Proceedings of Machine Learning Research 259:824–838, 2024 Machine Learning for Health (ML4H) 2024

Topological Machine Learning for Low Data Medical Imaging

Brighton Nuwagira BRIGHTON.NUWAGIRA@UTDALLAS.EDU
University of Texas at Dallas, USA

Korkmaz Caner C.KORKMAZ23@IMPERIAL.AC.UK
Imperial College, London, UK

Koung Philmore Fan-Hsi PHILMORE.KOUNG@UTDALLAS.EDU
University of Texas at Dallas, USA

Baris Coskunuzer COSKUNUZ@UTDALLAS.EDU
University of Texas at Dallas, USA

Abstract

Deep Learning (DL) has revolutionized medical image analysis by providing automated techniques to extract valuable insights from large datasets. However, challenges such as interpretability and reliance on extensive labeled data persist. Topological Data Analysis (TDA) has emerged as a complementary approach that captures underlying topological structures in data, potentially enhancing the performance of DL models.

In this paper, we present a comprehensive evaluation of TDA methods for computer-aided diagnosis from two perspectives. First, we examine the effectiveness of topological methods in data-limited settings by comparing the standalone performance of DL models, TDA approaches, and their fusion. Our results demonstrate that integrating topological features into DL models significantly improves performance when labeled data are scarce. Second, we assess the standalone performance of TDA methods in data-rich environments using the MedMNIST collection, which includes over 600K images across 12 2D and 6 3D medical imaging datasets. Our experiments reveal that while TDA methods do not outperform DL models on 2D datasets, they achieve competitive results on 3D imaging tasks. These findings suggest that the fusion of TDA and DL methods can enhance the accuracy and robustness of computer-aided diagnosis, particularly in low-data or 3D imaging scenarios.

Keywords: Medical Image Analysis, Topological Data Analysis, Cubical Persistence, CNNs, Computer-aided Diagnosis, MedMNIST

Data and Code Availability Our datasets are publicly available at https://medmnist.com. We provide our code at https://github.com/Kausta/topo-net

Institutional Review Board (IRB) Our study does not require IRB approval.

1. Introduction

Deep Learning (DL) has become a cornerstone in medical image analysis, providing efficient, effective, and automated techniques to extract valuable insights from vast amounts of medical image data (Chan et al., 2020). These models excel at learning from large datasets, identifying patterns in medical images that are often imperceptible to the human eye, and improving the accuracy of disease detection and diagnosis. The rapid progress in machine learning has significantly enhanced the identification, classification, and quantification of patterns in medical imaging, achieving high precision in computer-aided diagnosis (Fujita, 2020). Despite these advances, challenges such as interpretability and the reliance on vast amounts of labeled data remain persistent. Therefore, complementary approaches that enhance computational feasibility, interpretability, and robustness are critical (Sarvamangala et al., 2022).

In the past decade, topological data analysis (TDA) has emerged as a promising approach that complements existing ML models in medical image analysis (Skaf et al., 2022). TDA focuses on extracting and analyzing the underlying topological structures in data, enabling a deeper understanding of complex relationships and patterns. By capturing topological features, TDA provides an additional layer of information, enriching the overall analysis (Choe and Ramanna, 2022; Hensel et al., 2021). Inspired by its successful application in various domains, TDA has shown potential as a powerful feature extraction tool that can enhance the performance of state-of-the-art (SOTA) models in medical image analysis (Singh et al., 2023).

In this paper, we present a comprehensive analysis of topological machine learning methods for computer-aided diagnosis from two distinct perspectives. First, we examine the effectiveness of topological methods in data-limited settings by comparing the standalone performance of several deep learning models, topological approaches, and their fusion. Our results show that in low-data scenarios, incorporating topological features significantly enhances the performance of deep learning models. Second, we assess the standalone performance of topological methods in data-rich environments across various medical imaging domains. To conduct this evaluation, we utilize the MedMNIST collection, which consists of 12 sets of 2D and 6 sets of 3D medical imaging benchmark datasets, comprising over 600,000 images. Our experiments reveal that while topological methods do not outperform DL models on 2D datasets, they achieve competitive results on 3D imaging tasks.

Our contributions:

- We provide a comprehensive evaluation of topological methods in medical image analysis, comparing its performance with deep learning models in both data-limited and data-rich settings.
- Our experiments show that integrating topological features into DL models significantly improves performance in data-limited environments, highlighting the value of TDA in scenarios with scarce labeled data.
- We assess TDA's performance in data-rich environments using the MedMNIST collection, demonstrating competitive results on 3D imaging tasks.
- We offer insights into the fusion of TDA and DL methods, enhancing computer-aided diagnosis with improved accuracy and robustness.

2. Related Work

CNNs in Computer-Aided Diagnosis. Medical image analysis has witnessed a transformative revolution with the advent of DL techniques. By harnessing various computer vision algorithms, DL has emerged as a key tool in extracting valuable insights from medical images. From accurate segmentation and classification to detecting subtle abnormalities, DL algorithms are empowering clinicians and researchers to unlock a wealth of information and make more informed decisions. In recent years, the dominance of traditional hand-crafted feature approaches has been overshadowed by the emergence of CNNs (Zhou et al., 2021; Halalli and Makandar, 2018). Numerous successful deep convolutional architectures have emerged for this purpose, including AlexNet (Krizhevsky et al., 2017), as well as the ResNet (He et al., 2016), Inception (Szegedy et al., 2017), MobileNet (Howard et al., 2017), EfficientNet (Tan and Le, 2019) and many more. Following the success of the attention mechanism and Transformer networks (Vaswani et al., 2017) in other domains, Vision Transformers (Dosovitskiy et al., 2021) achieved significant success in the domain of image analysis, leading to models like Swin Transformer (Liu et al., 2021, 2022). Despite their success, persistent challenges include the interpretability of results and the ongoing requirement for substantial volumes of labeled data, as highlighted in the study by Salehi et al. (2023). For a thorough review of deep learning methods in medical imaging, and computer-aided diagnosis, see excellent reviews by Aggarwal et al. (2021); Shamshad et al. (2023); Abdou (2022).

Recent advances in medical image analysis in limited data settings have focused on mitigating the challenges of overfitting and enhancing generalization by leveraging techniques such as data augmentation (Chlap et al., 2021), transfer learning (Kora et al., 2022), and synthetic data generation using generative adversarial networks (GANs) (Chen et al., 2022). Few-shot learning has gained significant traction in this space, enabling models to learn from a small number of labeled examples by utilizing meta-learning techniques or adapting pre-trained models to new tasks with minimal data (Singh et al., 2021; Feng et al., 2021). Self-supervised learning and contrastive learning approaches have also shown promise, leveraging large amounts of unlabeled data to extract meaningful representations, which are then fine-tuned for specific medical imaging tasks (Shurrab et al., 2022). These approaches, combined with careful model regularization, have proven effective in improving the robustness and accuracy of models in limited-data scenarios (Jiao et al., 2023).

TDA in Medical Image Analysis. Topological Data Analysis (TDA) is a novel approach for studying complex data by identifying both local and global structures at various scales, addressing challenges related to data dimensionality, differences in data collection methodologies, and varying scales (Coskunuzer and Akcora, 2024; Hensel et al., 2021). In the past decade, TDA has found widespread applications across multiple domains, including image analysis, neurology, cardiology, hepatology, gene-level and single-cell transcriptomics, drug discovery, evolution, and protein structural analysis (Skaf et al., 2022). By leveraging inherent topological features, and with its ability to capture hidden patterns of images, TDA opens up new opportunities for tasks such as image segmentation, object recognition, image registration, and image reconstruction (Choe and Ramanna, 2022). One commonly used tool from TDA in image analysis is persistent homology (PH) has shown its effectiveness in several medical imaging domains, e.g., histopathological images (Yadav et al., 2023), MRI images (François et al., 2024), and CT scans (Somasundaram et al., 2021). Recently, TDA methods have been increasingly utilized to improve the performance of deep learning models across various medical imaging domains, as demonstrated in studies such as those by Peng et al. (2024); McGuire et al. (2023). A comprehensive review of TDA methods in medical imaging and biomedicine can be found in the excellent survey (Singh et al., 2023).

3. Methodology

Our methodology leverages topological techniques, specifically cubical persistence, to extract image embeddings. We then develop two ML models: *Topo-Med*, which evaluates the standalone performance of topological features, and *Topo-Net*, which combines topological vectors with deep learning methods.

3.1. Cubical Persistence

Persistent homology (PH) is a powerful mathematical tool used in TDA to study the shape and structure of complex datasets. The main idea behind PH is to systematically assess the evolution of various hidden patterns in the data as we vary a scale parameter (Coskunuzer and Akcora, 2024). While PH is used as a very effective feature extraction method for various data types (e.g., point clouds, networks), in this paper, we focus only on the image setting, in particular *cubical persistence*. While we describe PH construction in simple terms for non-experts, one can find more details in the textbook by Dey and Wang (2022). Basically, PH can be described as a 3-step process: Inducing a nested sequence of topological spaces from data *Filtration*, recording the evolution of topological changes in this sequence *Persistence Diagram*, and finally, converting persistence diagrams into vectors *Vectorization*.

Step 1 - Constructing Filtrations. We first describe the process for 2D images. Then, we explain how to adapt this construction to 3D images. For a given 2D image $\mathcal{X}$ (say $r \times s$ resolution), the filtration step is to create a nested sequence of binary images (aka cubical complexes) (Kaji et al., 2020). To create such a sequence, one can use grayscale (or other color channels) values $\gamma_{ij} \in [0, 255]$ of each pixel $\Delta_{ij} \subset \mathcal{X}$. In particular, for a sequence of grayscale values, called *thresholds*, $(0 \leq t_1 < t_2 < \cdots < t_N \leq 255)$, one obtains a nested sequence of binary images $\mathcal{X}_1 \subset \mathcal{X}_2 \subset \cdots \subset \mathcal{X}_N$ such that $\mathcal{X}_n = \{\Delta_{ij} \subset \mathcal{X} \mid \gamma_{ij} \leq t_n\}$ (See Figure 1). In other words, we start with an empty $r \times s$ image and start activating (coloring black) pixels when their grayscale value reaches the given threshold. This is called *sublevel filtration* for $\mathcal{X}$ with respect to a given function (grayscale color values in this case). Similarly, when $\mathcal{X}$ is a 3D image (say $r \times s \times t$ resolution), we consider $\mathcal{X}$ as a 3D cubical complex of size $r \times s \times t$. Again, to create our nested sequence of 3D binary images, we use grayscale values $\gamma_{ijk} \in [0, 255]$ of each voxel $\Delta_{ijk} \subset \mathcal{X}$. Then, for a sequence of grayscale values $(0 \leq t_1 < t_2 < \cdots < t_N \leq 255)$, one obtains 3D binary images $\mathcal{X}_1 \subset \mathcal{X}_2 \subset \cdots \subset \mathcal{X}_N$ where $\mathcal{X}_n = \{\Delta_{ijk} \subset \mathcal{X} \mid \gamma_{ijk} \leq t_n\}$ as before.

Step 2 - Persistence Diagrams. PH keeps track of the evolution of topological features in the filtration sequence, and records as *persistence diagram*. In particular, if a topological feature σ first appears in $\mathcal{X}_m$ and disappears in $\mathcal{X}_n$ with $1 \leq m < n \leq N$, we call the corresponding thresholds t_m as *birth time* b_σ and t_n *the death time* d_σ of the topological feature σ,

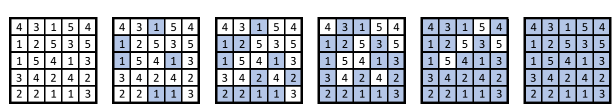

Figure 1: For the 5×5 image $\mathcal{X}$ with the given pixel values, **the sublevel filtration** is the sequence of binary images $\mathcal{X}_1 \subset \mathcal{X}_2 \subset \mathcal{X}_3 \subset \mathcal{X}_4 \subset \mathcal{X}_5$.

i.e. $b_\sigma = t_m$ and $d_\sigma = t_n$. Then, PD is the collection of all such 2-tuples $\mathrm{PD}_k(\mathcal{X}) = \{(b_\sigma, d_\sigma)\}$ where k represent the dimension of the topological features. The difference $d_\sigma - b_\sigma$ is called *lifespan* of σ. More formally, k^{th} persistence diagram can be defined as $\mathrm{PD}_k(\mathcal{X}) = \{(b_\sigma, d_\sigma) \mid \sigma \in H_k(\mathcal{X}_i) \text{ for} b_\sigma \leq i < d_\sigma\}$ where $H_k(\mathcal{X}_i)$ is the k^{th} homology group of the cubical complex $\mathcal{X}_i$. Hence, $\mathrm{PD}_k(\mathcal{X})$ is the collection of 2-tuples marking the birth and death times of k-dimensional holes $\{\sigma\}$ (connected components, loops, and voids) in the filtration sequence $\{\mathcal{X}_i\}$. e.g., for Figure 1, $\mathrm{PD}_0(\mathcal{X}) = \{(1,\infty),(1,2),(1,2),(1,3)\}$ marking the connected components, while $\mathrm{PD}_1(\mathcal{X}) = \{(2,4),(3,5),(4,5)\}$ marking the holes in respective binary images in Figure 1.

Step 3 - Vectorizations. Persistence Diagrams (PDs) which are collections of 2-tuples, are not particularly practical for use with ML tools. Instead, a widely adopted approach is to convert the PD information into a vector or a function, a process known as "vectorization" (Ali et al., 2023). A commonly used function for this purpose is the *Betti function* which keeps track of the number of "alive" topological features at a given threshold. In particular, the Betti function is a step function with $\beta_0(t_n)$ the count of connected components in the binary image $\mathcal{X}_n$, and $\beta_1(t_n)$ the number of holes (loops) in $\mathcal{X}_n$. In ML applications, Betti functions are usually taken as a vector $\overrightarrow{\beta_k}$ of size N with entries $\beta_k(t_n)$ for $1 \leq n \leq N$. i.e., $\overrightarrow{\beta_k}(\mathcal{X}) = [\beta_k(t_1) \ \dots \ \beta_k(t_N)]$. e.g., for Figure 1, we have $\overrightarrow{\beta_0}(\mathcal{X}) = [4\ 2\ 1\ 1\ 1]$ which are the count of connected components in binary image $\{\mathcal{X}_i\}$ while $\overrightarrow{\beta_1}(\mathcal{X}) = [0\ 1\ 2\ 2\ 0]$ are the counts of holes in $\{\mathcal{X}_i\}$. i.e., $\beta_0(1) = 4$ is the count of components in $\mathcal{X}_1$ and $\beta_1(3) = 2$ is the count of holes (loops) in $\mathcal{X}_3$.

There are various other ways to convert PDs into a vector, e.g., persistence images, persistence landscapes, and kernel methods (Ali et al., 2023).

Topological Vectors. For a given 2D (or 3D) image $\mathcal{X}$, say at $r \times s$ (or $r \times\ s \times t$) size, to produce its topological feature vectors, we first obtain a sublevel filtration, a sequence of binary images, $\mathcal{X}_1 \subset \mathcal{X}_2 \dots \subset \mathcal{X}_N$ by using the grayscale values of each pixel. If $\mathcal{X}$ is a color image, one can replace grayscale values with RGB color channels. Here, N is the number of thresholds equally dividing $[0, 255]$ color interval. In practice, $N = 50$ or 100 works well. The next step is to get the persistence diagram of this sublevel filtration. The final step is to choose a good vectorization. In our experiments, we use Betti functions and Silhouette functions with different powers. If the image is 2D, only meaningful dimensions are $k = 0$ and $k = 1$. If the image is 3D, then the only meaningful dimensions are $k = 0, 1, 2$. e.g., assuming $N = 100$ and Betti functions as vectorization, for each image, we obtain $\overrightarrow{\beta_0}$ and $\overrightarrow{\beta_1}$ vectors where each is 100 dimensional. After concatenation, this results in 200-dimensional embeddings for 2D images, and 300-dimensional embeddings for 3D images.

3.2. ML Models

Once we have obtained the topological feature vectors, the next step involves applying ML tools to analyze the effectiveness of these features. Essentially, each medical image is represented as a 200-dimensional vector (300-dimensional for 3D images). The classification task treats these embeddings as a point cloud and aims to identify distinct clusters corresponding to different classes. To evaluate the utility of these vector in computer-aided diagnosis, we employed two different ML models. In the first model, Topo-Net, we utilized topological vectors in existing deep learning models. We developed simpler and scalable second model, Topo-Med, to test the standalone performance of these topological vectors.

Topo-Net for Limited Data. In our first model, Topo-Net (Figure 2), we assess the impact of incorporating topological features into CNN architectures trained from scratch on datasets with limited samples. The motivation behind this approach lies in the complementary roles of topological features and convolutional layers: topological features capture the global structure of an image, providing embeddings

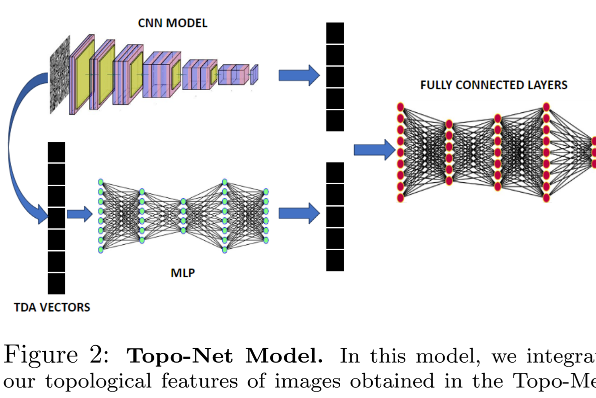

Figure 2: **Topo-Net Model.** In this model, we integrate our topological features of images obtained in the Topo-Med model with convolutional vectors generated by a CNN backbone, followed by processing through a fully connected layer.

that are less dependent on data quantity, while convolutional layers focus on localized patterns, where learning is highly influenced by the amount of available data. By merging these two perspectives, we hypothesize that the model will exhibit improved performance. Furthermore, we deliberately chose a simple architecture (Figure 2) to isolate and evaluate the contribution of topological vectors in enhancing DL models under limited-data settings.

For the 2D tasks, the Topo-Net model is crafted by selecting a base model from a range of state-of-the-art CNN models, including ResNet50 (He et al., 2016) as a baseline similar to the original MedMnist experiments (Yang et al., 2023a) and EfficientNet-B0 (Tan and Le, 2019) and Swin Transformer V2-Tiny (Liu et al., 2022) as more up-to-date alternatives. These were chosen for their robust feature extraction capabilities and performance in image classification tasks.

Topo-Net extends these models with an auxiliary feature network, taking Betti features as input. This Betti network uses a Multilayer Perceptron (MLP) to learn and extract features from the Betti numbers. Then, a classifier head consisting of an MLP is used to merge the CNN-based features and the auxiliary features from the Betti network and predict the output class. This allows the topological features to be incorporated with the convolutional features in a simple and network-agnostic way, allowing easy modification of other networks for limited-data scenarios where topological features can be obtained without additional training or data.

The 3D models are constructed similarly with an auxiliary Betti number MLP and a classifier MLP. These MLPs are used with a ResNet-18 (He et al., 2016) backbone where the 2D convolutions are converted into 2.5D, 3D, or ACS convolutions from the previous work ACSConv (Yang et al., 2021a), similar to the original MedMnist experiments. Similar to the 2D case, this allows the usage of topological features like Betti numbers without requiring significant network architecture changes, as the auxiliary networks can be added to any 3D CNN-based backbone.

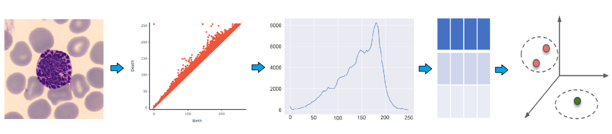

Figure 3: **Topo-Med model.** In our basic model, we simply extract topological feature vectors from images and apply standard ML models.

Topo-Med model. To evaluate the standalone performance of topological vectors across various medical domains, we introduce a simple ML model, Topo-Med (Figure 3). In this model, we use two types of classifiers with our topological vectors. The first is eXtreme Gradient Boosting (XGBoost), a tree-based method widely recognized for its strong performance. The second is the Multilayer Perceptron (MLP), a feedforward neural network known for its flexibility and robustness. These models offer diverse strengths for classification tasks, particularly in high-dimensional settings. MLP typically performs well with large and imbalanced datasets, while XGBoost consistently delivers reliable results across different scenarios. Further details on our ML procedures and hyperparameter tuning are provided in Section 4.

4. Experiments

4.1. Experimental Setup

Datasets. In our experiments, we utilized the publicly available MedMNIST dataset collection, which is widely recognized in the medical imaging community (https://medmnist.com). MedMNIST comprises twelve pre-processed 2D datasets and six pre-processed 3D datasets sourced from various modalities, representing diverse classification tasks and varying data scales. The dataset includes predefined training and test splits and offers multiple resolutions, ranging from 28x28 to 224x224 for 2D images, and from 28x28x28 to 64x64x64 for 3D images (Yang et al., 2023b). Detailed information about these datasets is provided in Table 9 in Appendix.

Topological Vectors. For both our models, Topo-Net and Topo-Med, we first generated topological feature vectors. For each dataset, we applied sublevel filtration using 100 evenly spaced thresholds within the range [0, 255] for the grayscale channel. Betti vectorization was then applied to the resulting persistence diagrams. This process yielded 100-dimensional Betti vectors, $\vec{\beta_i}(\mathcal{X})$, for each dimension of an image $\mathcal{X}$. After concatenation, we obtained a 200-dimensional vector for each MedMNIST2D dataset and a 300-dimensional vector for each MedMNIST3D dataset.

Topo-Net Hyperparameters. We followed the hyperparameters from the original MedMNIST experiments as closely as possible (Yang et al., 2023b). We used a batch size of 128 for all 2D models with 224x224 images and a batch size of 16 for all 3D mod-

Table 1: **2D Limited Data Performances.** AUC results for vanilla-CNN and Topo-CNN models using ResNet-50, EfficientNetB0, and Swin Transformer backbones in limited data settings on the MedMNIST 2D datasets. The TDA columns display the standalone performance of topological vectors (Topo-Med). For each backbone, the top AUC **improvements** are highlighted in **blue**, and highest performance for each sample training size is **underlined**. Accuracy results are given in Table 5.

	DermaMNIST							TissueMNIST							PneuMNIST						
#	TDA	Res	T-Res	Eff	T-Eff	Swin	T-Swin	TDA	Res	T-Res	Eff	T-Eff	Swin	T-Swin	TDA	Res	T-Res	Eff	T-Eff	Swin	T-Swin
100	**0.8163**	0.7068	0.6783	0.6775	0.6790	0.6149	0.6576	**0.7273**	0.6169	0.6316	0.5973	0.6336	0.5670	0.6303	0.7466	0.8968	**0.9221**	0.9092	0.8862	0.8509	0.8215
250	**0.8217**	0.7942	0.7829	0.7687	0.7520	**0.6339**	**0.7432**	**0.7345**	0.6356	0.6800	0.6253	0.6583	0.5851	0.6686	0.7532	**0.8960**	**0.9296**	0.9157	0.9178	**0.6352**	**0.8690**
500	**0.8342**	**0.8181**	**0.8261**	0.8073	0.7894	0.6737	0.7169	**0.7421**	**0.6393**	**0.6947**	**0.6233**	**0.6731**	0.5980	0.6841	0.8061	0.9213	**0.9446**	0.9340	0.9210	0.8351	0.8678
1000	0.8451	**0.8625**	0.8446	0.8454	0.8511	0.6719	0.7467	**0.7322**	0.6967	0.7177	0.6911	0.7033	0.5788	0.6994	0.8029	0.9443	**0.9586**	0.9359	0.9252	0.7868	0.8944
2000	0.8665	**0.8784**	0.8634	**0.8502**	**0.8766**	0.7900	0.7898	0.7289	0.7260	**0.7440**	0.7087	0.7328	**0.5659**	**0.7118**	0.8525	0.9532	**0.9605**	**0.9178**	**0.9413**	0.7369	0.8912

	BloodMNIST							PathMNIST							OCTMNIST						
#	TDA	Res	T-Res	Eff	T-Eff	Swin	T-Swin	TDA	Res	T-Res	Eff	T-Eff	Swin	T-Swin	TDA	Res	T-Res	Eff	T-Eff	Swin	T-Swin
100	0.9129	**0.9403**	**0.9639**	**0.8474**	**0.9527**	0.8080	0.9549	0.7660	**0.8379**	**0.8745**	**0.8270**	**0.9042**	0.7179	0.8818	0.5379	**0.6956**	**0.6909**	0.5079	0.6220	0.5149	0.5821
250	0.9487	0.9822	**0.9902**	0.9249	0.9823	0.8262	0.9832	0.8379	**0.9380**	0.9310	0.8943	0.9147	**0.5664**	**0.9084**	0.5691	**0.8178**	0.8006	**0.6283**	**0.7717**	0.5943	0.6979
500	0.9629	0.9865	**0.9926**	0.9621	0.9865	0.8750	0.9863	0.8223	**0.9560**	0.9284	0.9144	0.9390	0.6658	0.9268	0.6043	**0.8734**	0.8414	0.7536	0.8273	0.5789	0.7118
1000	0.9778	**0.9951**	0.9946	0.9883	0.9915	0.8668	0.9907	0.8622	**0.9648**	0.9553	0.9449	0.8924	0.6697	0.9280	0.6315	**0.9330**	0.8946	0.8526	0.8531	**0.5394**	**0.7182**
2000	0.9778	**0.9978**	0.9964	0.9965	0.9957	**0.8198**	**0.9929**	0.9088	**0.9756**	0.9691	0.9648	0.9579	0.7918	0.9237	0.6381	**0.9636**	0.9442	0.8983	0.9400	0.5981	0.7648

Table 2: **3D Limited Data Performances.** AUC results for vanilla-CNN and Topo-CNN models using ResNet-18 + 2.5D, ResNet-18 + 3D, and ResNet-18 + ACS backbones in limited data settings on the MedMNIST 3D datasets.

	AdrenalMNIST3D							VesselMNIST3D						
#	TDA	R18-ACS	T-R18-ACS	R18-3D	T-R18-3D	R18-2.5D	T-R18-2.5D	TDA	R18-ACS	T-R18-ACS	R18-3D	T-R18-3D	R18-2.5D	T-R18-2.5D
50	**0.8426**	0.8358	0.7259	**0.7970**	**0.8349**	0.6734	0.7472	**0.7733**	0.7330	0.7236	0.7725	0.6890	**0.6562**	**0.6861**
100	**0.8529**	**0.7987**	**0.8472**	0.8396	0.8453	0.7352	0.7160	**0.8388**	0.6505	0.7290	0.7679	0.7546	0.7420	0.6629
200	0.8566	0.7999	0.8343	0.8540	**0.8577**	**0.6697**	**0.7607**	**0.8801**	**0.6678**	**0.7866**	**0.7628**	**0.7910**	0.7700	0.7860
250	0.8511	0.7863	0.7932	0.8570	0.8113	0.6962	0.7748	**0.9194**	0.8050	0.8175	0.8184	0.8065	0.7819	0.8029
500	0.8592	0.8958	0.8497	0.8636	0.8048	0.8079	0.8169	**0.9171**	0.8175	0.8458	0.9274	0.8542	0.8024	0.8584

els with 64x64x64 inputs. We used cross-entropy loss with Adam optimizer with an initial learning rate of 0.001, and we trained the models for 100 epochs, multiplying the learning rate by 0.1 after 50 epochs and 75 epochs. Key performance metrics are incorporated, including accuracy and AUC.

For the Topo-Net variants, an auxiliary three-layer Betti MLP with 256 hidden layer dimensions and a two-layer classifier MLP with 256 hidden layer dimension were used. In addition, a feature size of 128 for CNN and 128 for the auxiliary Betti MLP was chosen based on initial hyperparameter tests comparing different ratios of feature sizes for the CNN and the Betti MLP. For the CNNs, original normalization and activations were used, and ReLU activations were used for all MLPs. The data augmentations followed were the same as the MedMnist experiments (no augmentation for the 2D datasets, multiplying the training set by a random value in [0, 1] during training and multiplying the images by a fixed coefficient of 0.5 during evaluation for the 3D datasets).

Topo-Med Hyperparameters. We trained both XGBoost and MLP models on our datasets. The XGBoost model was set up for multi-class classification with *multi:softmax* and *binary:logistic* for binary classification as the objective and the number of classes specified by MedMNIST in each dataset. For the MLP model, the architecture included one input layer, three hidden layers with the first two having 256 units each and ReLU activation, and the third with 128 units and ReLU activation, and a dropout layer with a rate of 0.5. The output layer used a softmax activation function. The model was compiled with the categorical crossentropy loss function and the Adam optimizer with a learning rate of 0.001. We trained the MLP models for 100 epochs with a batch size of 32.

Computational Complexity & Runtime. PH calculation can be computationally intensive, especially in high-dimensional data (Otter et al., 2017). However, for 2D images, it is relatively efficient, with a time complexity of approximately $\mathcal{O}(|\mathcal{P}|^{2.3})$, where$|\mathcal{P}|$ denotes the total number of pixels (Milosavljević et al., 2011). This means the computational effort grows almost quadratically with image size. In contrast, subsequent steps like vectorization and ML are comparatively lightweight.

We conducted all experiments on a high-performance computing system configured with one node, 20 tasks, and the 256i partition. The 256i par-

Table 3: **Topo-Med vs. DL.** Performance comparison of seven deep learning models vs. standalone performance of topological vectors on MedMNIST collection using full datasets. Detailed results for the seven baseline models are provided in Table 8, while additional performance metrics for Topo-Med are presented in Table 7 in the Appendix.

	PathM		**ChestM**		**DermaM**		**OCTM**		**PneuM**		**RetinaM**	
Methods	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**
Best Baseline	0.989	0.909	0.778	0.948	0.920	0.754	0.963	0.771	0.991	0.946	0.750	0.531
Topo-Med	0.942	0.683	0.787	0.530	0.904	0.669	0.710	0.450	0.845	0.762	0.728	0.458
	BreastM		**BloodM**		**TissueM**		**OrganAM**		**OrganCM**		**OrganSM**	
Methods	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**
Best Baseline	0.919	0.861	0.998	0.966	0.941	0.703	0.997	0.935	0.994	0.920	0.974	0.813
Topo-Med	0.821	0.737	0.973	0.798	0.837	0.450	0.921	0.523	0.894	0.489	0.910	0.532
	OrganM3D		**NoduleM3D**		**FractureM3D**		**AdrenalM3D**		**VesselM3D**		**SynapseM3D**	
Methods	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**
Best Baseline	0.996	0.907	0.914	0.874	0.750	0.571	0.839	0.754	0.930	0.928	0.851	0.795
Topo-Med	0.837	0.554	0.808	0.736	0.653	0.480	0.864	0.769	0.920	0.887	0.730	0.730

tition is equipped with 20 Intel Ivy Bridge cores and 256 GB of memory. Generating both Betti-0 and Betti-1 vectors from the OCTMNIST dataset, which contains 109,309 images of size 28×28, took 50 minutes and 54 seconds. The remaining ML tasks were significantly faster, all completed in under a minute. For smaller datasets, the generation of topological vectors was much faster.

4.2. Results and Discussion

Limited Data Performance. In Tables 1 and 2, we present our results across various limited data settings. For the 2D datasets, we used 100, 250, 500, 1000, and 2000 training images with the backbones ResNet50, EfficientNetB0, and Swin Transformers. For the 3D datasets, we used 50, 100, 200, 250, and 500 training images with the backbones ResNet-18 + 2.5D, ResNet-18 + 3D, and ResNet-18 + ACS. Additionally, in Figure 4, we provide a comprehensive summary of AUC performance across selected MedMNIST datasets, illustrating the impact of TDA-enhanced models on CNN performance. This figure demonstrates how integrating topological data analysis (TDA) consistently improves model robustness and generalization, particularly in limited data scenarios. *All models were evaluated on the original test set of the corresponding MedMNIST dataset*, consisting of 2,005 images for DermaMNIST, 47,280 images for TissueMNIST, 624 images for PneuMNIST, 3,421 images for BloodMNIST, 7,180 images for PathMNIST, 1,000 images for OCTMNIST, 298 images for AdrenalMNIST3D, and 382 images for VesselMNIST3D.

Corresponding accuracy results are detailed in Tables 5 and 6, and additional performance metrics for the Topo-Med model are presented in Table 7 in the Appendix. For the 2D datasets, our results show that topological vectors consistently enhance deep learning models in limited data scenarios. In particular, the performance of Swin Transformers improves by as much as 30%. Similarly, the EfficientB0 model shows significant gains, with improvements of up to 15%. Although the synergy between ResNet50 and topological vectors is less substantial, we still observe improvements of up to 6%. When considering the best results for each limited data setting, topological vectors, either on their own or in combination with a model, achieve the highest performance in nearly all settings, with few exceptions in the PathMNIST and OCTMNIST datasets.

In 3D datasets, the performance of topological vectors is even more remarkable. In nearly all limited data scenarios, topological vectors outperform deep learning (DL) models when used alone. This difference is especially significant in the VesselMNIST3D dataset. Furthermore, the improvement that topological vectors provide to various backbone models is consistent, ranging from 3% to 12%.

These results support our hypothesis that in data-limited settings, topological vectors provide critical complementary information to deep learning models, showing great promise in addressing this challenge. Notably, even state-of-the-art models like Swin Transformers struggle with limited data, and our results suggest that integrating topological vectors has strong potential to mitigate this issue.

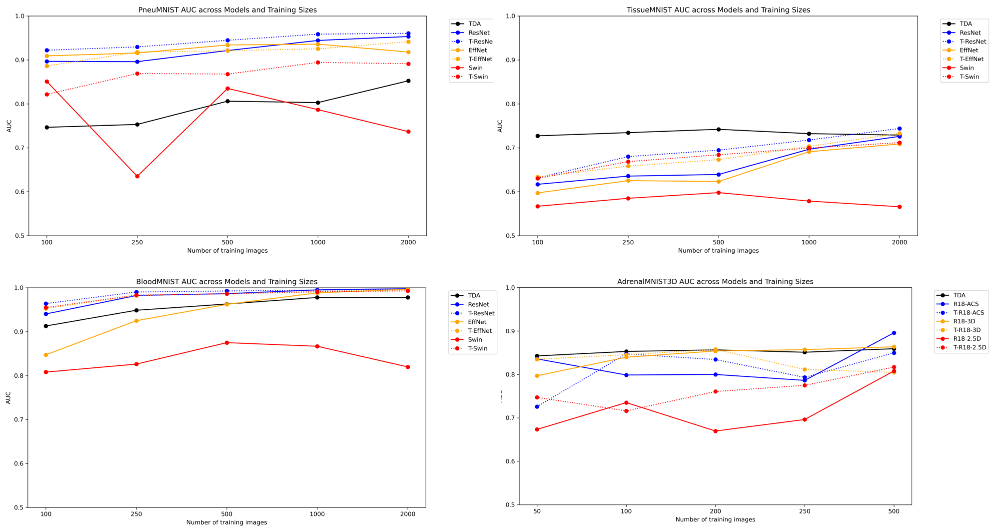

Figure 4: Summary of AUC performance across selected MedMNIST datasets.

Table 4: **Effect of Resolution & ML Classifiers on TDA performance.** Performance comparison of topological vectors on high (128×128) vs. low (28×28) resolution images with XGBoost and MLP Classifiers. Colored bottom rows give the difference between the best performances of low vs. high-resolution models. Blue numbers represent high res model is better, and red numbers represent low res model is better.

	PathM		ChestM		DermaM		OCTM		PneuM		RetinaM		BreastM		BloodM		TissueM	
Methods	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**
Low + XGB	0.943	0.712	0.567	0.638	0.684	0.665	0.820	0.505	0.838	0.732	0.618	0.453	0.795	0.752	0.972	0.810	0.813	0.485
Low + MLP	0.942	0.683	0.787	0.530	0.904	0.669	0.710	0.450	0.845	0.762	0.728	0.458	0.821	0.737	0.973	0.798	0.837	0.450
High + XGB	0.962	0.746	0.623	0.635	0.796	0.685	0.852	0.475	0.888	0.789	0.614	0.433	0.768	0.821	0.997	0.944	0.829	0.511
High + MLP	0.949	0.689	0.779	0.632	0.921	0.678	0.673	0.433	0.894	0.832	0.671	0.395	0.854	0.794	0.996	0.936	0.862	0.489
High vs. Low	**1.9%**	**3.4%**	**0.8%**	**0.3%**	**1.7%**	**1.6%**	**3.2%**	**3.0%**	**4.9 %**	**7.0 %**	**5.7%**	**2.5%**	**3.3%**	**6.9 %**	**2.4 %**	**13.4 %**	**2.5 %**	**2.6 %**

	OrganAM		OrganCM		OrganSM		OrganM3D		NoduleM3D		FractureM3D		AdrenalM3D		VesselM3D		SynapseM3D	
Methods	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**
Low + XGB	0.924	0.565	0.900	0.543	0.894	0.546	0.933	0.585	0.768	0.823	0.625	0.463	0.749	0.769	0.662	0.887	0.783	0.810
Low + MLP	0.921	0.523	0.894	0.489	0.910	0.532	0.837	0.554	0.808	0.736	0.653	0.48	0.864	0.769	0.920	0.887	0.730	0.730
High + XGB	0.949	0.630	0.918	0.585	0.912	0.587	0.941	0.616	0.762	0.800	0.582	0.454	0.743	0.758	0.694	0.887	0.786	0.790
High + MLP	0.950	0.604	0.926	0.580	0.924	0.585	0.848	0.590	0.866	0.819	0.602	0.408	0.867	0.769	0.923	0.887	0.743	0.730
High vs. Low	**2.6%**	**6.5%**	**2.6%**	**4.2%**	**1.4%**	**4.1%**	**0.8%**	**3.1%**	**5.8%**	**0.4%**	**5.1%**	**2.6%**	**0.3%**	0%	**0.3%**	0%	**0.3%**	**2%**

Topo-Med & Entire Dataset. In Tables 3 and 8, we present the standalone performance of topological vectors using the full dataset and compare it with SOTA benchmarks provided by MedMNIST (Yang et al., 2023b) for 28x28 (or 28x28x28) size images. For 2D datasets, the seven baselines include two pre-trained CNNs (for two image sizes): ResNet-18 and ResNet-50 (He et al., 2016), and three AutoML models: auto-sklearn (Feurer et al., 2022), AutoKeras (Jin et al., 2019), and Google AutoML Vision (Bisong, 2019). For 3D datasets, eight baselines include ResNet-18 and ResNet-50 with 2.5D/3D/ACS convolutions (Yang et al., 2021b), and the AutoML methods auto-sklearn and AutoKeras.

While we report only the top-performing baselines in Table 3, comprehensive performance metrics for all baselines can be found in Appendix B. Our results establish a new benchmark for the MedMNIST collection, highlighting that in data-rich scenarios, topological vectors offer competitive performance but generally do not outperform DL models across most medical imaging domains. With a few exceptions, the Topo-Med model achieves results within 1-10% of the best-performing DL models. This indicates that while topological vectors are a promising alternative in limited data settings, they often fall short in data-rich environments. However, effective integration of topological vectors with DL models holds significant potential for developing more robust and accurate models, a direction we plan to explore in our future work.

High vs. Low Resolution. We also analyzed how image resolution affects the performance of our topological model. As shown in Section 4.2, while higher resolution generally leads to improved performance, this is not always the case. Blue percentages indicate instances where the high-resolution model outperforms, while red percentages highlight where the low-resolution model does better. Overall, we interpret that while TDA methods offer comparable performance with greater computational efficiency on low-resolution images, higher resolution is preferable when performance is the priority, as it better captures topological patterns in medical imaging.

ML Classifiers. Comparing the performances of the ML part of our model, we observe that MLP is giving slightly better performance than XGBoost for 2D datasets. For 3D datasets, XGBoost is consistently better. The rough interpretation is that for small datasets, XGBoost would be a better choice as MLP needs a large training set for good performance.

Limitation. A key limitation of the TDA approach in medical imaging is its strong dependence on the specific domain. While it excels in certain diagnostic tasks, such as dermal, retinal, and blood-related imaging, it falls short when compared to deep learning (DL) models in other domains, like tissue or organ imaging (e.g., abdominal CT scans). When sufficient data is available, DL methods generally learn more effectively, adapting better to downstream tasks and delivering more accurate results. The reliance on topological feature vectors makes the TDA approach highly domain-specific, limiting its applicability across general medical imaging tasks. However, our experiments indicate that TDA vectors capture fundamentally different information from DL models and, when properly integrated, have the potential to enhance the performance of existing DL models.

5. Conclusion

In this study, we evaluated the effectiveness of topological features across various medical imaging domains. Our findings demonstrate that topological methods hold significant promise in addressing the limitations of DL models in data-limited settings, a common challenge in medical imaging. Notably, in 3D settings, topological vectors showed remarkable standalone performance, outperforming all DL models, even when they were augmented with topological features. However, in data-rich environments, DL models generally exhibited superior performance compared to TDA approaches. Importantly, our results highlight that integrating topological feature vectors can enhance the accuracy and robustness of DL models. Future research will focus on harnessing this potential to develop specialized clinical decision support systems tailored to specific medical domains by combining TDA outputs with DL models, ultimately aiming for more advanced and reliable medical image analysis systems.

Acknowledgements

This work was partially supported by the National Science Foundation under grants DMS-2202584, 2229417, and DMS-2220613 and by Simons Foundation under grant # 579977. The authors acknowledge the Texas Advanced Computing Center (TACC) at UT Austin for computational resources which contributed to the research results reported within this paper.

References

Mohamed A Abdou. Literature review: Efficient deep neural networks techniques for medical image analysis. *Neural Computing and Applications*, 34(8): 5791–5812, 2022.

Ravi Aggarwal et al. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. *NPJ digital medicine*, 4(1):65, 2021.

Dashti Ali, Aras Asaad, Maria-Jose Jimenez, Vidit Nanda, Eduardo Paluzo-Hidalgo, and Manuel Soriano-Trigueros. A survey of vectorization methods in topological data analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

Ekaba Bisong. Google automl: cloud vision. *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners*, pages 581–598, 2019.

Heang-Ping Chan, Ravi K Samala, Lubomir M Hadjiiski, and Chuan Zhou. Deep learning in medical image analysis. *Deep Learning in Medical Image Analysis: Challenges and Applications*, pages 3–21, 2020.

Yizhou Chen, Xu-Hua Yang, Zihan Wei, Ali Asghar Heidari, Nenggan Zheng, Zhicheng Li, Huiling Chen, Haigen Hu, Qianwei Zhou, and Qiu Guan. Generative adversarial networks in medical image augmentation: a review. *Computers in Biology and Medicine*, 144:105382, 2022.

Phillip Chlap, Hang Min, Nym Vandenberg, Jason Dowling, Lois Holloway, and Annette Haworth. A review of medical image data augmentation techniques for deep learning applications. *Journal of Medical Imaging and Radiation Oncology*, 65(5): 545–563, 2021.

Seungho Choe and Sheela Ramanna. Cubical homology-based machine learning: An application in image classification. *Axioms*, 11(3):112, 2022.

Baris Coskunuzer and Cuneyt Gurcan Akcora. Topological methods in machine learning: A tutorial for practitioners. *arXiv preprint arXiv:2409.02901*, 2024.

Tamal Krishna Dey and Yusu Wang. *Computational Topology for Data Analysis*. Cambridge University Press, 2022.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. *ICLR*, 2021.

Ruiwei Feng, Xiangshang Zheng, Tianxiang Gao, Jintai Chen, Wenzhe Wang, Danny Z Chen, and Jian Wu. Interactive few-shot learning: Limited supervision, better medical image segmentation. *IEEE Transactions on Medical Imaging*, 40(10): 2575–2588, 2021.

Matthias Feurer, Katharina Eggensperger, Stefan Falkner, Marius Lindauer, and Frank Hutter. Auto-sklearn 2.0: Hands-free automl via meta-learning. *The Journal of Machine Learning Research*, 23(1):11936–11996, 2022.

Anton François et al. Train-free segmentation in mri with cubical persistent homology. *arXiv preprint arXiv:2401.01160*, 2024.

Hiroshi Fujita. Ai-based computer-aided diagnosis (ai-cad): the latest review to read first. *Radiological physics and technology*, 13(1):6–19, 2020.

Bhagirathi Halalli and Aziz Makandar. Computer aided diagnosis-medical image analysis techniques. *Breast imaging*, 85:85–109, 2018.

Kaiming He et al. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016.

Felix Hensel, Michael Moor, and Bastian Rieck. A survey of topological machine learning methods. *Frontiers in Artificial Intelligence*, 4:52, 2021.

Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

Rushi Jiao, Yichi Zhang, Le Ding, Bingsen Xue, Jicong Zhang, Rong Cai, and Cheng Jin. Learning with limited annotations: a survey on deep semi-supervised learning for medical image segmentation. *Computers in Biology and Medicine*, page 107840, 2023.

Haifeng Jin, Qingquan Song, and Xia Hu. Autokeras: An efficient neural architecture search system. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 1946–1956, 2019.

Shizuo Kaji, Takeki Sudo, and Kazushi Ahara. Cubical ripser: Software for computing persistent homology of image and volume data. *arXiv preprint arXiv:2005.12692*, 2020.

Padmavathi Kora, Chui Ping Ooi, Oliver Faust, U Raghavendra, Anjan Gudigar, Wai Yee Chan, K Meenakshi, K Swaraja, Pawel Plawiak, and U Rajendra Acharya. Transfer learning techniques for medical image analysis: A review. *Biocybernetics and Biomedical Engineering*, 42(1):79–107, 2022.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.

Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021.

Ze Liu, Han Hu, Yutong Lin, Zhuliang Yao, Zhenda Xie, Yixuan Wei, Jia Ning, Yue Cao, Zheng Zhang, Li Dong, Furu Wei, and Baining Guo. Swin transformer v2: Scaling up capacity and resolution. In *International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

Sarah McGuire, Shane Jackson, Tegan Emerson, and Henry Kvinge. Do neural networks trained with topological features learn different internal representations? In *NeurIPS Workshop on Symmetry and Geometry in Neural Representations*, pages 122–136. PMLR, 2023.

Nikola Milosavljević, Dmitriy Morozov, and Primoz Skraba. Zigzag persistent homology in matrix multiplication time. In *SoCG*, pages 216–225, 2011.

Nina Otter, Mason A Porter, Ulrike Tillmann, Peter Grindrod, and Heather A Harrington. A roadmap for the computation of persistent homology. *EPJ Data Science*, 6:1–38, 2017.

Yaopeng Peng, Hongxiao Wang, Milan Sonka, and Danny Z Chen. Phg-net: Persistent homology guided medical image classification. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 7583–7592, 2024.

Ahmad Waleed Salehi, Shakir Khan, Gaurav Gupta, Bayan Ibrahimm Alabduallah, Abrar Almjally, Hadeel Alsolai, Tamanna Siddiqui, and Adel Mellit. A study of CNN and transfer learning in medical imaging: Advantages, challenges, future scope. *Sustainability*, 15(7):5930, 2023.

DR Sarvamangala et al. Convolutional neural networks in medical image understanding: a survey. *Evolutionary intelligence*, 15(1):1–22, 2022.

Fahad Shamshad, Salman Khan, Syed Waqas Zamir, Muhammad Haris Khan, Munawar Hayat, Fahad Shahbaz Khan, and Huazhu Fu. Transformers in medical imaging: A survey. *Medical Image Analysis*, page 102802, 2023.

Saeed Shurrab et al. Self-supervised learning methods and applications in medical imaging analysis: A survey. *PeerJ Computer Science*, 8:e1045, 2022.

Rishav Singh, Vandana Bharti, Vishal Purohit, Abhinav Kumar, Amit Kumar Singh, and Sanjay Kumar Singh. Metamed: Few-shot medical image classification using gradient-based meta-learning. *Pattern Recognition*, 120:108111, 2021.

Yashbir Singh, Colleen M Farrelly, Quincy A Hathaway, Tim Leiner, Jaidip Jagtap, Gunnar E Carlsson, and Bradley J Erickson. Topological data analysis in medical imaging: current state of the art. *Insights into Imaging*, 14(1):1–10, 2023.

Yara Skaf et al. Topological data analysis in biomedicine: A review. *Journal of Biomedical Informatics*, page 104082, 2022.

Eashwar Somasundaram, Adam Litzler, Raoul Wadhwa, Steph Owen, and Jacob Scott. Persistent homology of tumor ct scans is associated with survival in lung cancer. *Medical physics*, 48(11):7043–7051, 2021.

Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31, 2017.

Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *ICML*, pages 6105–6114. PMLR, 2019.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Ankur Yadav, Faisal Ahmed, Ovidiu Daescu, Reyhan Gedik, and Baris Coskunuzer. Histopathological cancer detection with topological signatures. In *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1610–1619. IEEE, 2023.

Jiancheng Yang, Xiaoyang Huang, Yi He, Jingwei Xu, Canqian Yang, Guozheng Xu, and Bingbing Ni. Reinventing 2d convolutions for 3d images. *IEEE Journal of Biomedical and Health Informatics*, 25 (8):3009–3018, August 2021a. ISSN 2168-2208. doi:

10.1109/jbhi.2021.3049452. URL http://dx.doi.org/10.1109/JBHI.2021.3049452.

Jiancheng Yang, Xiaoyang Huang, Yi He, Jingwei Xu, Canqian Yang, Guozheng Xu, and Bingbing Ni. Reinventing 2d convolutions for 3d images. *IEEE Journal of Biomedical and Health Informatics*, 25 (8):3009–3018, 2021b.

Jiancheng Yang, Rui Shi, Donglai Wei, Zequan Liu, Lin Zhao, Bilian Ke, Hanspeter Pfister, and Bingbing Ni. Medmnist v2-a large-scale lightweight benchmark for 2D and 3D biomedical image classification. *Scientific Data*, 10(1):41, 2023a. https://medmnist.com/.

Jiancheng Yang et al. MedMNIST v2-a large-scale lightweight benchmark for 2d and 3d biomedical image classification. *Scientific Data*, 10(1):41, 2023b. URL https://medmnist.com.

S Kevin Zhou et al. A review of deep learning in medical imaging. *Proceedings of the IEEE*, 109(5): 820–838, 2021.

Appendix A. Further Performance Metrics for Limited Data

In Tables 5 and 6, we present the accuracy results of our models in limited data settings. We further present the additional performance metrics for Topo-Med model in these datasets for limited data settings in Table 7.

Table 5: **2D Limited Data.** Accuracy results for vanilla-CNN and Topo-CNN models using ResNet-50, EfficientNetB0, and Swin Transformer backbones in limited data settings on the MedMNIST 2D datasets. The TDA columns display the standalone performance of topological vectors using an MLP (Topo-Med). For each backbone, the top AUC **improvements** are highlighted in **blue**, and the highest performance within each row for each dataset is **underlined**.

	BloodMNIST						
#	TDA	Res	T-Res	Eff	T-Eff	Swin	T-Swin
100	0.7366	0.7653	**0.7866**	**0.5077**	**0.7463**	0.4291	0.7638
250	0.8158	0.8834	**0.9062**	0.6942	0.8375	0.4505	0.8690
500	0.8603	**0.8942**	**0.9220**	0.7662	0.8787	0.4744	0.8781
1000	0.8948	**0.9395**	0.9304	0.9035	0.9073	0.4616	0.8994
2000	0.9176	**0.9594**	0.9494	0.9532	0.9281	**0.3975**	**0.9182**

	OCTMNIST						
#	TDA	Res	T-Res	Eff	T-Eff	Swin	T-Swin
100	0.3000	**0.4710**	0.4420	0.2480	0.3720	0.2390	0.3250
250	0.3410	**0.5630**	0.5060	0.3270	0.4810	0.314	0.402
500	0.3830	0.6230	**0.6310**	0.4830	0.5200	0.2500	0.3980
1000	0.4040	**0.7290**	0.6990	0.5750	0.6200	0.2510	0.4050
2000	0.3980	**0.7820**	0.7760	**0.6770**	**0.7820**	**0.2500**	**0.4120**

	PathMNIST						
#	TDA	Res	T-Res	Eff	T-Eff	Swin	T-Swin
100	0.4187	**0.6029**	**0.6418**	**0.5387**	**0.7077**	0.3050	0.6659
250	0.6383	0.6974	**0.7113**	0.6109	0.7072	0.1180	0.6760
500	0.5840	**0.8194**	0.7155	0.6469	0.7234	**0.0825**	**0.6947**
1000	0.6164	**0.8187**	0.7253	0.7227	0.6985	0.2056	0.6845
2000	0.6448	**0.8191**	0.7765	0.7926	0.7467	0.4721	0.6669

	DermaMNIST						
#	TDA	Res	T-Res	Eff	T-Eff	Swin	T-Swin
100	0.6050	**0.5840**	**0.6160**	**0.6643**	0.6195	0.6459	0.5800
250	0.5786	0.6324	0.6185	0.6464	0.6589	**0.6668**	0.6613
500	0.6030	0.6459	0.6429	0.6688	0.6264	**0.6693**	0.6524
1000	0.6115	0.6693	**0.6948**	0.6763	0.6788	0.6688	0.6579
2000	0.6289	0.7122	0.7032	**0.6863**	**0.7187**	0.6838	0.6633

	TissueMNIST						
#	TDA	Res	T-Res	Eff	T-Eff	Swin	T-Swin
100	0.3886	0.4007	0.394	0.3886	**0.4074**	0.3226	0.4014
250	0.3912	0.4181	0.3896	0.3789	0.3931	**0.1487**	**0.4358**
500	0.4033	0.4336	**0.4643**	**0.4018**	**0.4505**	0.3355	0.4568
1000	0.3977	**0.3711**	**0.4625**	**0.4630**	0.4625	0.3360	0.4583
2000	0.4088	0.4575	**0.5213**	0.4448	0.4738	0.3341	0.4685

	PneuMNIST						
#	TDA	Res	T-Res	Eff	T-Eff	Swin	T-Swin
100	0.6955	0.7612	0.7821	**0.6282**	**0.7949**	**0.8045**	0.7276
250	0.7532	0.8157	**0.8574**	0.8013	0.8237	**0.3750**	**0.8061**
500	0.8061	**0.8285**	**0.8750**	0.8189	0.8478	0.7837	0.8029
1000	0.8029	**0.8654**	0.8638	0.8317	0.8510	0.7308	0.8269
2000	0.8077	**0.8574**	0.8478	0.7965	0.8365	0.6250	0.8061

Table 6: **3D Limited Data Performances.** Accuracy results for vanilla-CNN and Topo-CNN models using ResNet-18 + 2.5D, ResNet-18 + 3D, and ResNet-18 + ACS backbones in limited data settings on the MedMNIST 3D datasets.

	AdrenalMNIST3D						
#	TDA	R18-ACS	T-R18-ACS	R18-3D	T-R18-3D	R18-2.5D	T-R18-2.5D
50	0.7282	0.7752	0.7685	**0.8054**	0.2819	0.5503	0.3926
100	0.7483	**0.8255**	0.2617	0.7819	0.2416	**0.7785**	**0.7886**
200	0.7685	0.5503	0.2919	**0.8054**	0.7550	0.6477	0.7685

	VesselMNIST3D						
#	TDA	R18-ACS	T-R18-ACS	R18-3D	T-R18-3D	R18-2.5D	T-R18-2.5D
50	0.7016	**0.8874**	0.8063	0.8848	0.8298	**0.8874**	0.8743
100	0.8717	**0.7880**	**0.8874**	**0.6545**	**0.8874**	**0.8874**	**0.8874**
200	0.8717	0.8874	**0.9031**	0.8115	0.8796	0.8874	0.8874

Table 7: Standalone performance of topological vectors with different sample sizes on MedMNIST datasets.

	TissueMNIST (8)				OCTMNIST (4)				BloodMNIST (8)			
Samples	Acc	AUC	Precision	Recall	Acc	AUC	Precision	Recall	Acc	AUC	Precision	Recall
100	0.3886	0.7273	0.4044	0.3746	0.3000	0.5379	0.2986	0.2980	0.7366	0.9129	0.7436	0.7308
250	0.3912	0.7345	0.4062	0.3776	0.3410	0.5691	0.3417	0.3410	0.8158	0.9487	0.8218	0.8118
500	0.4033	0.7421	0.4178	0.3887	0.3830	0.6043	0.3841	0.3810	0.8603	0.9629	0.8639	0.8591
1000	0.3977	0.7322	0.4073	0.3869	0.4040	0.6315	0.4124	0.4000	0.8948	0.9714	0.8961	0.8921
2000	0.4088	0.7289	0.4154	0.4018	0.3980	0.6381	0.4074	0.3940	0.9176	0.9778	0.9186	0.9170

	BreastMNIST (2)				PneuMNIST (2)				PathMNIST (9)			
Samples	Acc	AUC	Precision	Recall	Acc	AUC	Precision	Recall	Acc	AUC	Precision	Recall
100	0.7115	0.7617	0.7115	0.7115	0.6955	0.7466	0.6955	0.6955	0.5187	0.7660	0.5194	0.5174
250	0.7051	0.7490	0.7051	0.7051	0.7532	0.8283	0.7532	0.7532	0.6383	0.8379	0.6411	0.6376
500	0.7308	0.8490	0.7308	0.7308	0.8061	0.8587	0.8061	0.8061	0.5840	0.8223	0.5862	0.5816
1000	0.7372	0.8126	0.7372	0.7372	0.8029	0.8747	0.8029	0.8029	0.6164	0.8622	0.6246	0.6109
2000	0.7885	0.8307	0.7885	0.7885	0.8077	0.8525	0.8077	0.8077	0.6448	0.9088	0.6603	0.6312

Appendix B. Topo-Med vs. SOTA Baselines on Entire Dataset

In Table 8, we present the performance metrics of seven baseline models on the MedMNIST datasets (using 28x28 and 28x28x28 sized images), compared with our Topo-Med model, which incorporates topological vectors and an MLP. This comparison evaluates the standalone performance of topological vectors against SOTA models when applied to the entire dataset (Table 9). Baseline performances are reported from Yang et al. (2023b).

Table 8: **Topo-Med vs. DL.** Performance comparison across various DL models and Topo-Med model on the MedMNIST2D (28x28) and MedMNIST3D (28x28x28) datasets. The best performance is given in bold, and the second best is underlined.

	PathM		ChestM		DermaM		OCTM		PneuM		RetinaM	
Methods	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**
ResNet-18 (28)	0.983	0.907	0.768	0.947	0.917	0.735	0.943	0.743	0.944	0.854	0.717	0.524
ResNet-18 (224)	0.989	**0.909**	0.773	0.947	**0.920**	0.754	0.958	0.763	0.956	0.864	0.710	0.493
ResNet-50 (28)	**0.990**	0.911	0.769	0.947	0.913	0.735	0.952	0.762	0.948	0.854	0.726	0.528
ResNet-50 (224)	0.989	0.892	0.773	0.948	0.912	0.731	0.958	**0.776**	0.962	0.884	0.716	0.511
auto-sklearn	0.934	0.716	0.649	0.779	0.902	0.719	0.887	0.601	0.942	0.855	0.690	0.515
AutoKeras	0.959	0.834	0.742	0.937	0.915	0.749	0.955	0.763	0.947	0.878	0.719	0.503
Google AutoML	0.944	0.728	0.778	**0.948**	0.914	**0.768**	**0.963**	0.771	**0.991**	**0.946**	**0.750**	**0.531**
Topo-Med	0.942	0.683	**0.787**	0.530	0.904	0.669	0.710	0.450	0.845	0.762	0.728	0.458

	BreastM		BloodM		TissueM		OrganAM		OrganCM		OrganSM	
Methods	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**
ResNet-18 (28)	0.901	**0.863**	0.998	0.958	0.930	0.676	**0.997**	**0.935**	0.992	0.900	0.972	0.782
ResNet-18 (224)	0.891	0.833	0.998	0.963	0.933	0.681	0.998	0.951	**0.994**	**0.920**	0.974	0.778
ResNet-50 (28)	0.857	0.812	0.997	0.956	0.931	0.680	0.997	0.935	0.992	0.905	0.972	0.770
ResNet-50 (224)	0.866	0.842	0.997	0.950	0.932	0.680	0.998	0.947	0.993	0.911	**0.975**	0.785
auto-sklearn	0.836	0.803	0.984	0.878	0.828	0.532	0.963	0.762	0.976	0.829	0.945	0.672
AutoKeras	0.871	0.831	0.998	0.961	**0.941**	**0.703**	0.994	0.905	0.990	0.879	0.974	**0.813**
Google AutoML	**0.919**	0.861	**0.998**	**0.966**	0.924	0.673	0.990	0.886	0.988	0.877	0.964	0.749
Topo-Med	0.821	0.737	0.973	0.798	0.837	0.450	0.921	0.523	0.894	0.489	0.910	0.532

	OrganM3D		NoduleM3D		FractureM3D		AdrenalM3D		VesselM3D		SynapseM3D	
Methods	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**
ResNet-18 + 2.5D	0.977	0.788	0.838	0.835	0.587	0.451	0.718	0.772	0.748	0.846	0.634	0.696
ResNet-18 + 3D	**0.996**	**0.907**	0.863	0.844	0.712	0.508	0.827	0.721	0.874	0.877	0.820	0.745
ResNet-18 + ACS	0.994	0.900	0.873	0.847	0.714	0.497	0.839	0.754	**0.930**	**0.928**	0.705	0.722
ResNet-50 + 2.5D	0.974	0.769	0.835	0.848	0.552	0.397	0.732	0.763	0.751	0.877	0.669	0.735
ResNet-50 + 3D	0.994	0.883	0.875	0.847	0.725	0.494	0.828	0.745	0.907	0.918	**0.851**	**0.795**
ResNet-50 + ACS	0.994	0.889	0.886	0.841	**0.750**	**0.517**	0.828	0.758	0.912	0.858	0.719	0.709
auto-sklearn	0.977	0.814	**0.914**	**0.874**	0.628	0.453	0.828	**0.802**	0.910	0.915	0.631	0.730
AutoKeras	0.979	0.804	0.844	0.834	0.642	0.458	0.804	0.705	0.773	0.894	0.538	0.724
Topo-Med	0.837	0.554	0.808	0.736	0.653	0.480	**0.864**	0.769	0.920	0.887	0.730	0.730

Appendix C. Dataset Details

In our experiments, we utilized the publicly available MedMNIST dataset, which covers a diverse array of medical imaging domains and formats, comprising over 600,000 images. MedMNIST is a standardized and extensive collection of biomedical image datasets, supporting a range of tasks and scales—from 100 to 100,000 samples. It provides uniform preprocessing, predefined train-validation-test splits, and image sizes varying from 28x28 to 224x224 for 2D, and from 28x28x28 to 64x64x64 for 3D data. Detailed information about the dataset is presented in Table 9. More information can be found at https://medmnist.com

Table 9: MedMNIST2D and MedMNIST3D Dataset Details.

MedMNIST2D	**Data Modality**	**# Classes**	**# Samples**	**Train/Valid/Test**
PathMNIST	Colon Pathology	9	107,180	89,996 / 10,004 / 7,180
ChestMNIST	Chest X-ray	14	112,120	78,468 / 11,219 / 22,433
DermaMNIST	Dermatoscope	7	10,015	7,007 / 1,003 / 2,005
OCTMNIST	Retinal OCT	4	109,309	97,477 / 10,832 / 1,000
PneuMNIST	Chest X-ray	2	5,856	4,708 / 524 / 624
RetinaMNIST	Fundus Camera	5	1,600	1,080 / 120 / 400
BreastMNIST	Breast Ultrasound	2	780	546 / 78 / 156
BloodMNIST	Blood Cell Microscope	8	17,092	11,959 / 1,712 / 3,421
TissueMNIST	Kidney Cortex Microscope	8	236,386	165,466 / 23,640 / 47,280
OrganAMNIST	Abdominal CT	11	58,850	34,581 / 6,491 / 17,778
OrganCMNIST	Abdominal CT	11	23,660	13,000 / 2,392 / 8,268
OrganSMNIST	Abdominal CT	11	25,221	13,940 / 2,452 / 8,829
OrganMNIST3D	Abdominal CT	11	1,743	972 / 161 / 610
NoduleMNIST3D	Chest CT	2	1,633	1,158 / 165 / 310
AdrenalMNIST3D	Abdominal CT	2	1,584	1,188 / 98 / 298
FractureMNIST3D	Chest CT	3	1,370	1,027 / 103 / 240
VesselMNIST3D	Brain MRA	2	1,909	1,335 / 192 / 382
SynapseMNIST3D	Electron Microscope	2	1,759	1,230 / 177 / 352