PHG-Net: Persistent Homology Guided Medical Image Classification*

Yaopeng Peng
University of Notre Dame
ypeng4@nd.edu

Hongxiao Wang
University of Notre Dame
hwang21@nd.edu

Milan Sonka
University of Iowa
milan-sonka@uiowa.edu

Danny Z. Chen
University of Notre Dame
dchen@nd.edu

Abstract

Modern deep neural networks have achieved great successes in medical image analysis. However, the features captured by convolutional neural networks (CNNs) or Transformers tend to be optimized for pixel intensities and neglect key anatomical structures such as connected components and loops. In this paper, we propose a persistent homology guided approach (PHG-Net) that explores topological features of objects for medical image classification. For an input image, we first compute its cubical persistence diagram and extract topological features into a vector representation using a small neural network (called the PH module). The extracted topological features are then incorporated into the feature map generated by CNN or Transformer for feature fusion. The PH module is lightweight and capable of integrating topological features into any CNN or Transformer architectures in an end-to-end fashion. We evaluate our PHG-Net on three public datasets and demonstrate its considerable improvements on the target classification tasks over state-of-the-art methods.

1. Introduction

Deep neural networks (DNNs) are capable of learning useful image features based on their potent representations, and are widely used in medical image analysis. From AlexNet [17], VGG [25], to DenseNet [26, 10, 15], many convolutional neural network (CNN) architectures have been proposed for rich feature representations. Due to their nature, CNNs primarily focus on capturing local features. Vision Transformers [7, 21] have been proposed to procure global dependencies and long-range relationships by leveraging self-attention mechanism, which allows models to attend to different patches and learn their dependen-

cies. However, these models tend to neglect key global and robust anatomical structures (e.g., topological structures), such as connected components, loops, and voids. Medical images commonly contain tissues, organs, and lesions as connected components with specific patterns (e.g., loops and voids), but such structures and typologies are often overlooked by deep learning (DL) models.

In recent years, topological data analysis (TDA) [9] has been applied as a powerful methodology to analyze data in chemistry [20], medicine [6], biology [29], and other fields. Persistent homology (PH) is a most widely-used method of TDA. It tracks topological changes of object dynamics during the filtration process, where a lifespan is associated with these changes in the form of entity birth or death. The collection of such birth-death time pairs forms a persistence diagram (PD). But, effectively utilizing persistence diagrams in machine learning is not straightforward due to the multiset nature of the persistent homology building process.

A technique was proposed to input topological signatures into DNNs for 2D object shape and social network classification [12]. Additionally, a readout operation to aggregate node features into a graph representation was introduced [11]. Persistent landscapes were presented [2] as a method to summarize topological data into vectors to use in machine learning. In [1], persistent images were proposed as a stable representation that converts a persistence diagram into a finite-dimensional vector representation. A framework in [3] first encoded a persistence diagram into a vector and then learned the vectorization using a neural network. Another method was developed in [16] to first encode a persistence diagram into a list of persistent landscapes and then vectorize them for use in DL models

Medical images often exhibit intricate marked topological structures/patterns of clinical targets, thereby rendering persistent homology (PH) a promising computational technique that provides supplementary topological information and insights alongside pixel-oriented CNNs [9]. More
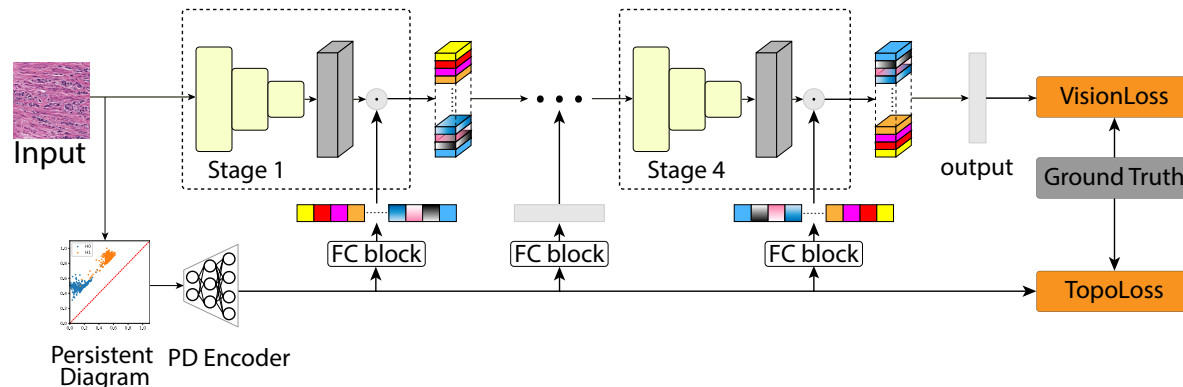
Figure 1. Illustrating the pipeline of our proposed PHG-Net. The details of the PD encoder are shown in Fig. 3(a). A fully connected block (FC block) is a two-layer MLP shown in Fig. 3(b). The blue points in a persistence diagram denote the 0-dimensional persistent homology (H0), and the orange points denote the 1-dimensional persistent homology (H1). VisionLoss and TopoLoss represent the vision loss and topological loss, respectively. ⊙ denotes matrix multiplication. We use a CNN backbone for illustration.

specifically, cubical persistence leverages image intensity values as filtration functions in its analytical process, which serves as a persistent homology tool that has been proven to be highly useful for medical image analysis. For example, in [13], it computed persistent curves and statistics, and these features were integrated into ResNet for skin lesion classification. In [18], the persistence of each region of interest (e.g., *birth_time-death_time*) was ranked, and the images were clustered into sub-architectural groups. In [23], topological features were combined with features extracted from the last CNN layer for tumor segmentation in histology slides. In [27], it first masked out the critical positions of breast MRI images using persistent homology and classified the masked images. In [8], a persistence diagram was first encoded into Betti curves and then integrated into a CNN network. In [28], it built topological-attention of consecutive slices for 3D anisotropic image segmentation. These methods often involved encoding a persistence diagram into a vector using mathematical tools, which is usually a non-trivial process. Another aspect that could be improved is that they concatenated topological features with the last CNN layer, which incorporates another type of (CNN) features, but lacks interactions between the topological features and multi-scale CNN features. Furthermore, the topological features are not involved when updating the gradients of the CNN or Transformer weights.

To address these issues, we propose a persistent homology-guided approach (PHG-Net) for medical image classification. PHG-Net directly processes persistence diagrams with a neural network and leverages the extracted topological features in multiple layers of CNN or vision Transformer (not just the last layer) to refine vision features at multiple scales. We treat persistence diagrams as point clouds and process them with a network (called the PH module or topological

branch) inspired by PointNet [24]. Code is available at https://github.com/yaoppeng/TopoClassification

Our main contributions are three-fold:

1. Inspired by PointNet [24], we develop a new approach to process persistence diagrams as point clouds rather than vectorized features, using a neural network. The motivation for our work is that the differences among topological structures of images can be represented by persistence diagrams and captured by a neural network, thus guiding the learning of CNNs or Transformers. This process is data-driven and learnable.

2. Our PHG-Net is capable of incorporating topological features into any base vision models, including CNNs and Transformers.

3. Our new PHG-Net approach achieves considerable improvements on three public datasets compared to state-of-the-art medical image classification methods.

2. Method

We use a CNN backbone to illustrate our PHG-Net. The overall pipeline of our PHG-Net approach is shown in Fig. 1. Given an image I, the persistence diagram (PD) of I is first computed, and a PointNet-like neural network is then applied to encode the PD into a feature vector. The feature vector is integrated into feature maps generated by the CNN at each of its ConvBlocks for feature map refinement.

In this section, we first provide a brief review of cubical persistent homology (the readers may refer to [9] for further exposition). We then describe how we design a neural network encoder to directly convert a PD into a feature vector. Finally, we present the process of using persistent homology to guide the CNN for feature refinement in medical image classification.
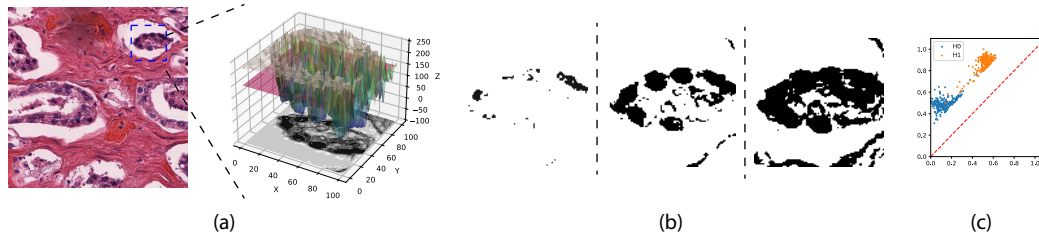
Figure 2. Illustrating the process of sub-level filtration. (a) A 3D plot of the sub-level set filtration of pixel intensities of a crop in an H&E stained image. The corresponding hematoxylin channel of the crop is projected on the xy plane. The z-axis is for threshold values. The red plane corresponds to one 2D slice of the 3D plot at threshold = 150. (b) Three thresholded binary images of the cropped image. As the threshold value increases or decreases, some connected components or loops are born or die. (c) The computed persistence diagram of the image crop (blue points denote the 0-D persistent homology (H0), and orange points denote the 1-D persistent homology (H1)).

2.1. Cubical Persistent Homology

Persistent homology (PH) is a powerful mathematical tool for analyzing topological properties of data. It is capable of identifying and quantifying important features of data that persist over a range of different spatial scales, thus providing insight into the underlying structures of the data. Fig. 2 illustrates the filtration process of PH. Given an image I and a series of threshold values $(\tau_0, \tau_1, \ldots, \tau_h)$, $\tau_0 < \tau_1 < \cdots < \tau_h$, in each step i, we threshold I as a sub-level image: $S_{\tau_i} = \{x : x \in I, f(x) \leq \tau_i\}$, where $f: I \to R$ is a function of pixel intensities. That is, for each τ_i, a sub-image S_{τ_i} is generated, and its connected components, loops, and voids are recorded. The sequence of $\{S_{\tau_i} \mid i = 0, 1, \ldots, h\}$, with $S_{\tau_0} \subseteq S_{\tau_1} \subseteq \cdots \subseteq S_{\tau_h}$, forms the filtration. Following the evolution of these sub-level images through the threshold sequence, the homology groups are induced as $\{H(S_{\tau_0}), H(S_{\tau_1}), \ldots, H(S_{\tau_h})\}$, where $H(S_{\tau_i})$ records the topological features of S_{τ_i} (e.g., connected component, loops, voids). When a new topological structure appears or is "born" at threshold τ_i and disappears, "dies", or merges with another topological structure at threshold τ_j, a tuple (τ_i, τ_j) is recorded and plotted as a 2D point with τ_i on the x-axis and τ_j on the y-axis; $\tau_j - \tau_i$ denotes the lifespan (i.e., persistence) of that topological structure. A structure with a long persistence means that the differences between it and its surroundings are significant and tend to be more salient (e.g., the boundary region of an image).

2.2. Persistence Diagram Encoder

Since PD is a most commonly used descriptor for persistent homology, there are an increasing number of methods to map PDs into a vector representation for machine learning tasks. However, these methods often rely on a mathematical encoding step, which is usually non-trivial and specifically designed, and may result in loss of certain information of the PDs. To address this issue, we propose a data-driven method to encode PDs.

Specifically, we treat each data point in a PD as a 2D fea-

ture, denoted as (f_{born}, f_{death}). A straightforward method would be to aggregate all the data points and obtain a 2D vector, for example, through max pooling or average pooling. However, such an operation may lead to losing too much information of the PD and may not effectively represent the entire filtration process. Therefore, we first increase the dimension of each data point so that the points can interact with one another. Then, we aggregate the information of all the points into a vector representation.

One key property of the points in a PD is permutation invariance, which means that the feature vector of the PD remains the same if some of the points exchange positions with one another. Note that this property is different from that of processing image data with a CNN, where exchanging pixels may result in different outputs. Based on this observation, we formulate the PD-encoding problem as:

$$f(p_1, p_2, \ldots, p_n) = g(m(p_1), m(p_2), \ldots, m(p_n)), \quad (1)$$

where p_i denotes a data point in a PD, $f(p_0, p_1, \ldots, p_{n-1})$ represents the output feature vector, and $m(p_i)$ is a function that maps a data point into a higher dimension so that it contains interactions of the data points and the information of the entire PD. To account for the permutation invariance property, we use a multi-layer perceptron (MLP) instead of using convolution. In Eq. (1), g is a function that aggregates information from all the points, such as through max pooling or average pooling. The entire vector learning process is carried out with a neural network, which is data-driven, and the vectorization process is supervised by the target task. Fig. 3(a) illustrates the details of the persistence diagram encoder that we design.

For persistent homology, let H_0 denote the connected components, H_1 denote the loops, and H_2 denote the voids in the 3D representation. However, a neural network taking points from these groups as input cannot differentiate which homology group that a point belongs to. Therefore, we add an additional one-hot vector to label each point. In other words, a point $(birth, death)$ in H_i is represented as
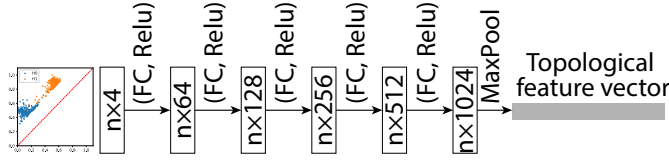
(a) The structure of our persistence diagram encoder (PDE).
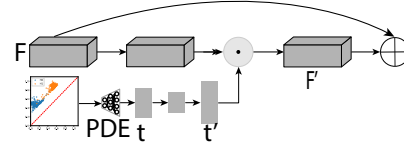
(b) A persistent homology guided residual block, PDE denotes persistence diagram encoder.

Figure 3. Illustrating (a) our persistence diagram encoder and (b) a persistent homology guided residual block (FC block). $\odot$ and $\oplus$ are matrix multiplication and summation, respectively.

$(birth, death, 0, \ldots, 0, 1, 0, \ldots, 0)$, in which the only 1 is located in the i-th position of the 0-1 sequence.

2.3. Persistent Homology Guidance (PHG)

The persistent homology features capture global topological information of an image. Previous methods typically utilized these features by integrating them with features learned by CNNs or Transformers. For example, in [13], it concatenated PH features and CNN features just before the classification step. In [8], PH features were treated as a teacher for adjusting CNN features. However, in concatenation based methods, gradient values of multiple branches during back-propagation are computed separately, and there are no direct interactions between the gradients of different branches. Moreover, concatenation just before the classification step lacks interactions between topological features and CNN features at multi-scale levels. To tackle this issue, we propose integrating topological supervision with each CNN or Transformer block. Additionally, we add a fully connected block (FC block) to regulate the intervention of the topological branch in each CNN or Transformer block (see Fig. 3(b)). This also enables us to refine the feature maps in multi-scales using topological features.

Given an intermediate CNN feature map F, $F \in \mathbb{R}^{H \times W \times C}$, our PHG-Net integrates a topological feature vector t into F, where $t \in \mathbb{R}^M$ is the output of the persistent homology encoder (see Eq. (1)). To ensure flexibility and applicability to any feature map, we employ a gate mechanism in the process:

$$t' = f(t, W) = \sigma(W_1(Relu(W_2(t)))), \quad (2)$$

where $t' \in R^C$ is the processed topological feature vector that will be used to refine the feature map F, t denotes the output of the PD encoder, $W_1 \in R^{\frac{C}{r} \times C}$ and $W_2 \in R^{M \times \frac{C}{r}}$ are the parameters of two MLP layers, σ is the sigmoid activation function, and r is a parameter for balancing the tradeoff between the performance and complexity. Afterwards, the topological guidance is formulated as:

$$F' = F \otimes t', \quad (3)$$

where F' is the refined feature map and $\otimes$ is element-wise multiplication. During the refinement, the feature vector

$t' \in R^C$ will first be broadcast along the spatial dimension of F'. Fig. 3(b) illustrates a persistent homology guided residual block.

2.4. Loss Function

The whole model is optimized with the following loss $\mathcal{L}$:

$$\mathcal{L} = \mathcal{L}_V(\tilde{y}_v, y) + \alpha \mathcal{L}_{Topo}(\tilde{y}_{topo}, y), \quad (4)$$

where $\mathcal{L}_V$ and $\mathcal{L}_{Topo}$ are the cross-entropy losses of the vision model and topological branches respectively; $\tilde{y}_v$, $\tilde{y}_{topo}$, and y denote the vision branch output, topological branch output, and ground truth respectively, and α is a hyperparameter for balancing the CNN and topological losses. See Fig. 1 for more details.

3. Experiments

3.1. Datasets

We evaluate our proposed PHG-Net approach using the following three datasets.

ISIC 2018: A skin lesion dataset for predicting 7 classes of skin disease lesions. It consists of 10,015 training images and 193 validation images [5]; the test set is unavailable. For fair comparison, we follow the dataset split strategy in [30], with which five-fold cross validation is conducted.

Prostate Cancer Classification: A dataset of hematoxylin and eosin (H&E) stained prostate cancer images that consists of 77 whole slide images (WSIs) from 19 patients [18]. The WSIs are divided into regions of interest (RoIs) of 512×512 pixels each. 5,182 RoIs are generated and three classes of prostate cancer are to be predicted. We also conduct five-fold cross validation.

CBIS-DDSM: A Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM). It contains 1,566 participants and 6,775 studies, which are categorized as benign or malignant [19]. We use the train/test split given by the dataset provider, in which 20% of the cases are for testing and the rest for training.

3.2. Experimental Setup

Common data augmentation techniques such as random flipping, color jittering, random rotation, and cropping are

Table 1. Experimental results on the ISIC 2018 dataset. Reported p-values reflect statistical significance of the improvement achieved by SwinV2-B + PHG over SwinV2-B, assessed by paired t-test; "–" is reported if SwinV2-B + PHG performance is lower than SwinV2-B for a specific evaluation metric.

Method	Accuracy	AUC	Sensitivity	Specificity
ResNet152 [10]	88.83±0.45	97.52±0.15	80.85±0.16	97.01±0.28
ResNet152 [10] + PHG	90.03±0.24	98.14±0.23	**83.75±0.30**	97.12±0.33
SENet154 [14]	89.80±0.37	97.92±0.27	81.96±0.37	96.90±0.25
SENet154 [14] + PHG	90.83±0.23	98.67±0.30	82.04±0.41	97.12±0.19
SwinV2-B [21]	90.85±0.34	97.99±0.38	82.23±0.27	**97.32±0.27**
SwinV2-B [21] + PHG	**91.92±0.27**	**98.97±0.42**	83.14±0.36	97.28±0.35
p-value	0.0006	0.005	0.002	–

Table 2. Experimental results on the Prostate Cancer dataset. Reported p-values reflect statistical significance of the improvement achieved by SwinV2-B + PHG over SwinV2-B, assessed by paired t-test.

Method	Accuracy	AUC	Sensitivity	Specificity
ResNet152 [10]	92.96±0.28	98.02±0.21	96.54±0.18	97.44±0.34
ResNet152 [10] + PHG	94.01±0.28	98.80±0.30	96.91±0.22	96.67±0.41
SENet154 [14]	93.73±0.39	98.67±0.34	94.84±0.50	96.38±0.28
SENet154 [14] + PHG	97.99±0.21	99.72±0.29	96.55±0.33	**98.26±0.40**
SwinV2-B [21]	95.21±0.33	98.61±0.31	97.22±0.34	97.99±0.29
SwinV2-B [21] + PHG	**98.64±0.27**	**99.83±0.24**	**98.34±0.29**	98.17±0.17
p-value	$\ll 0.001$	0.0001	0.0005	0.265

Table 3. Experimental results on the CBIS-DDSM dataset demonstrate that SwinV2-B + PHG significantly outperforms all the other tested methods for all the evaluation metrics. The p-values are obtained by paired t-test between SwinV2-B + PHG and SwinV2-B.

Method	Accuracy	AUC	Sensitivity	Specificity
ResNet152 [10]	71.16±0.34	78.75±0.45	73.19±0.55	71.65±1.08
ResNet152 [10] + PHG	74.34±0.31	79.35±0.36	71.74±1.10	72.03±0.49
SENet154 [14]	71.96±0.19	79.01±0.29	72.02±0.34	71.52±0.62
SENet154 [14] + PHG	75.66±0.26	82.23±0.42	73.53±0.29	73.07±0.21
SwinV2-B [21]	73.51±0.22	81.58±0.54	72.92±0.31	72.28±0.20
SwinV2-B [21] + PHG	**77.23±0.37**	**83.39±0.43**	**75.89±0.29**	**74.83±0.38**
p-value	$\ll 0.001$	0.0004	$\ll 0.001$	$\ll 0.001$

used. The network is optimized using the Adam optimizer with an initial learning rate of 0.0001, β_1 value of 0.9, and β_2 value of 0.999. A polynomial learning rate decay with a power of 0.9 is applied. The maximum number of training epochs is set to 1000. The value of α in Eq. (4) is set to 0.1. The parameter r for balancing the trade-off between performance and complexity is empirically set to 8. All experiments are conducted on an NVIDIA P100 GPU using PyTorch. Four common evaluation metrics are used: accuracy (Acc), area under the receiver operating characteristic (ROC) curve (AUC), sensitivity (Sen), and specificity (Spe).

We utilize the GUDHI package [22] to compute persistence diagrams. The points in a persistence diagram are first scaled to the range of $[0, 1]$ and then normalized. Additionally, a one-hot marker indicating which homology group each point belongs to is added. Any persistence value smaller than 10 is ignored. The points within each homology group are sorted in decreasing order based on their persistence values. For each persistence diagram, we experi-

mented with selecting the top n = 50, 100, 150, 200, and 250 points (see Eq. (1)). We found that selecting n = 150 points yields saturated performance while maintaining relatively low computation, as most of the points are very close to the diagonal line and are considered as noise points.

Moreover, we add dummy points of $(0, 0)$ to a persistence diagram if the number of points in it is less than 150. All the points are given an additional coordinate to indicate that a point is an added one (marked by 0) or originally presented one (marked by 1).

3.3. Results and Analysis

We evaluate the effectiveness and robustness of our PHG-Net approach by utilizing two common CNN backbones (ResNet152 [10] and SENet154 [14]), which performed well on medical image datasets (e.g., used by the team attaining top-1 score in the ISIC 2018 Challenge), as well as a latest Transformer, Swin Transformer v2 (SwinV2-B) [21]. These models are evaluated without us-

Table 4. Ablation study on the ISIC 2018 dataset. Reported p-value$_1$ was obtained using paired t-test between settings 12 and 13 in this table, measuring the statistical significance of the differences between KD and our PHG. "–" is reported since the Spe score of 13 is lower than 12. Reported p-value$_2$ was obtained using paired t-test between settings 10 and 13, measuring the statistical significance between PLLay and our PD encoder. "Concat" means concatenating CNN features and topological features at the last CNN layer.

	Backbone		PD Module			PD Fusion			Acc	AUC	Sen	Spe
	SENet154	SwinV2-B	PersLay	PLLay	Our PD	Concat	KD	PHG				
1			✓						65.69±0.26	82.30±0.55	36.48±0.21	91.72±0.36
2				✓					67.13±0.26	82.02±0.49	32.12±0.25	91.97±0.48
3					✓				70.72±0.43	91.66±0.52	53.01±0.72	94.62±0.36
4	✓		✓					✓	90.02±0.68	97.37±0.38	81.80±0.61	97.01±0.70
5	✓			✓				✓	90.27±0.21	97.04±0.65	80.66±0.71	96.82±0.45
6	✓				✓	✓			89.97±0.34	97.34±0.47	82.21±0.57	96.50±0.39
7	✓				✓		✓		90.27±0.29	96.94±0.48	82.09±0.37	96.47±0.64
8	✓				✓			✓	90.83±0.23	98.67±0.30	82.04±0.41	97.12±0.19
9		✓	✓					✓	91.00±0.44	98.23±0.17	81.76±0.19	97.40±0.25
10		✓		✓				✓	91.12±0.29	98.46±0.31	82.33±0.52	96.78±0.41
11		✓			✓	✓			91.18±0.19	98.09±0.26	81.99±0.24	**97.68±0.37**
12		✓			✓		✓		91.00±0.44	98.20±0.38	82.42±0.59	97.45±0.40
13		✓			✓			✓	**91.92±0.27**	**98.97±0.42**	**83.14±0.36**	97.28±0.35
	p-value$_1$								0.016	0.004	0.048	–
	p-value$_2$								0.002	0.06	0.021	0.072

Table 5. Ablation study on the Prostate Cancer dataset. Reported p-value$_1$ was obtained using paired t-test between settings 12 and 13 in this table, measuring the statistical significance between KD and our PHG. Reported p-value$_2$ was obtained using paired t-test between settings 10 and 13, measuring the statistical significance between PLLay and our PD encoder. "Concat" means concatenating CNN features and topological features at the last CNN layer.

	Backbone		PD Module			PD Fusion			Acc	AUC	Sen	Spe
	SENet154	SwinV2-B	PersLay	PLLay	Our PD	Concat	KD	PHG				
1			✓						75.31±0.46	86.43±0.53	67.37±0.57	84.42±0.38
2				✓					77.34±0.30	89.84±0.56	66.58±0.33	85.75±0.22
3					✓				82.35±0.46	92.53±0.20	76.37±0.48	87.46±0.58
4	✓		✓					✓	96.91±0.43	99.68±0.25	95.92±0.40	98.09±0.44
5	✓			✓				✓	97.11±0.30	99.68±0.43	96.13±0.47	98.21±0.39
6	✓				✓	✓			97.02±0.55	99.45±0.39	96.28±0.30	**98.40±0.47**
7	✓				✓		✓		96.57±0.29	99.55±0.27	96.57±0.29	97.91±0.28
8	✓				✓			✓	97.99±0.21	99.72±0.29	96.55±0.33	98.26±0.40
9		✓	✓					✓	98.21±0.31	99.54±0.46	97.54±0.38	97.59±0.24
10		✓		✓				✓	98.04±0.26	99.06±0.27	97.31±0.24	97.54±0.19
11		✓			✓	✓			98.04±0.29	98.12±0.25	**98.45±0.41**	97.12±0.28
12		✓			✓		✓		98.11±0.40	99.75±0.21	97.24±0.32	97.77±0.22
13		✓			✓			✓	**98.64±0.27**	**99.83±0.24**	98.34±0.29	98.17±0.17
	p-value$_1$								0.039	0.59	0.0004	0.012
	p-value$_2$								0.007	0.001	0.0002	≪0.001

ing extra training data. To demonstrate the statistical significance of the performance improvements yielded by our PHG-Net, we compute p-values (with paired t-test) between SwinV2-B + PHG (our model) and SwinV2-B by running each experiment 5 rounds. For each round, we select 5 epochs that achieve the top-5 best accuracy and take their average as the final result for robust evaluation.

The ISIC 2018 and Prostate Cancer datasets are for multi-class classification tasks. Thus, we report the OvR (One-vs-Rest) scores for AUC, Specificity, and Sensitivity. In each time, we take one class as positive and the other classes as negative, under which we compute the score for each class c. Finally, the average score of all the classes is reported.

Tables 1, 2, and 3 present the experimental results. We find that our PHG-Net approach improves, for example, the accuracy by 1.07%, 3.43%, and 3.72% on the ISIC 2018, prostate cancer, and CBIS-DDSM datasets, respectively,

Table 6. Ablation study on the CBIS-DDSM dataset. Reported p-value$_1$ was obtained using paired t-test between settings 12 and 13 in this table, measuring the statistical significance between KD and our PHG. Reported p-value$_2$ was obtained using paired t-test between settings 10 and 13, measuring the statistical significance between PLLay and our PD encoder. "Concat" means concatenating CNN features and topological features at the last CNN layer.

	Backbone		PD Module			PD Fusion			Acc	AUC	Sen	Spe
	SENet154	SwinV2-B	PersLay	PLLay	Our PD	Concat	KD	PHG				
1			✓						61.38±0.32	55.05±0.10	55.60±0.25	57.72±0.14
2				✓					63.23±0.26	63.06±0.21	60.36±0.37	59.40±0.35
3					✓				64.02±0.31	62.44±0.48	58.29±0.44	60.17±0.19
4	✓		✓					✓	74.87±0.30	79.71±0.33	74.00±0.35	73.13±1.31
5	✓			✓				✓	74.60±0.40	81.66±0.32	75.02±0.25	73.69±0.35
6	✓				✓	✓			72.36±0.64	80.27±0.57	73.58±0.47	72.89±0.46
7	✓				✓		✓		73.01±0.47	81.90±0.35	74.40±0.46	73.01±0.77
8	✓				✓			✓	75.66±0.26	82.23±0.42	73.53±0.29	73.07±0.21
9		✓	✓					✓	73.71±0.30	81.26±0.27	72.64±0.29	72.54±0.30
10		✓		✓				✓	74.06±0.45	82.21±0.53	73.45±0.47	73.04±0.42
11		✓			✓	✓			73.58±0.27	82.04±0.51	74.11±0.45	73.44±0.51
12		✓			✓		✓		74.34±0.40	82.41±0.43	74.03±0.46	73.87±0.29
13		✓			✓			✓	**77.23±0.37**	**83.39±0.43**	**75.89±0.29**	**74.83±0.38**
	p-value$_1$								≪0.001	0.007	≪0.001	0.002
	p-value$_1$								≪0.001	0.005	≪0.001	0.0001

Table 7. Time and parameter complexities of our PHG-Net approach on an NVIDIA P100 GPU; w/ share and w/o share represent using one PD encoder and multiple PD encoders for multi-scale vision feature fusion, respectively.

Method	Input Size	Acc			# Params.	FLOPs	FPS
		ISIC 2018	Prostate	CBIS-DDSM			
SwinV2-B	(3, 224, 224)	90.85	95.21	73.51	86.913 M	20.370 G	14.937
SwinV2-B+PLLay+PHG	(3, 224, 224) (4, 300)	91.12	98.04	74.06	95.073 M	20.375 G	20.787
SwinV2-B+PD+PHG w/ share	(3, 224, 224) (4, 300)	91.92	98.64	77.23	98.065 M	20.796 G	15.914
SwinV2-B+PD+PHG w/o share	(3, 224, 224) (4, 300)	91.95	98.70	77.04	108.892 M	22.062 G	17.059

when using the SwinV2-B backbone. These performance improvements and the p-values verify the effectiveness of our approach, especially on the prostate cancer and breast mammography datasets, where discernible changes in appearances or topological structures occur when cancer is present. For instance, in an H&E-stained prostate image, the number of nuclei is likely to increase if cancer cells are found in the image. In breast mammography images, tumor areas are often characterized by dense masses or clusters of micro-calcifications. In some cases, they may be present as tiny calcium deposits that appear as small white specks or clusters in the images. These topological structures, such as connected components or loops, can differentiate the images containing them from normal images.

3.4. Ablation Study

To examine the effects of our new persistence diagram (PD) encoder and PHG mechanism, we compare our methods with two common persistence diagram vectorization-based methods: (1) PersLay [3] (it encodes a persistence diagram and then refines the PD using an FC layer, where the persistence silhouette setting [4] is applied), and (2) PLLay [16] (it utilizes a weighted persistent landscape to extract topological features). Specifically, we pass the topological features extracted by these methods through the PH branch and compare their performances. Furthermore, we investigate the effect of using only topological features in the models. Additionally, we compare our approach with a typical knowledge distillation method (KD) [8], but replace its backbone with SENet154 and SwinV2-B, respectively, for fair comparison.

Tables 4, 5, and 6 show that our PD encoder outperforms the PersLay and PLLay methods, while our PHG outperforms the KD and concatenation (Concate, which concatenates CNN features and topological features at the last CNN layer) methods, since our PHG refines CNN features at multiple scales. This confirms the effectiveness of our schemes for constructing the PD encoder and PH guidance. It is worth noting that the methods relying **solely** on topological information (PersLay, PLLay, and our PD encoder) do not perform well on these datasets. Note that PersLay is a general form of persistence landscape [2], persistence silhouette [4], and persistence images [1] by utilizing different hyper-parameters, and we found that there is no obvious dif-
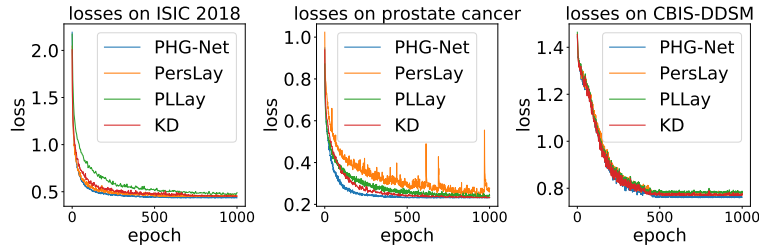
Figure 4. Validation losses of different persistent homology based methods and the KD (with our PD encoder) method.
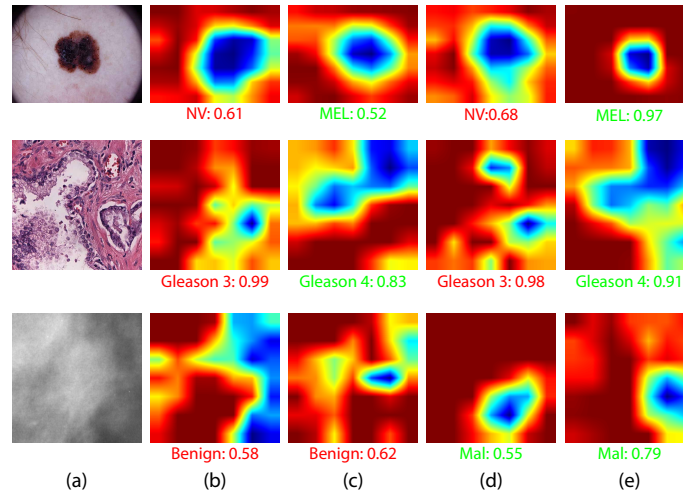

Figure 5. Qualitative examples of different persistent homology based methods and the KD (with our PD encoder) method. The 1st, 2nd, and 3rd rows are for samples from the ISIC 2018, Prostate Cancer, and CBIS-DDSM datasets, respectively. Columns (a)-(e) are for input images, heat-maps by PLLay, PersLay, KD, and PHG-Net, respectively. **Green** and **red** correspond to correct and incorrect predictions, respectively. The green or red floating-point number under each heat-map represents the top-1 prediction confidence (the higher the better).

ference among the settings of these PH methods. Thus, our approach also outperforms these PH methods. The p-values in Tables 4, 5, and 6 demonstrate the statistical significance of our PD encoder and PHG module.

3.5. Time and Parameter Complexity

To examine the time and parameter complexities of our PHG-Net, we present the number of parameters, floating point operations (FLOPs), and frames per second (FPS) for our approach and the SwinV2-B model in Table 7 for comparison.

In Table 7, $(3, 224, 224)$ denotes the size of an input image, while $(4, 300)$ stands for the size of the corresponding persistence diagram. It can be observed from Table 7 that the additional parameters and computation costs caused by our PHG-Net are quite limited. Furthermore, sharing the PD Encoder at multiple scales will result in fewer parameters and FLOPs without reducing the performance.

3.6. Qualitative Results

The validation losses of different persistent homology based methods and the KD method on the three datasets

using the backbone of SwinV2-B [14] are shown in Fig. 4.

Some qualitative examples based on the backbone of SwinV2-B on the three datasets are shown in Fig. 5, which demonstrate that our PHG-Net is capable of helping the network concentrate on the correct regions.

4. Conclusions

In this paper, we proposed to extract and encode persistence diagrams of medical images as important topological features for classification tasks using a neural network (PD encoder) that we designed. Our new approach, PHG-Net, is data-driven, learnable, and superior in performance over known topology based methods. We further developed a persistence diagram (PD) guided mechanism, which incorporates PD features into CNN or Transformer for co-training of DL models. Experiments on three datasets validated the effectiveness of our PHG-Net. Complexity analysis showed that the extra costs introduced by our approach are quite small. It is expected that our PHG-Net approach will provide a new perspective and paradigm for combining topological features with CNNs/Transformers to improve AI model performances for medical image analysis tasks.

References

[1] Henry Adams, Tegan Emerson, Michael Kirby, Rachel Neville, Chris Peterson, Patrick Shipman, Sofya Chepushtanova, Eric Hanson, Francis Motta, and Lori Ziegelmeier. Persistence images: A stable vector representation of persistent homology. *Journal of Machine Learning Research*, 18(8):1–35, 2017.

[2] Peter Bubenik et al. Statistical topological data analysis using persistence landscapes. *Journal of Machine Learning Research*, 16(1):77–102, 2015.

[3] Mathieu Carrière, Frédéric Chazal, Yuichi Ike, Théo Lacombe, Martin Royer, and Yuhei Umeda. PersLay: A neural network layer for persistence diagrams and new graph topological signatures. In *International Conference on Artificial Intelligence and Statistics*, pages 2786–2796. PMLR, 2020.

[4] Frédéric Chazal, Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, and Larry Wasserman. Stochastic convergence of persistence landscapes and silhouettes. In *Proceedings of the 30th Annual Symposium on Computational Geometry*, pages 474–483, 2014.

[5] Noel Codella, Veronica Rotemberg, Philipp Tschandl, M Emre Celebi, Stephen Dusza, David Gutman, Brian Helba, Aadi Kalloo, Konstantinos Liopyris, Michael Marchetti, et al. Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (ISIC). *arXiv preprint arXiv:1902.03368*, 2019.

[6] Meryll Dindin, Yuhei Umeda, and Frederic Chazal. Topological data analysis for arrhythmia detection through modular neural networks. In *Advances in Artificial Intelligence: 33rd Canadian Conference on Artificial Intelligence*, pages 177–188. Springer, 2020.

[7] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[8] Shiyi Du, Qicheng Lao, Qingbo Kang, Yiyue Li, Zekun Jiang, Yanfeng Zhao, and Kang Li. Distilling knowledge from topological representations for pathological complete response prediction. In *Medical Image Computing and Computer Assisted Intervention–MICCAI, Part II*, pages 56–65. Springer, 2022.

[9] Herbert Edelsbrunner and John Harer. *Computational Topology: An Introduction*. American Mathematical Soc, 2010.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[11] Christoph Hofer, Florian Graf, Bastian Rieck, Marc Niethammer, and Roland Kwitt. Graph filtration learning. In *International Conference on Machine Learning*, pages 4314–4323. PMLR, 2020.

[12] Christoph Hofer, Roland Kwitt, Marc Niethammer, and Andreas Uhl. Deep learning with topological signatures. *Advances in Neural Information Processing Systems*, 30, 2017.

[13] Chuan-Shen Hu, Austin Lawson, Jung-Sheng Chen, Yu-Min Chung, Clifford Smyth, and Shih-Min Yang. TopoResNet: A hybrid deep learning architecture and its application to skin lesion classification. *Mathematics*, 9(22):2924, 2021.

[14] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7132–7141, 2018.

[15] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4700–4708, 2017.

[16] Kwangho Kim, Jisu Kim, Manzil Zaheer, Joon Sik Kim, Frédéric Chazal, and Larry Wasserman. PLLay: Efficient topological layer based on persistence landscapes. *Advances in Neural Information Processing Systems*, 2020.

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

[18] Peter Lawson, Andrew B Sholl, J Quincy Brown, Brittany Terese Fasy, and Carola Wenk. Persistent homology for the quantitative evaluation of architectural features in prostate cancer histology. *Scientific Reports*, 9(1):1139, 2019.

[19] Rebecca Sawyer Lee, Francisco Gimenez, Assaf Hoogi, Kanae Kawai Miyake, Mia Gorovoy, and Daniel L Rubin. A curated mammography data set for use in computer-aided detection and diagnosis research. *Scientific Data*, 4(1):1–9, 2017.

[20] Yongjin Lee, Senja D Barthel, Paweł Dłotko, S Mohamad Moosavi, Kathryn Hess, and Berend Smit. Quantifying similarity of pore-geometry in nanoporous materials. *Nature Communications*, 8(1):1–8, 2017.

[21] Ze Liu, Han Hu, Yutong Lin, Zhuliang Yao, Zhenda Xie, Yixuan Wei, Jia Ning, Yue Cao, Zheng Zhang, Li Dong, et al. Swin Transformer v2: Scaling up capacity and resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12009–12019, 2022.

[22] Clément Maria, Jean-Daniel Boissonnat, Marc Glisse, and Mariette Yvinec. The Gudhi library: Simplicial complexes and persistent homology. In *Mathematical Software–ICMS 2014: 4th International Congress*, pages 167–174. Springer, 2014.

[23] Talha Qaiser, Yee-Wah Tsang, Daiki Taniyama, Naoya Sakamoto, Kazuaki Nakane, David Epstein, and Nasir Rajpoot. Fast and accurate tumor segmentation of histology images using persistent homology and deep convolutional features. *Medical Image Analysis*, 55:1–14, 2019.

[24] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. PointNet: Deep learning on point sets for 3D classification and segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 652–660, 2017.

[25] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[26] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent

Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–9, 2015.

[27] Fan Wang, Saarthak Kapse, Steven Liu, Prateek Prasanna, and Chao Chen. TopoTxR: A topological biomarker for predicting treatment response in breast cancer. In *Information Processing in Medical Imaging: 27th International Conference*, pages 386–397. Springer, 2021.

[28] Jiaqi Yang, Xiaoling Hu, Chao Chen, and Chialing Tsai. A topological-attention ConvLSTM network and its application to EM images. In *Medical Image Computing and Computer Assisted Intervention–MICCAI, Part I*, pages 217–228. Springer, 2021.

[29] Yuan Yao, Jian Sun, Xuhui Huang, Gregory R Bowman, Gurjeet Singh, Michael Lesnick, Leonidas J Guibas, Vijay S Pande, and Gunnar Carlsson. Topological methods for exploring low-density states in biomolecular folding pathways. *The Journal of Chemical Physics*, 130(14):04B614, 2009.

[30] Jiaxin Zhuang, Weipeng Li, Siyamalan Manivannan, Roy Wang, JianGuo Zhang, Jihan Liu, Jiahui Pan, Gongfa Jiang, and Ziyu Yin. Skin lesion analysis towards melanoma detection using deep neural network ensemble. *ISIC Challenge*, 2018(2):1–6, 2018.