy, as seen in our trend of decreasing ellipticity but increasing anisotropy with void size. Our observation that larger voids trend toward sphericity but yet never achieve perfect roundness–is because of the cosmic dance between voids and their environment. As voids expand, they inevitably encounter the dense filaments and walls that scaffold the cosmic web, halting their journey toward ideal spheres. This boundary confinement, noted by Platen and colleagues, explains why our largest voids plateau despite their growth.

More profoundly, tidal forces which are the gravitational pull from surrounding overdensities–continuously sculpt void shapes. These external influences induce anisotropies that can collapse void axes, creating the elongated forms we observe in intermediate-sized voids. The tension kind of relationship between internal expansion and external distortion creates a morphological spectrum and there each void tells a story of its cosmic neighborhood. Our latent space captures this narrative–not as discrete categories, but as a continuum of shapes reflecting the eternal tug-of-war between emptiness and matter.

With our incorporation of voids statistics with machine learning, we showed that the use of machine learning can be very informative in voids studies.

Acknowledgments

We thank the SDSS collaboration for making the data publicly available and acknowledge the use of the REVOLVER void catalog from Douglass et al. (2023). We acknowledge Prof. Walter Saurer of Innsbruck University, Austria for his suggestion on data reduction. We would also like to thank the anonymous referee for their feedback on this manuscript. We would also like to express our gratitude towards our family and friends for their encouragement and support till date.

References

[1] M. Argudo-Fernández et al. Morphologies of galaxies within voids. *Astron. Astrophys.*, 692:A258, 2024. doi:10.1051/0004-6361/202450809.

[2] J. R. Bond et al. How filaments of galaxies are woven into the cosmic web. *Nature*, 380:603–606, 1996. doi:10.1038/380603a0.

[3] Q. Mao et al. A cosmic void catalog of SDSS DR12 BOSS galaxies. *Astrophys. J.*, 835:161, 2017. doi:10.3847/1538-4357/835/2/161.

[4] Y. C. Cai et al. The integrated Sachs-Wolfe effect in f(r) gravity. *Mon. Not. R. Astron. Soc.*, 439:2978–2989, 2014. doi:10.1093/mnras/stu154.

[5] C. D. Kreisch et al. The GIGANTES data set: Precision cosmology from voids in the machine-learning era. *Astrophys. J.*, 935:100, 2022. doi:10.3847/1538-4357/ac7d4b.

[6] B. Y. Wang et al. Machine-learning cosmology from void properties. *Astrophys. J.*, 955:131, 2023. doi:10.3847/1538-4357/aceaf6.

[7] K. A. Douglass et al. Updated void catalogs of the SDSS DR7 main sample. *Astrophys. J. Suppl. Ser.*, 265:7, 2023. doi:10.3847/1538-4365/acabcf.

[8] Sloan Digital Sky Survey. SDSS sky survey, 2024. Accessed: 2024-04-07. URL: https://www.sdss3.org/science/gallery_sdss_pie2.php.

[9] A. Kravtsov and A. Klypin. Formation of the cosmic web, 2024. Accessed: 2024-04-07. URL: http://cosmicweb.uchicago.edu/filaments.html.

[10] E. Platen et al. Alignment of voids in the cosmic web. *Mon. Not. R. Astron. Soc.*, 387:128–136, 2008. doi:10.1111/j.1365-2966.2008.13019.x.

[11] D. Park and J. Lee. Void ellipticity distribution as a probe of cosmology. *Phys. Rev. Lett.*, 98:081301, 2007. doi:10.1103/PhysRevLett.98.081301.

[12] S. Nadathur. Testing cosmology with a catalogue of voids in the BOSS galaxy surveys. *Mon. Not. R. Astron. Soc.*, 461:stw1340, 2016. doi:10.1093/mnras/stw1340.

[13] E. Adermann et al. Cosmic voids in evolving dark sector cosmologies: the low redshift universe. *Mon. Not. R. Astron. Soc.*, 468:3381–3394, 2017. doi:10.1093/mnras/stx657.

[14] D. Bank et al. Autoencoders. In L. Rokach et al., editors, *Machine Learning for Data Science Handbook*, pages 353–374. 2023. doi:10.1007/978-3-031-24628-9_16.

[15] F. L. Gewers et al. Principal component analysis: A natural approach to data exploration. *ACM Comput. Surv.*, 54:70:1–70:34, 2021. doi:10.1145/3447755.

[16] I. T. Jolliffe and J. Cadima. Principal component analysis: a review and recent developments. *Philos. Trans. R. Soc. A*, 374:20150202, 2016. doi:10.1098/rsta.2015.0202.