NJIT Grace Hopper AI Research Institute Seed Program
Tier 1 Spark Grant

Training AI Models to Predict Permeability of Porous Media

Ebru Dagdelen, Aakash Karlekar, Manav Arora, Zhaoshu Cao, Catherin Lalu, Matthew Illingworth, Jonathan Jaquette, Linda Cummings, Lou Kondic

Department of Mathematical Sciences, New Jersey Institute of Technology

NSF Grant
DMS-2201627
DMR-2410985

Introduction

Permeability determines how easily fluids move through porous materials, controlling flow in natural and engineered systems such as groundwater filtration, enhanced oil recovery, and CO_2 sequestration. Traditional approaches to calculate permeability, based on direct experiments or numerical flow simulations, are accurate but computationally expensive.

In this work, we explore whether machine learning informed by topology and network descriptors can predict permeability efficiently while maintaining interpretability. Our approach combines geometric analysis, pore-network modeling, and topological data analysis (TDA) to build predictive models that are both data-driven and physically meaningful.

Data Generation

Synthetic 3D porous structures were generated using PuMA [1] with randomly distributed overlapping spheres (Figure 1). The dataset includes 1000 samples with controlled porosity (0.5) and variable particle diameters (5-15 voxels). Each voxel is assigned an integer value, where 0 denotes the void phase and 1 denotes the solid phase, producing a binary image of the structure. In figure 1, solid regions appear gray, while the void phase appears in white.

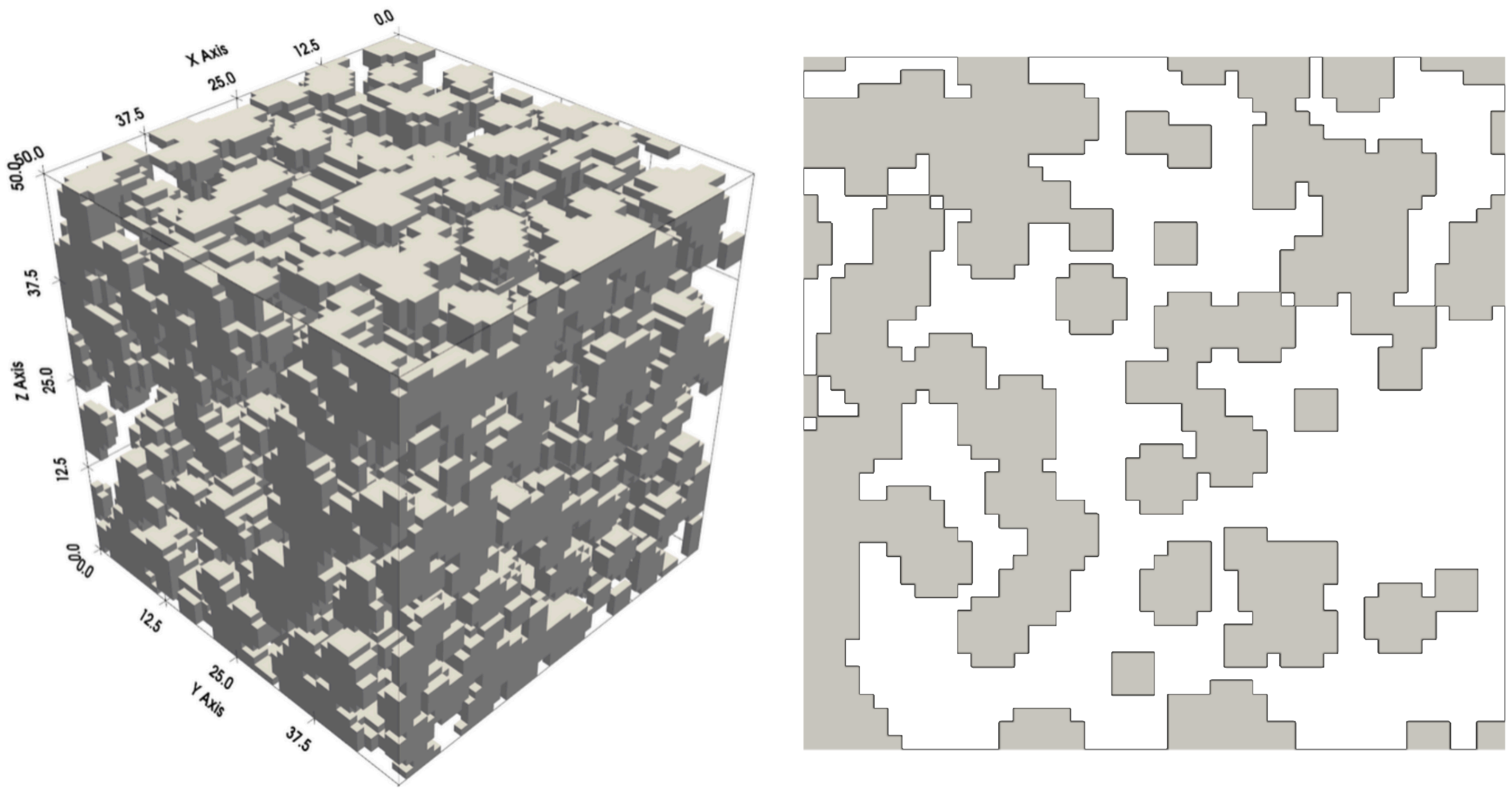

Figure 1: Simulated 3D dataset, 50 x 50 x 50 voxels; 2D cross-section of the same dataset

Methods

Topological Feature Extraction:

Persistence Homology (PH) quantifies connected components, loops, and voids across as a threshold parameter (α) varies. We computed measures of PH using two approaches:

→*Alpha Complexes - GUDHI [8]:* A geometric filtration constructed using Delaunay triangulation, where points are connected when their circumscribed spheres radius is less than or equal to α. As α increases, features such as connected components and cavities appear and merge, capturing the topology of the structure.

→ *Distance Transform - HomCloud [7]:* Computes PH directly from binary image using distance transform of the void space where each voxel's value equals its distance to the nearest solid. Increasing the threshold fills the pores and connects void regions, revealing the birth and death of topological features like connected components and cavities.

Pore Network Extraction:

The SNOW2 algorithm from Porespy [4] converts the binary void space into a pore-network graph by identifying pores (nodes) and throats (edges). Figure 2 illustrates an example of the SNOW2 method. Figure 2a illustrates the segmentation of the void space, and Figure 2b shows the resulting pore network overlay. This representation preserves connectivity information while greatly reducing computational complexity. From these networks, we computed geometric and graph descriptors such as degree distribution, throat length, and

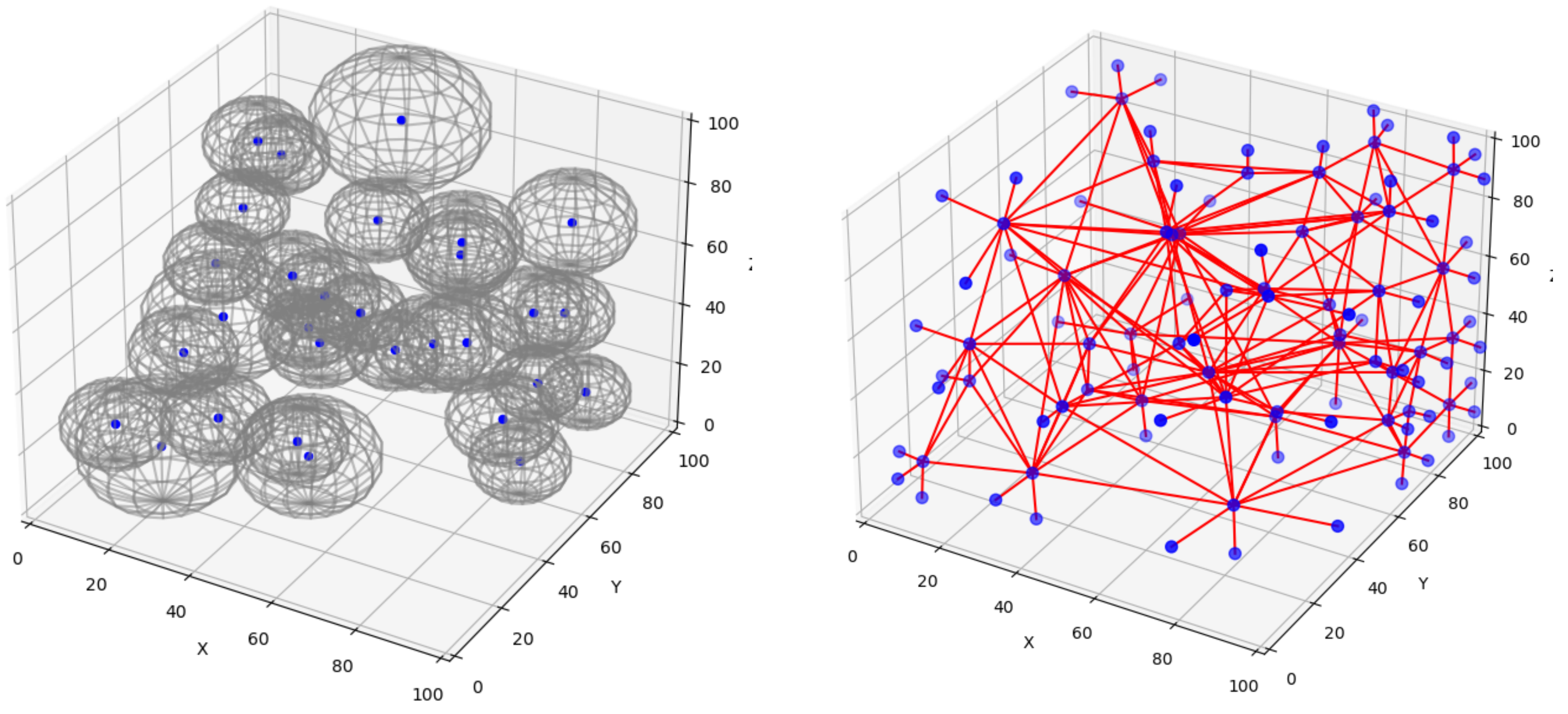

Figure 2: Pore-network extraction example using SNOW2. Pore centers are highlighted in blue.

Permeability Computation:

- For the 3D voxel datasets, permeability was computed using PuMA via direct flow simulations on the full geometry based on Darcy's Law.
- For the pore-network representations, permeability was obtained using OpenPNM which solves Stokes flow on the extracted networks.

Machine Learning & Neural Network

A Neural Network model was trained to predict permeability using structural (diameter, diffusivity, surface area, tortuosity), network (triads, path length, edge length, pore distances, centrality), and topological features (total persistence from Alpha Complex and Distance Transform filtrations). The model was based on Röding et al. [2] with only minor changes. The network has one input layer, three dense layers of 128 nodes each, and one output layer that predicts permeability. In total, this creates roughly 50,000 trainable parameters, giving the model enough flexibility to learn patterns in the data.

The dataset of 1000 samples was divided into 700 training, 150 validation, and 150 testing samples. The model uses a rectified linear unit (ReLU) activation function, mean squared error (MSE) as the loss function, and the Adam optimizer. Training was performed under 1000 epochs with a batch size of 32, which helped prevent overfitting while keeping computation reasonable.

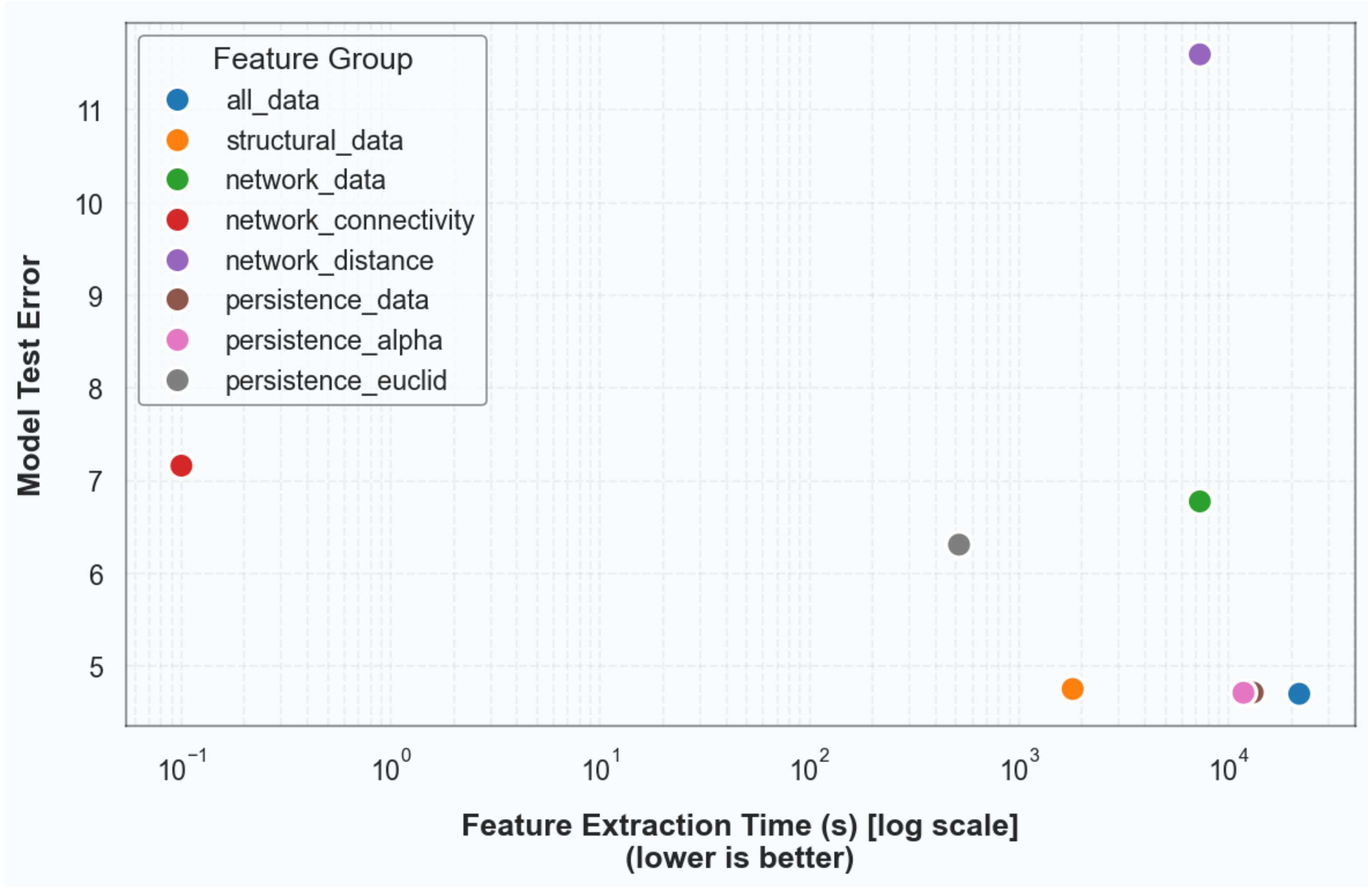

Figure 3: Model Test Error (%) vs. Feature Extraction Time (s) [log scale] Graph. Computations were carried out on a machine equipped with an 11th Gen Intel(R) Core(TM) i5-1145G7 CPU @ 2.60 GHz and 16 GB of RAM

Results & Conclusions

The neural network achieved the highest accuracy (MAPE ≈ 4.7%) when trained on all features combined, showing that using a combination of geometric, network, and topological information provides the most complete description of the porous structure. When considered separately, persistence features calculated through Alpha Complexes and the structural descriptors produced nearly identical results (MAPE ≈ 4.71% and 4.76%, respectively). In contrast, network-based features derived from SNOW2 resulted in higher errors (~6–12%) but were faster to compute.

Overall, the differences between structural and topological features were relatively small, suggesting that both entail permeability-relevant information effectively, while network metrics alone are not sufficient to obtain permeability.

Future Work

We plan to extend our method to more complex datasets, such as the image of the coquina sample shown in Figure 7, provided by experimental collaborators [4]. These structures present greater computational challenges due to their size and complexity. By applying our method to these datasets, we aim to investigate whether machine learning methods are able to robustly predict permeability.

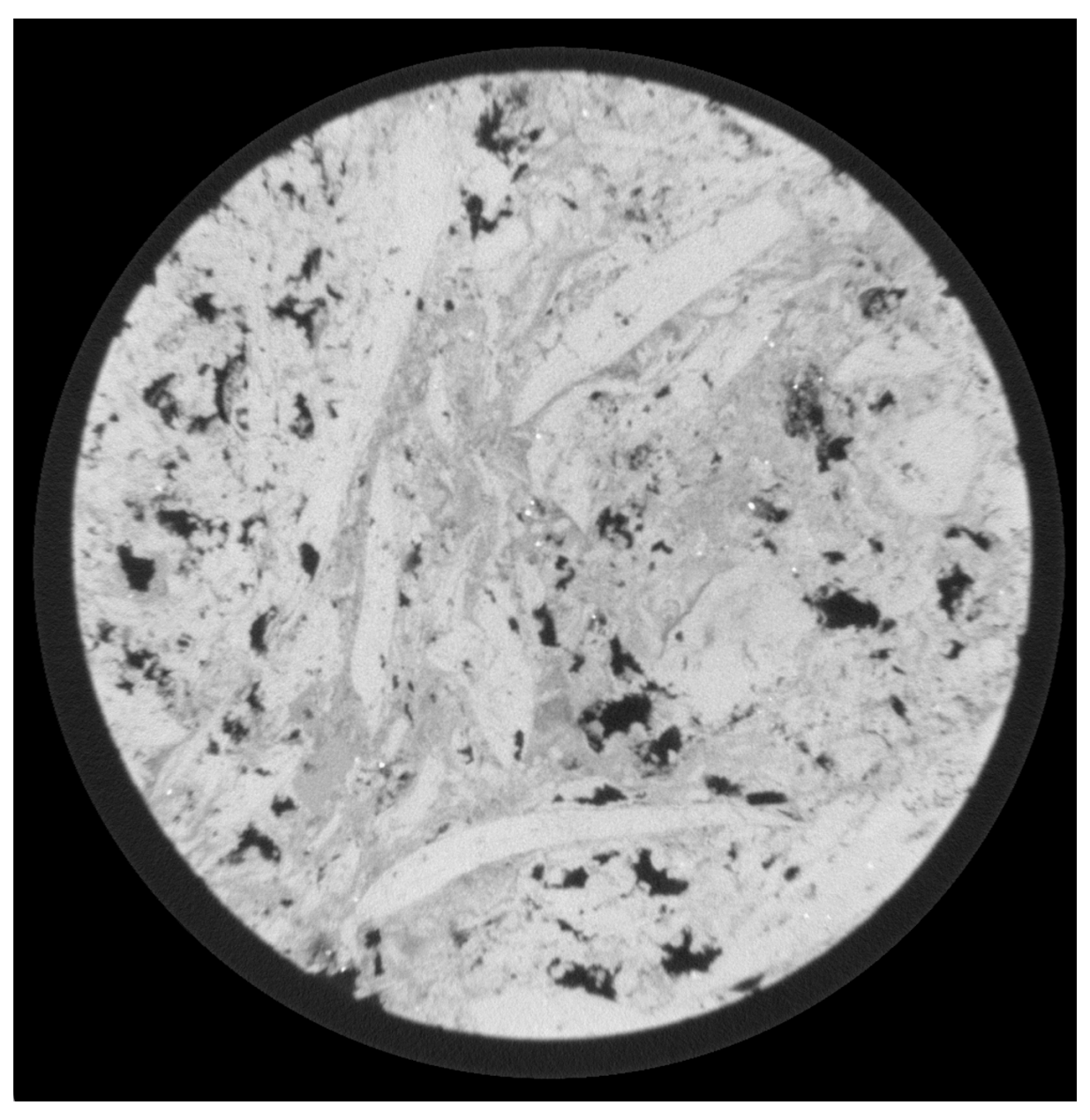

Figure 4: Cross-section of man-made coquina sample, courtesy of Prof. M. Carvalho, PUC-Rio, Brazil. Resolution: 40 microns; dimensions: 1001 x 1024 pixels.

References

[1] Ferguson, J. C., etal. (2021). Update 3.0 to PuMA: The Porous Microstructure Analysis software. SoftwareX, 15, 100775.

[2] Röding, M., etal. (2020). Predicting permeability via statistical learning on higher-order microstructural information. Scientific Reports, 10, 15239.

[3] Vianna, R. S., etal. (2020). Computing effective permeability of porous media with FEM and Micro-CT: An educational approach. Fluids, 5(1), 16.

[4] Gostick, J., etal. (2019). PoreSpy: A Python toolkit for quantitative analysis of porous media images. Journal of Open Source Software, 4(37), 1296.

[5] Gostick, J., etal. (2016). OpenPNM: A pore network modeling package. Computing in Science & Engineering, 18(4), 60–74.

[6] Zhang, J., etal. (2024). Knowledge extraction via machine learning guides a topology-based permeability prediction model. Water Resources Research, 60(7), e2024WR037124.

[7] Suzuki, A., etal. (2021). Flow estimation solely from image data through persistent homology analysis. Scientific Reports, 11, 17948.

[8] The GUDHI Project. (2025). GUDHI User and Reference Manual (Edition 3.11.0). GUDHI Editorial Board.