



Article TopoResNet: A Hybrid Deep Learning Architecture and Its Application to Skin Lesion Classification [†]

Chuan-Shen Hu ¹^(b), Austin Lawson ², Jung-Sheng Chen ^{3,*}^(b), Yu-Min Chung ^{4,*,‡}^(b), Clifford Smyth ⁵^(b) and Shih-Min Yang ⁶

- ¹ Department of Mathematics, National Taiwan Normal University, Taipei City 11365, Taiwan; peterbill26@hotmail.com
- ² Department of Mathematics, University of Tennessee Knoxville, Knoxville, TN 37916, USA; alawso50@utk.edu
- ³ Department of Medical Research, E-Da Hospital, Kaohsiung City 824410, Taiwan
- ⁴ Eli Lilly and Company, Indianapolis, IN 46225, USA
- ⁵ Department of Mathematics and Statistics, University of North Carolina at Greensboro, Greensboro, NC 27412, USA; cdsmyth@uncg.edu
- ⁶ Department of Computer Science and Information Engineering, National Taiwan Normal University, Taipei City 11365, Taiwan; ethansmin.yang@gmail.com
- Correspondence: ed113187@edah.org.tw (J.-S.C.); yumchung@alumni.iu.edu (Y.-M.C.); Tel.: +886-7-6151806 (J.-S.C.)
- † This paper is an extended version of our paper published in 2018 IEEE International Conference on Big Data, Seattle, WA, USA, 10–13 December 2018; pp. 100–105. doi:10.1109/BigData.2018.8622175.
- ‡ The work was done when Y.-M. Chung was employed at University of North Carolina at Greensboro.

Abstract: The application of artificial intelligence (AI) to various medical subfields has been a popular topic of research in recent years. In particular, deep learning has been widely used and has proven effective in many cases. Topological data analysis (TDA)—a rising field at the intersection of mathematics, statistics, and computer science—offers new insights into data. In this work, we develop a novel deep learning architecture that we call *TopoResNet* that integrates topological information into the residual neural network architecture. To demonstrate TopoResNet, we apply it to a skin lesion classification problem. We find that TopoResNet improves the accuracy and the stability of the training process.

Keywords: deep learning; topological data analysis; persistent homology; persistence statistics; persistence curves; hybrid models

1. Introduction

Early medical image analysis mainly focused on the interpretation and numerical analysis of images. For example, Statistical Parametric Mapping provides medical staff with reference values for images such as MRI and PET images to assist physicians in making treatment decisions [1–3]. Due to the rapid development of AI over the past decade, numerous research teams have developed computer diagnostic systems to assist physicians [4–7]. There are two main tasks that computer-aided medical image analysis tools perform: segmentation and diagnosis. In the medical field, image analysis is usually performed for specific regions of the body, such as tumors, organs, or the brain [8–13]. The analyzed images are then segmented. Notably, various automated image segmentation methods [14,15] have been developed in recent years. These methods still suffer from human biases and fail to identify some differences in real data [16]. However, the latest deep learning methods have shown reliable results in addressing these problems. In this regard, Convolutional Neural Networks (CNN) are the most groundbreaking application platforms. They play a dominant role in the field of image analysis [17,18]. However, the main disadvantage of CNNs is that they require large numbers of of data for training, and



Citation: Hu, C.-S.; Lawson, A.; Chen, J.-S.; Chung, Y.-M.; Clifford, S.; Yang, S.-M. TopoResNet: A Hybrid Deep Learning Architecture and Its Application to Skin Lesion Classification. *Mathematics* **2021**, *9*, 2924. https://doi.org/10.3390/ math9222924

Academic Editor: Qingshan Jiang

Received: 29 September 2021 Accepted: 12 November 2021 Published: 17 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). the acquisition of medical images is difficult and expensive. Furthermore, the training process relies on subjective judgments for obtaining accurate information. In addition, medical images are geometrically and biologically complex. For example, angiogenesis, which commonly accompanies cancer, is mainly affected by protein signalling in the overall microenvironment. Although there are correlations between them, CNN analysis is unable to resolve homology problems [19].

Topological data analysis (TDA) offers a different viewpoint from CNN analysis. TDA has been used for the classification of biomolecular data [19–21]. In topological studies, independent entities, rings, and higher-dimensional faces in a space are characterized through the connectivity of different components in space. Many biological problems exhibit topology-function relationships in biomolecular systems. For example, topological analysis can predict protein–ligand binding affinities for 3D biological protein molecular structures [19]. It can also provide reliable immunohistochemical (IHC) data for diagnosis of pathological slides [22].

The application of deep learning to the field of medical image analysis has made significant progress in recent years. Hybrid methods have been employed in numerous studies to improve the accuracy of deep learning classification [23,24]. For example, Mahbod et al. (2019) pointed out that when a combination of AlexNet, VGG-16 and ResNet-18 is applied for classification, the accuracy can be as high as 90.69%, and if the images are only classified for seborrheic keratosis disease, the accuracy can reach up to 97.55% [23]. This suggests that a combination of multiple algorithms can provide high accuracy in skin lesion classification.

Topological features calculated from images can improve classification accuracy. In [19], the, authors applied topology-based deep learning methods to successfully predict biomolecular properties. The main advantage is that topology allows for effective structural classification, mainly via the application of homology. The deep learning method in [19] was combined with topology to successfully predict protein–ligand binding affinity. It was demonstrated that topology results in accurate classification when the classes in the classification task are distinguished by their structure.

We hypothesized that skin lesions can be identified by their topological structure, so we developed a novel deep learning architecture, called *TopoResNet-101*, that combines topological features computed based on the persistence curve framework (defined in Section 2.3) and persistence statistics (defined in Section 2.2) with features produced by ResNet-101 [25]. We use PC and PS as abbreviations for persistence curves and persistence statistics, respectively. These can be viewed as summaries of the topological features in the images that are invariant under smooth transformations of the image, such as rotations and stretchings. The features generated by ResNet-101 are often local and geometric (e.g., gradients, edges) information. On the other hand, topological ones are global information. Hence, they can be used as additional information for the original neural network model. As shown in Section 3, TopoResNet-101 has advantages in accuracy and stability over other models not using topology. To the best of our knowledge, our work is the first to combine such topological features with a convolutional neural network (CNN) such as ResNet-101 in a classification task. To measure the stability of our classification model's performance, we utilize the *top-n accuracy* evaluation metric of testing results among training epochs. In deep learning, the weights of parameters in an architecture are usually determined by observing the convergence behaviors and (local) maxima of the accuracy curve on a validation dataset. However, this method has a significant drawback in that the chosen weights would strongly depend on validation datasets. In our experiments, we show that TopoResNet-101 has a higher top-*n* accuracy than ResNet-101, and the accuracy curves between validation and testing sets are more compatible than the same for pure ResNet-101. This phenomenon shows that PS and PC may provide more robust features of skin lesions.

The outline of this paper is as follows. In Section 2, we discuss the mathematical background needed to properly define PC and PS. In Section 3, we introduce the TopoResNet-101 and its topological rate α . The main classification results are shown in Section 3 and the conclusion is in Section 5.

2. Mathematical Background

In this section, we introduce our topological features: persistence curves (PCs) and persistence statistics (PS). We provide some of the necessary mathematical background for this. Because these features are based on persistence diagrams and persistent homology, we review those in Section 2.1. The PS and PC features will be presented in Sections 2.2 and 2.3, respectively. Persistence diagrams contain topological information about the image. However, they cannot be used in machine learning algorithms directly. In fact, transforming persistence diagrams into vectors is one of the main research areas in TDA [26–29]. PCs were proven to be useful for classification of texture data sets in [30]. Persistence statistics were used in [31] to classify sleep stages from heart rate signals.

In addition to feature engineering, our previous work [32] proposed an intuitive method for segmenting the lesion part of the image. See the left part of Figure 1 for an example. There are deep learning methods to perform image segmentation, such as [33,34]. It would be interesting to explore those segmentation methods in the context of skin images, but this will be beyond the scope of this paper. The focus of this paper is to design topological features and combine them with ResNet-101.



Figure 1. The architecture of TopoResNet-101. The blue bar is the concatenation of the PS and PC topological data and $\alpha \in [0, 1]$ is the topological rate. The yellow bar is the output vector of ResNet-101 before a final fully-connected layer. The part of ResNet-101 architecture used in this figure is generated by the Deep Network Designer in the Matlab toolbox. Reproduced with permission from [32]; published by IEEE, 2018.

2.1. Persistent Homology

Algebraic topology is a classical subject and has a long history within mathematics. *Persistent homology*, formally introduced in [35], brings the power of algebraic topology to bear on real world data. The field has proven useful in many applications, such as neuroscience [36], medical biology [37], sensor networks [38], and social networks [39]. Here, we give a brief overview of homology and persistent homology for images.

Homology is a tool in topology that allows us to associate an algebraic object (such as a group) to a topological space. In this work, we are concerned with topological spaces that are built as finite unions of *n*-cubes glued together at points, edges, or faces. Such a space is called a cubical complex. For instance, a digital image fits naturally in this framework. A pixel can be expressed as a unit square, i.e., a 2-cube, and a collection of pixels in a digital image forms a cubical complex. Given a cubical complex *X*, we

denote its corresponding k-th homology group by $H_k(X)$. For a more formal introduction and development of cubical homology, we refer the reader to [40]. Informally, homology counts topological features such as connected components (0-dimensional homological features), loops (1-dimensional homological features), voids (2-dimensional homological features), and more. Mathematically, 3- or higher-dimensional homological features exist, but they are difficult to visualize. In practice, 0-dimensional and 1-dimensional homology already provide useful information. The counts of such *k*-dimensional features are the well-known Betti numbers. In binary images, a black pixel is indicated by a value of 0 and a white pixel by a value of 1. We interpret 0- and 1-dimensional homological features in binary images as follows. We count connected clusters of white pixels as 0-dimensional homological features and connected clusters of black pixels (surrounded by white pixels) as 1-dimensional homological features. For example, the Betti numbers of the binary image shown in Figure 2 are $\beta_0 = 4$ because there are four separate clusters of white pixels and $\beta_1 = 2$ because there are two clusters of black pixels enclosed by the white ones. Let *X* be a binary image. We treat X as a cubical set (made up of a finite union of pixels or squares) and denote the *k*-th Betti number of X by $\beta_k(X)$. Note that the connectivity in the 2D cubical complex is four-connectivity.



Figure 2. The Betti numbers of a binary image. By convention, a binary image represents the cubical complex *X* of white pixels in the image. Its Betti numbers are $\beta_0(X) = 4$ and $\beta_1(X) = 2$. Note that if the image is surrounded by a boundary of white pixels, then $\beta_0(X) = 5$ and $\beta_1(X) = 3$. Reproduced with permission from [32]; published by IEEE, 2018.

To generalize the idea of homology, consider a filtration, an increasing sequence of cubical complexes $\{X_i\}$ that satisfies

$$X_0 \subseteq X_1 \subseteq \ldots \subseteq X_n. \tag{1}$$

One could count Betti numbers for each X_i and the inclusion allows one to track the changes of the Betti numbers. More precisely, each inclusion $f_i : X_i \to X_{i+1}$ (sending x in X_i to the same x in X_{i+1}) extends linearly to a homomorphism, also denoted f_i , between the corresponding homology groups (see [41]) so that for each k we have

$$H_k(X_1) \xrightarrow{f_1} H_k(X_2) \xrightarrow{f_2} \dots \xrightarrow{f_{n-1}} H_k(X_n).$$

Furthermore, if $j \ge i$, the inclusion of $X_i \subseteq X_j$ induces a map $f_{i,j}$ on the corresponding homology groups. Functoriality of homology concludes that these maps satisfy the relation $f_{k,j} \circ f_{i,k} = f_{i,j}$ for $i \le j \le k$. We say a homology class α is *born* at *b* if we have $\alpha \in H_k(X_b)$ and $\alpha \notin \inf f_{b-1,b}$. We say that α is born at *b* and *dies* at *d* (with $d \ge b$), if $f_{b,d-1}(\alpha) \notin$ im $f_{b-1,d-1}$, but $f_{b,d}(\alpha) \in \inf f_{b-1,d}$, i.e., if α merges with a previous class. The ranks $\beta_{b,d}^k = \operatorname{rank} \inf f_{b,d}$ for $d \ge b$ form the *k*-th persistent Betti numbers of the filtration. These persistent Betti numbers count the number of classes that were born at or before *b* and are still alive at *d*. Inclusion–exclusion allows us to count exactly the number $\mu_{b,d}^k$ of classes born at *b* and that die at *d* by $\mu_{b,d}^k = \beta_{b,d-1}^k - \beta_{b-1,d-1}^k + \beta_{b-1,d}^k - \beta_{b,d}^k$ [42]. Every homology feature α has a birth time. However, certain classes might not have death times. For such classes, we assign the "death time" as ∞ . This procedure allows one to define a unique multi-set of points, one point (b, d) for each homology class where *b* is the birth time of the class and *d* is its death time. By collecting these pairs accounting for their multiplicity, we obtain a summary of the data called a *persistence diagram* for each dimension k. To summarize, given a filtration $\{X_i\}$, its k-dimensional persistence diagram is denoted by $\mathcal{P}_k(\{X_i\})$. These diagrams are an integral part of persistence statistics and persistence curves as described in Sections 2.2 and 2.3.

To conclude this subsection, we describe how to find a persistence diagram for a digital image. Consider a grayscale image g as a function $g : \mathbb{Z}^2 \to \mathbb{R}$. It is straightforward to show that its sublevel sets form a filtration. More precisely, for any $t_1, t_2 \in \mathbb{R}$

$$g^{-1}(t_1) \subseteq g^{-1}(t_2), \ \forall t_1 \le t_2,$$
(2)

where $g^{-1}(t)$ denotes the sublevel set of g at the value t, i.e., $g^{-1}(t) := \{(x, y) \in \mathbb{Z}^2 : g(x, y) \leq t\}$. Such a filtration is also known as *sublevel set filtration*. In this work, we use the sublevel set filtration of a grayscale image to produce the persistence diagrams, denoted by $\mathcal{P}_k(\{g^{-1}(t_i)\}_{i=0}^N)$ for k = 0, 1. Typically, when considering an 8-bit grayscale image, a natural choice of sublevel set filtration would be $\{g^{-1}(i)\}_{i=0}^{255}$. See Figure 3 for an illustration of a grayscale image (Top), some of its sublevel sets (Middle), and the corresponding persistence diagrams (Bottom). If the image is not a gray-scale image, we consider each color channel (such as RGB) separately. In other words, let $f = (f_R, f_G, f_B)$ be a colored image, and we consider $\mathcal{P}_k(\{f_R^{-1}(t_i)\}_{i=0}^N), \mathcal{P}_k(\{f_G^{-1}(t_i)\}_{i=0}^N), \mathcal{P}_k(\{f_B^{-1}(t_i)\}_{i=0}^N), for <math>k = 0, 1$





Figure 3. (**Top**): An image from the ISIC dataset. (**Middle**): A small sample of the full filtration of the top image. (**Bottom**): The 0 and 1 dimensional persistence diagrams corresponding to the full filtration of the top image. Reproduced with permission from [32]; published by IEEE, 2018.

2.2. Persistence Statistics

Persistence statistics are statistical measurements of the birth and death coordinates of the points in a persistence diagram. Recall that from Section 2.1, persistence diagrams are multi-sets of pairs of points, (b, d), where b and d indicate the birth and death values of a homological feature, respectively. For each 2D image, we compute the 0- and 1-dimensional

persistence diagrams, \mathcal{P}_0 and \mathcal{P}_1 . We consider only those nontrivial pairs, i.e., (b, d) where b < d. Hence our diagrams contain only finitely many points. For a birth–death pair (b, d), the quantity, d - b, represents the life of the corresponding generator. We consider sets of numbers

$$L_{k} = \{d - b | (b, d) \in \mathcal{P}_{k}\}, \ M_{k} = \left\{\frac{(b + d)}{2}, \ (b, d) \in \mathcal{P}_{k}\right\},$$
(3)

for k = 0, 1. L_k is the set of lifespans. In some sense, it measures the robustness of homological features. M_k is the set of the midlives of the features. Thus, M_k describes the locations of points in a persistence diagram. Our PSs are a set of statistical measurements of M_k and L_k . In particular, we use

- 1. means of M_k , and L_k ;
- 2. standard deviations of M_k , and L_k ;
- 3. coefficient of variation of M_k and L_k ;
- 4. skewness of M_k , and L_k ;
- 5. kurtosis of M_k , and L_k ;
- 6. 25-th percentiles of M_k , and L_k ;
- 7. medians of M_k , and L_k ;
- 8. 75-th percentiles of M_k , and L_k ;
- 9. interquartile ranges of M_k , and L_k ;
- 10. persistent entropy of L_k .

where k = 0, 1. Note that *persistent entropy* was introduced in [43] and is defined as

$$-\sum_{(b,d)\in P}\frac{d-b}{\sum(d-b)}\log\left(\frac{d-b}{\sum(d-b)}\right),\tag{4}$$

where *P* is a given persistence diagram. Persistent entropy can be viewed as the diversity of the lifespans. Note also that the PS is a 19-dimensional vector ($9 \times 2 + 1$). As an example, Table 1 shows samples of PS used in the article.

Table 1. Sample M_0 persistence statistics from the X channel of the XYZ color space of images in Figure 4. Reproduced with permission from [32]; published by IEEE, 2018.

Disease	Mean	std	Skewness	Kurtosis	Median	iqr
MEL	2.2533	1.6644	3.6107	3.0519	2.4897	2.0668
NV	2.6123	2.3389	2.0343	2.2425	2.7211	3.4245
BCC	6.8147	3.0709	3.2841	2.1271	10.9705	2.7159
AKIEC	3.3722	3.3452	3.8697	2.7465	4.4496	6.8388
BKL	4.2876	2.7614	3.7254	7.5341	3.5813	4.6003
DF	1.8916	6.4557	1.9783	3.4724	2.3310	2.8247
VASC	2.5901	2.0824	4.6341	2.7230	2.5502	1.8615

2.3. Persistence Curves

Although the term "persistence curve" has traditionally been used to describe the count of persistence pairs (b, d) in a persistence diagram with persistence d - b higher than a threshold t as a function of t [42,44], in this work, we use a generalization that allows flexible vectorization of persistence diagrams [30]. The motivation for this class of curves lies in the Fundamental Lemma of Persistent Homology [42]. Suppose we have a k-dimensional persistence diagram P corresponding to a filtration $X_1 \subseteq ... \subseteq X_n$. Then this lemma states that we can recover the k-th Betti number of any member of the filtration, say X_t , corresponding to a threshold value t by counting the number of points that lie in the upper left quadrant whose corner lies on the diagonal at (t, t). That is, $\beta_k(X_t) = |\{(b, d) \in P \mid b \leq t, d > t\}|$. We recall the formal definition of PCs from [30].





Figure 4. Sample images of the seven types of skin lesions present in the ISIC training datasets. They are melanoma (MEL), melanocytic nevus (NV), basal cell carcinoma (BCC), actinic keratosis (AKIEC), benign keratosis (BKL), dermatofibroma (DF), and vascular lesion (VASC). The number of images for MEL, NV, BCC, AKIEC, BKL, DF, and VASC are 1113, 6705, 514, 327, 1099, 115, and 142, respectively. Reproduced with permission from [32]; published by IEEE, 2018.

Let \mathcal{D} represent the set of all persistence diagrams. Let \mathcal{F} represent the set of all functions $\psi : \mathcal{D} \times \mathbb{R}^3 \to \mathbb{R}$ so that $\psi(D; x, x, t) = 0$ for all $x \in \mathbb{R}$. Let \mathcal{T} represent the set of operators that map multi-sets to the reals, and finally let \mathcal{R} represent the set of functions on \mathbb{R} . We define a map $P : \mathcal{D} \times \mathcal{F} \times \mathcal{T} \to \mathcal{R}$ where

$$P(D, \psi, T)(t) = T(\{\psi(D; b, d, t) \mid b \le t, d > t\}).$$
(5)

The function $P(D, \psi, T)$ is called a *persistence curve* on *D* with respect to ψ and *T*. In [30], it is shown that persistence landscapes [29] are a special case of PCs.

In the present application, all filtrations have exactly 255 spaces. Thus, for each diagram, a persistence curve is a vector in \mathbb{R}^{255} . The two functions that were of greatest use were the functions $\psi(b, d, t) = 1$, giving rise to the Betti curve $\beta(t)$ and $e(b, d, t) = -\frac{d-b}{\Sigma(d-b)} \log \frac{d-b}{\Sigma(d-b)}$ giving rise to a variant of the entropy summary (curve) E(t). The entropy summary and its stability are discussed in [43,45]. In [30], a general stability result for an entire class of PC is given. We calculate the curves for the 0- and 1-dimensional persistence diagrams for each channel in our color space. Finally, we fed these features into machine learning models. The persistence curves we used in our final model are

- 1. $\beta_0(t)$ and $\beta_1(t)$.
- 2. $E_0(t)$ and $E_1(t)$.

The $\beta_0(t)$ and $\beta_1(t)$ curves and the $E_0(t)$ and $E_1(t)$ curves are the $\beta(t)$ and E(t) curves that correspond to the 0 and 1 dimensional diagrams, respectively.

3. TopoResNet

Convolutional neural networks (CNN) have become important tools in deep learning. Some of the most important models have been AlexNet [46], VGG [47], and ResNet [25]. These models have been very successful at image recognition tasks [25,33,34]. In this work, we will base our neural nets on the residual neural network (ResNet); in particular, we will use ResNet-101 [25], which provides an end-to-end architecture for image classification. ResNet optimizes the residuals between the input and the desired convolution features. The desired features can be extracted in a more efficient way than in other CNN models. Therefore, this optimization of residues can be applied to reduce the number of parameters in the network. Because of the benefit of reduction of parameters, the number of layers can be increased. We created the Topological ResNet-101 neural net (or TopoResNet-101).

3.1. Topological Features

In our implementation, these features are described in the RGB and XYZ color spaces with the following dimensions:

- PS-RGB (dimension = $19 \times 3 \times 2 = 114$);
- PS-XYZ (dimension = $19 \times 3 \times 2 = 114$);
- PC-RGB (dimension = $255 \times 3 \times 2 = 1530$);
- PC-XYZ (dimension = $255 \times 2 \times 2 = 1020$).

These features serve as part of the input for TopoResNet-101.

Note that for both PS-RGB, and PS-XYZ, each channel produces two persistence diagrams (0th and 1st level), and each persistence diagrams summarizes to PS as a 19dimensional vector. Therefore, both PS-RGB and PS-XYZ are of dimension 114. PC-RGB contains six PCs in total, which are $\beta_0(t)$ and $\beta_1(t)$ for each channel. We found that the X component of the XYZ color space performed well in our experiments. Hence, PC-XYZ contains four PCs in the X channel: $\beta_0(t)$, $\beta_1(t)$, $E_0(t)$, and $E_1(t)$. Figure 5 illustrates samples of persistence curves.

There are nine models of the TopoResNet-101 type. Each model uses a different combination of topological features. These models and the topological features they use are listed below:

Model 1: none (original ResNet-101); Model 2: PS-RGB; Model 3: PS-XYZ; Model 4: PC-RGB; Model 5: PC-XYZ; Model 6: Reduced PC-RGB (dimension = 512); Model 7: Reduced PC-XYZ (dimension = 512); Model 8: Reduced {PS-RGB, PS-XYZ, PC-RGB, PC-XYZ} (dimension = 512); Model 9: Random noise data (dimension = 512).

Because features with large dimensions (PC-RGB and PC-XYZ) may result in instabilities in the training process, we also design Reduced PC-RGB (Model 6), Reduced PC-XYZ (Model 7), and Reduced ALL (Model 8). In those models, we reduce the dimension of the topological feature vector before concatenation. We replaced the blue layer shown in Figure 1 by a fully connected layer with 512 output nodes, and Figure 6 depicts this reduction step. The input feature vectors were min–max normalized to have all their coordinates in the interval [0, 1].

Note that in Model 8, we combined all features used in Models 2, 3, 4, and 5. Therefore, the total number of features prior to reduction is $2 \cdot 114 + 1530 + 1020 = 2778$. In addition, for the comparison purpose, we also include the Model 9 which substitutes random noise (the uniform distribution on [-1, 1]) for the topological features.



Figure 5. Sample PCs from X channel of XYZ color space of images from different classes. For each class, 30 images are selected at random to illustrate their persistence curves. (**a**–**g**) Betti number curves, $\beta_0(t)$ and $\beta_1(t)$, for each class. When $t \in [0, 255]$, $\beta_0(t)$ is shown; when $t \in [256, 512]$, $\beta_1(t)$ is shown. (**h**) Average of β curves over those 30 images. (**i**–**o**) Entropy curves, $E_0(t)$ and $E_1(t)$, for each class. When $t \in [0, 255]$, $E_0(t)$ is shown; when $t \in [256, 512]$, $E_1(t)$ is shown. (**p**) Average of *E* curves over those 30 images. In our earlier work [32], a support vector machine model was built based on these curves and the classification score was 67.2%.



Topological features

Figure 6. The fully connected sub-network that is used for reducing the topological features. For the reduced models (Models 6–8), we replace the blue layer in Figure 1 with this sub-network. The dimensional of this sub-network is 512. Weights of links in the sub-network are also optimized by the back-propagation algorithm.

3.2. TopoResNet Main Architecture Features

The main architecture of TopoResNet-101 is as follows. The image itself is fed to ResNet with 101 layers. Persistence curves and persistence statistics derived from the image are fed to a series of parallel layers. The outputs of these two branches are then processed by additional layers to give a final output. The architecture of TopoResNet-101 is depicted in Figure 1.

We introduce a parameter $\alpha \in [0, 1]$, called the *topological rate*, that weights the contributions of the convolutional features and the topological ones. We multiply each component in the topological feature and ResNet-101 output feature by α and $1 - \alpha$, respectively. Thus, the input vector before the last fully connected layer, i.e., the pink bar in Figure 1, is

$$v = (1 - \alpha) \cdot \mathbf{v}_{\text{ResNet}-101} \oplus \alpha \cdot \mathbf{v}_{\text{Topology}},$$
(6)

where $\mathbf{v}_{\text{ResNet}-101}$ is the vector of ResNet-101 output features (the yellow bar in Figure 1) and $\mathbf{v}_{\text{Topology}}$ is the vector of topological features (the blue bar in Figure 1), and \oplus is the concatenation operator. Formally, if $\mathbf{v}_{\text{ResNet}-101} = (x_1, \ldots, x_n)$ (n = 2048 in ResNet-101), $\mathbf{v}_{\text{Topology}} = (t_1, \ldots, t_m)$, and $\mathbf{w} = (y_1, \ldots, y_l)$ is the first layer of the fully connected network, then for each $k \in \{1, 2, \ldots, l\}$ we have

$$y_{j} = \beta_{k} + (1 - \alpha) \sum_{i=1}^{n} \omega_{ik} x_{i} + \alpha \sum_{j=n+1}^{m+n} \omega_{jk} t_{j}$$
(7)

where β_k is a bias and $\omega_{ik}, \omega_{jk}, i \in \{1, 2, ..., n\}, j \in \{n + 1, 2, ..., n + m\}$ are the weights of links. In particular, if $t_i = \epsilon_i$'s are random variables that are i.i.d. uniform distributions on [-1, 1], then the value y_i is defined by adding noise on $\mathbf{v}_{\text{ResNet}-101}$ from the random variable $\alpha \sum_{j=n+1}^{n+m} \omega_{jk} \epsilon_j$. Therefore, the Model 9 can be viewed as a modification of ResNet-101 by adding random noise on the input layer of the fully-connected network.

Since α is a parameter in the TopoResNet-101, it will be changed in the learning process. In practice, α was initially set to be $\sigma(0.5) \approx 0.6$, where $\sigma : \mathbb{R} \to (0,1)$ is the sigmoid function

$$\sigma(t) = \frac{1}{1 + e^{-t}}$$

We will always apply the sigmoid function to α to ensure that it is in the interval (0, 1). The topological rate, α , records the importance of topological features. We also allow α as a weight in the network so it can be optimized over the training epochs. The terms α and $1 - \alpha$ can be viewed as a layer of two nodes in the model; hence, it can be optimized by the back-propagation technique.

We summarize our approach as follows. First, we apply the segmentation algorithm based on our previous work [32] to obtain image masks. Second, we apply the mask to the original image. Third, we consider both the RGB and XYZ color space and treat each channel separately. Fourth, we use persistent homology software, specifically, Perseus [48] and CubicalRipser [49], to compute persistence diagrams for each channel. Finally, from each persistence diagram, we calculate persistence curves and persistence statistics as features. This schematic pipeline is shown as a data pre-processing stage in Figure 1.

4. Experiment Results

4.1. Description of the Data Set

In the United States, the five-year survival rate for treated melanoma in the United States is 98% among those with localized disease and 17% among those in whom spread has occurred [50]. The ISIC 2018 challenge [51] tasked competitors to use ISIC's archive of over 13,000 dermatoscopic images collected from a variety of sources [52] to design models with the goal of detecting melanoma. Each image belonged to one of the following seven types: melanoma (MEL), melanocytic nevus (NV), basal cell carcinoma (BCC), actinic keratosis (AKIEC), benign keratosis (BKL), dermatofibroma (DF), and vascular lesion (VASC). Figure 4 demonstrates sample images. Note that the number of available images varies widely per class, as shown in the caption of Figure 4. This imbalance makes classification a challenging task. The International Skin Imaging Collaboration (ISIC, [53]) has put forth a number of imaging challenges to the scientific community [52,54,55]. These challenges have presented unique opportunities for researchers to test novel computer vision ideas to improve the detection of skin cancer with the long-term goal of facilitating early treatment and greatly improving patient outcomes.

Since the beginning of the ISIC skin lesion classification challenge in 2018, numerous studies have adopted CNNs and proposed further improvements to this network. In some studies, transfer learning or entropy-controlled neighborhood component analysis (ECNCA) were used [23], whereas in others, features were extracted from the images in spatial and frequency domains to improve the sensitivity and accuracy of the deep learning methods [56,57]. Recently, to improve the accuracy and sensitivity of image classification, several studies have adopted the concept of hybrid methods by combining different calculation methods [24,58,59]). In view of the advantages of introducing topological homology in biological image analysis, this study developed a PC and PS classification method combined with a CNN algorithm to improve the accuracy and sensitivity of image analysis. The method adopted in this study can be combined with different deep learning or hybrid deep learning methods for optimizing deep learning classification analysis. Therefore, our team applied ResNet-101 and TopoResNet-101 in combination with PC and PS analysis methods based on the persistent homology (called TopoResNet) to compare and analyze skin lesion classification results.

We separated the 10,015 ISIC images into three parts: the training set, validation set, and testing set. To construct a balanced test set, we define a testing set of 350 images by collecting 50 images from each class. We used 70% of the remaining 9965 images for training and the others for validation. We report our scores from the training set.

4.2. Performance of TopoResNet-101

Hyperparameters of a neural network are determined by assessing the performance of a fully trained model on a validation set. If a validation set has a smaller cardinality than the training set or is unbalanced in certain classes of images, then the weight-choosing criterion may be heavily biased. Because our validation set is quite small, we opt to use an evaluation metric known as the *top-n accuracy*, which quantifies the stability of a model. The *top-n accuracy* of a model is defined to be the average value of balanced accuracy on the testing dataset among the training epochs. A classification model or a feature extractor that has a higher top-*n* accuracy has more possibilities to pick models that have high performance among the training epochs. This means the final performance is more stable and reliable.

Table 2 shows that, amongst the nine models, Reduced ALL has the best performance and PS-RGB has the second-best performance. We illustrate the effect of utilizing topological features in Figure 7. There, we plot the accuracies of ResNet-101, PS-RGB, and Reduced ALL on the validation and test data. We see that the performance of Reduced ALL, which uses all the topological data, is more stable than that of PS-RGB, which only uses the PS-RGB topological data. In addition, PS-RGB is more stable than ResNet-101, which uses none of the topological data.



Figure 7. The mean accuracy curves for the Models ResNet-101, PS-RGB, and Reduced ALL over the course of the training process. There are 100 training epochs in total. The mean accuracy on the testing set is balanced. (**a**) Validation dataset; (**b**) testing dataset.

Table 2. The Top 5, 10, 15, 20, 25, and 30 balanced accuracies of all models. The Std column tabulates
the standard deviations of the Top 30 accuracies. Values with bold denote the best in correspond-
ing columns.

Model	Top-5	Тор-10	Top-15	Тор-20	Тор-25	Тор-30	Std * 100
ResNet-101	0.829	0.823	0.820	0.818	0.816	0.815	0.788
PS-RGB	0.833	0.827	0.823	0.820	0.817	0.815	1.023
PS-XYZ	0.827	0.823	0.820	0.818	0.816	0.814	0.811
PC-RGB	0.762	0.759	0.756	0.754	0.752	0.751	0.707
PC-XYZ	0.778	0.774	0.772	0.770	0.769	0.767	0.622
Reduced PS-RGB	0.819	0.817	0.816	0.814	0.813	0.811	0.538
Reduced PS-XYZ	0.817	0.814	0.812	0.811	0.810	0.809	0.455
Reduced ALL	0.848	0.845	0.843	0.842	0.841	0.840	0.459
Noise	0.823	0.821	0.820	0.819	0.818	0.818	0.288

In Figure 8, we show the accuracy curves on the validation/testing datasets of ResNet-101 and TopoResNet-101. Comparing Figure 8a and Figure 8b, we see that the learning process of TopoResNet-101 shows more stability than that of ResNet-101. For example, by observing the accuracy curve of Reduced ALL on the testing dataset (the red curve in Figure 8b), we see that the accuracies are \geq 0.75 after epoch 30. However, the convergence behavior of ResNet-101 seems to be unstable as some epochs have testing accuracy < 0.75 after epoch number \geq 40 (Figure 8a). On the other hand, the curves in Figure 7 show that



the balanced accuracy of TopoResNet-101 on the testing data is more stable and reliable than that of ResNet-101.

Figure 8. Comparisons of accuracy curves on validation set and balanced testing set of (**a**) ResNet-101, (**b**) PS-RGB, and (**c**) Reduced ALL.

Secondly, we observe that the convergence behavior of the blue curve in Figure 8b is closer to the red curve than curves in Figure 8a. We see this in Figure 8a. Although the blue curves appear to be converging between epochs $10 \sim 50$ and $60 \sim 100$, the red curve is changing, especially in epochs $60 \sim 80$. In contrast with the neural model, which operates without topological assistance, the blue and red curves were more consistent and had a similar increasing property.

We list Top-5, 10, 15, 20, 25, and 30 accuracies of ResNet-101, TopoResNet-101 and ResNet-101, with noisy input in Table 2.

4.3. Topological Rates in TopoResNet-101

In TopoResNet-101, α is designed as a new layer for the model. and hence it can be optimized (by backpropagation) as a weight of a neuron of the model. The motivation is to let model automatically decide the importance of features from ResNet and topological ones. Figure 9 shows the behaviors of α in the training process of models with topological assistance (Models 2–9). We observe that topological rate α converges in all models, as shown in Figure 9.



Figure 9. Behaviors of topological rates α in the training process of models with topological assistance. The horizontal axis is the number of training epochs, and the vertical axis is α . (a) PS-RGB (dim = 114) and PS-XYZ (dim = 114). (b) PC-RGB (dim = 1530) and Reduced PC-RGB (dim = 512). (c) PC-XYZ (dim = 1020) and Reduced PC-XYZ (dim = 512). (d) Reduced ALL (dim = 512) and random noise (dim = 512).

As shown in Figure 9, α seems to be influenced by the dimension of the input features. Indeed, by comparing Figure 9b and Figure 9c to Figure 9a, we see that PC-RGB and PC-XYZ have higher feature dimension and higher α rates than PS-RGB and PS-XYZ. In addition, as shown in Figure 9b, Reduced PC-RGB utilizes the full PC-RGB data and has a higher α rate than Reduced ALL which only uses the reduced version of this data. We observe the same effect in the PC-XYZ feature in Figure 9c. There, we see that the model which uses the full PC-XYZ data has a higher α -rate than the model that uses the Reduced PC-XYZ data.

To investigate the effects of dimension reduction, we plot the α curves of Models 4–7 in Figure 9b,c. This figure shows that reduction does help the α rate converge more quickly, even if to a lower value. We also observe that curves of PC-RGB and PC-XYZ (the yellow curve and green curve) in Figure 9 also show more variation in the α rate in earlier epochs. We believe that it shows the reduction can benefit ResNet to learn topological features.

The fact that $\alpha \neq 0$ in Models 2–8 might suggest PS and PC features may be useful for recognizing skin lesions. To validate our hypothesis, we replace topological features with random noise. As we discussed above, α seems to be influenced by the dimensions of features. In comparable settings, we consider Models 6–9 where the dimensions of topological features are all 512. As shown in Figure 9b–d, the converged α for Reduced PC-RGB, Reduced PC-XYZ, and Reduced ALL are 0.220, 0.210, and 0.230, respectively, whereas α for the noise input is 0.191. It would be interesting to further investigate properties of α , such as the correlation between α and accuracy of the model or extending α to a vector.

Moreover, except for using the convergence and accuracy to observe the influence of α rates, other unknown measurements seem necessary for studying the behavior of α .

We also observe that the model with noisy input (Model 9) has the second-best performance in Table 3. The literature provides some explanation for this puzzling phenomenon. In short, noisy input could benefit the robustness of machine learning models. Indeed, certain dropout techniques regularize a neural network by adding noise to its hidden units. In [60], the authors analyze this random dropout and conclude that its can be used to prevent neural networks from overfitting. The paper [61] also shows that adding Gaussian noise boosted the robustness of neural models. As shown in (7), the main difference between [60,61] and Model 9 is the random variables. Adding different types of random noise in hidden layers is now a widely applied method and has been implemented in deep learning frameworks, e.g., Keras and Tensorflow. This technique is believed to be helpful for improving the robustness of deep learning models (e.g., [62], Figure 3).

Table 3. (a) Average accuracies and standard deviations for Epoch *n* to Epoch 100 for (n = 50, 60, 70, 80, 90) on the testing dataset of all models. (b) Average accuracies and standard deviations from Epoch 10 to Epoch *n* for (n = 20, 30, 40, 50) for ResNet-101. Values with bold denote the best scores in corresponding columns.

(a)							
Epochs	\geq 50	\geq 60	\geq 70	\geq 80	\geq 90		
ResNet-101	0.778 ± 0.033	0.786 ± 0.028	0.794 ± 0.021	0.805 ± 0.008	0.811 ± 0.004		
PS-RGB	0.795 ± 0.024	0.795 ± 0.022	0.798 ± 0.019	0.791 ± 0.016	0.787 ± 0.018		
PS-XYZ	0.791 ± 0.022	0.788 ± 0.021	0.787 ± 0.021	0.792 ± 0.015	0.790 ± 0.014		
PC-RGB	0.743 ± 0.011	0.743 ± 0.012	0.743 ± 0.012	0.743 ± 0.013	0.742 ± 0.017		
PC-XYZ	0.763 ± 0.008	0.764 ± 0.008	0.763 ± 0.008	0.764 ± 0.008	0.763 ± 0.006		
Reduced PC-RGB	0.800 ± 0.011	0.799 ± 0.010	0.801 ± 0.009	0.802 ± 0.009	0.802 ± 0.008		
Reduced PC-XYZ	0.803 ± 0.009	0.802 ± 0.010	0.803 ± 0.010	0.803 ± 0.010	0.802 ± 0.010		
Reduced ALL	$\textbf{0.832} \pm \textbf{0.011}$	$\textbf{0.832} \pm \textbf{0.011}$	$\textbf{0.832} \pm \textbf{0.011}$	$\textbf{0.830} \pm \textbf{0.012}$	$\textbf{0.831} \pm \textbf{0.010}$		
Noise	0.814 ± 0.005	0.814 ± 0.005	0.815 ± 0.005	0.815 ± 0.004	0.813 ± 0.003		
(b)							
Epochs	10~20	10~30	10~40	10~50			
ResNet-101	0.808 ± 0.02	0.806~0.02	0.804~0.02	$0.804 {\sim} 0.01$			

Alternatively, because the same training image may frequently appear in training epochs, adding random noise on layers can be viewed as a way to extend the training features. We think it would be an interesting future work to train the model by adding random noise to data, hidden layers, and topological features.

4.4. Online Testing on ISIC 2018 and 2019

We utilized the live leaderboard provided by ISIC 2018 and uploaded our TopoResNet-101 and ResNet-101 results on the system. The balanced accuracies of TopoResNet-101 and ResNet-101 were 0.728 and 0.711, respectively, as measured on the 1512 testing images.

ISIC 2019 provided 25,331 skin lesion images for training across eight different categories and 8238 images for testing (https://challenge2019.isic-archive.com/, accessed 11 November 2021). We also uploaded our classification result on that system and achieved a balanced accuracy of 0.518. The best accuracy on that leaderboard is 0.636 (DAISYLab, Hamburg University of Technology/University Medical Center; Website: https://daisylabs.github.io/, accessed 11 November 2021). That model used HAM10000 (https://www.kaggle.com/kmader/skin-cancer-mnist-ham10000, accessed 11 November 2021) images as an additional dataset for training. It was also an ensemble of 15 neural models. We also compared the result of using ResNet-152 (the experiment is provided by BMIT (Biomedical and Multimedia Information Technology, University of Sydney); website: http://bmit-network.org/, accessed 11 November 2021) and denseNet-201 architectures, which had balanced accuracies of 0.481 and 470.

5. Discussion and Conclusions

In this work, we developed a hybrid deep learning model that uses topological features in conjunction with original image features to improve the performance of deep learning classification. Up to the top-*n* accuracy, the results show that the inclusion of these topological features improves the balanced accuracy of the classification by $\sim 2\%$ over the same model excluding the topological features.

We used a weight (α), i.e., the topological rate, to determine the relative importance of topological features used in the method. In our architecture, α was a trained feature based on the output of ResNet-101 and the topological features we selected. We notice in Figure 9 that in each model, the trained value of α was relatively small (<0.5) but never close to 0. This indicates that, although the raw topological features may not be more important than the features returned by ResNet-101, they do play a role in improving classification performance. In particular, ResNet-101 + PS-RGB and ResNet-101 + all features (reduce) achieved the highest accuracy. In these models, α was 0.152 and 0.230, respectively. In Figure 9, we see that the topological rate converged quickly in the training process, in each case converging at around 30 epochs. We believe that a higher value of α indicates that a higher importance was placed on a topological feature. However, we also observe that although PS-XYZ has higher α rates than PS-RGB (Figure 9a), it does not seem to perform better than PS-RGB (Table 2). Note also that although the feature dimension of PC-XYZ has a lower α rate than PC-RGB in the earlier training epochs, the α rate curve of PC-XYZ (Figure 9c) is more unstable than the curve of PC-RGB (Figure 9b). We speculate that a possible reason for this is that the XYZ representation for images may provide additional information that a CNN model pre-trained on RGB images cannot easily pick up. We plan to investigate this phenomenon in our future work.

We also saw in Figure 8 that the convergence of the balanced accuracy from TopoRes-Net on the Validation and Testing sets occurred quickly after epoch 30. In comparison, ResNet-101 seems to converge, but stutters around epochs 50–80. In addition, Table 2 shows the top-*n* accuracy results for the three models: ResNet-101, TopoResNet-101, and ResNet-101 with noise. Notably, TopoResNet-101 was the most accurate and its standard deviation was the third lowest, and it exhibited strong stability. As a result, we believe TopoResNet-101 to be the better classification model in terms of accuracy and stability.

5.1. More on Performances of Models

As we can see in Figures 7 and 8, the mean accuracy curves on the testing sets are lower than those on the validation sets. For example, although ResNet-101 has 0.829 Top-5 accuracy on the testing dataset, the parameters that achieve this might not be the parameters in the final model. In fact, ResNet-101 achieves its best performance at epoch 98 with accuracy 0.903 on the validation set, while the corresponding model has an accuracy 0.814 on the testing set. On the other hand, if we choose Epoch 27 as a local maximum (accuracy 0.899) of the blue curve in Figure 8a, then the corresponding model will have an accuracy of 0.789.

In Table 3, we consider a better measure of the accuracy of the models, the average accuracy from Epoch *n* to Epoch 100 for n = 50, 60, 70, 80, and 90 on the testing dataset. Table 3a shows that the ResNet models with topological features perform better when the training curve converges earlier. For a fair comparison, in Table 3b, we see that the training curve of ResNet-101 that seems to converge between Epochs 10 and 50. By comparing (a) and (b) in Table 3, we see that Reduced ALL and the model with noise input (Model 9) perform more accurately and more stably than ResNet-101.

We also present the accuracy curve per class on the testing set. The curves in Figure 10 show that TopoResNet-101 may perform more accurately and more stably on Class 2 and Class 3 than on the other classes. On the other hand, we see that both ResNet and

TopoResNet do not perform as well as Class 1, Class 4, and Class 5. The PS-RGB features may help ResNet to achieve a higher accuracy (e.g., Epoch 80 in Figure 10d). A similar phenomenon also occurs on PS-XYZ, PC-RGB, and PC-XYZ in the performances on Class 1 and Class 3. For example, the curves in Figure 11c,d show that PC-RGB and PC-XYZ may contain important topological features useful for classifying BCC lesions (Class 3).







Figure 10. The accuracy curves of ResNet-101, PS-RGB, and Reduced ALL plotted versus training epoch on each of the 7 classes in the testing set. (**a**) Class 1 (melanoma, MEL). (**b**) Class 2 (melanocytic nevus, NV). (**c**) Class 3 (basal cell carcinoma, BCC). (**d**) Class 4 (actinic keratosis, AKIEC). (**e**) Class 5 (benign keratosis, BKL). (**f**) Class 6 (dermatofibroma, DF). (**g**) Class 7 (vascular lesion, VASC).



Figure 11. Examples of models that use topological features that achieve accuracy scores higher than those of ResNet-101 in at least some epochs of training. (a) Class 1 (melanoma, MEL). (b) Class 3 (basal cell carcinoma, BCC). (c) Class 3 (basal cell carcinoma, BCC). (d) Class 3 (basal cell carcinoma, BCC).

5.2. Future Work and Summary

In this method, topological features calculated from the image are used to improve classification accuracy. The proposed method can be applied to different deep learning algorithms. In [19], authors applied topology-based deep learning methods to successfully predict biomolecular properties. The main advantage is that topology allows for effective

structural classification, mainly via the application of homology. Their deep learning method was combined with topology to successfully predict protein–ligand binding affinity. It was demonstrated that topology results in accurate classification when the classes in the classification task are distinguished by their structure. In view of this and with the assumption that skin lesion diagnosis can be described by lesion (topological) structure, our study utilized the homology of skin lesion images to generate features that improve the accuracy of the original deep learning method.

Many studies on skin lesion classification have been published in recent years. For example, Mahbod et al., (2017) applied AlexNet, VGG-16, and the hybrid of these two to achieve an accuracy range of 79.9–89.2%, thereby demonstrating that different methods have their own limitations. However, they showed that the optimized Fusion mode had the highest accuracy (Mahbod et al., 2017). In addition, Mahbod et al. (2019) pointed out that when a combination of AlexNet, VGG-16, and ResNet-18 is applied for classification, the accuracy can be as high as 90.69%, and if the images are only classified for seborrheic keratosis disease, the accuracy can reach up to 97.55% [23]. This suggests that a combination of multiple algorithms can provide high accuracy in skin lesion classification. To the best of our knowledge, this study is the first to provide a combined method involving two concepts. In particular, in theory, topological features have been proved to be advantageous in image analysis. Therefore, it is believed that the use of the above-mentioned methods combined with topological features may lead to more comprehensive skin lesion classification and more satisfactory accuracy. This is conducive to the future classification of medical images using AI.

The appeal of including topological information such as persistence curves (PC) and persistence statistics (PS) lies in the simplicity of calculating these features. The features themselves do not require user defined parameters; thus, one only needs to tune the attached machine learning algorithm. In addition, these features give intuitive shape summaries of the original space. The generalized nature of the persistence curve definition allows for a rich library of usable curves. In this paper, we have chosen to combine the PCs and PSs with the Betti and entropy curves and then fed them into SVM and ResNet-101. The corresponding performances show that using PS and PC can boost performance over models that use only convolutional features. One future direction we will consider is the application of these features to other classification tasks. Furthermore, the phenomenon shown in Figures 10 and 11 shows that using PS and PC features may boost performance in specific classes (e.g., Class 1, 2, and 3). Understanding which PS/PC features best help with classification for images in each class is also an important direction for future research.

In summary, this study sufficiently provides an empirical basis for image classification based on PC/PS applied to various color spaces. ResNet-101 facilitates the automatic extraction of features from raw image input and the influence of the topological features can be evaluated by training the α parameter proposed in this study. This basic architecture, which appears in our proposed model, TopoResNet-101, is an architecture that can be used in different hybrid methods. The techniques presented in this study form a basis for more innovative ideas to incorporate topological information in hybrid deep learning algorithms. For example, we may seek to exchange the dense net at the head of our architecture for some other machine learning algorithm. Or, prior to the topological rate, we may wish to do further preprocessing of the topological features themselves. Our team expects that different hybrid deep learning methods combined with the proposed topological features will provide more promising results in the future.

Author Contributions: Y.-M.C., A.L. and C.S. initiated the project and wrote the paper. Y.-M.C. and A.L. devised the topological features. C.-S.H. proposed the architecture of TopoResNet-101 and the segmentation algorithm, designed the experiments, and wrote the paper. J.-S.C. organized the paper, provided opinions for the statistical results from a biomedical point of view and wrote the paper. S.-M.Y. implemented the TopoResNet-101 model. All authors have read and agreed to the published version of the manuscript.

Funding: Clifford Smyth was supported by a Simons Collaboration grant ID 360486. Jung-Sheng Chen was supported by Ministry of Science and Technology, Taiwan grant ID MOST 110-2314-B-650-011-MY2.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

- 1. Dale, A.M.; Liu, A.K.; Fischl, B.R.; Buckner, R.L.; Belliveau, J.W.; Lewine, J.D.; Halgren, E. Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron* **2000**, *26*, 55–67. [CrossRef]
- Montandon, M.L.; Slosman, D.O.; Zaidi, H. Assessment of the impact of model-based scatter correction on [18F]-FDG 3D brain PET in healthy subjects using statistical parametric mapping. *Neuroimage* 2003, 20, 1848–1856. [CrossRef] [PubMed]
- Huang, Q.; Nie, B.; Ma, C.; Wang, J.; Zhang, T.; Duan, S.; Wu, S.; Liang, S.; Li, P.; Liu, H.; et al. Stereotaxic 18F-FDG PET and MRI templates with three-dimensional digital atlas for statistical parametric mapping analysis of tree shrew brain. *J. Neurosci. Methods* 2018, 293, 105–116. [CrossRef]
- Wernick, M.N.; Yang, Y.; Brankov, J.G.; Yourganov, G.; Strother, S.C. Machine learning in medical imaging. *IEEE Signal Process.* Mag. 2010, 27, 25–38. [CrossRef]
- Ithapu, V.K.; Singh, V.; Okonkwo, O.C.; Chappell, R.J.; Dowling, N.M.; Johnson, S.C. Imaging-based enrichment criteria using deep learning algorithms for efficient clinical trials in mild cognitive impairment. *Alzheimers Dement.* 2015, *11*, 1489–1499. [CrossRef] [PubMed]
- 6. Giger, M.L. Machine learning in medical imaging. J. Am. Coll. Radiol. 2018, 15, 512–520. [CrossRef]
- Karimi, D.; Dou, H.; Warfield, S.K.; Gholipour, A. Deep learning with noisy labels: exploring techniques and remedies in medical image analysis. *Med. Image Anal.* 2020, 65, 101759. [CrossRef] [PubMed]
- 8. Hegde, R.B.; Prasad, K.; Hebbar, H.; Singh, B.M.K. Comparison of traditional image processing and deep learning approaches for classification of white blood cells in peripheral blood smear images. *Biocybern. Biomed. Eng.* **2019**, *39*, 382–392. [CrossRef]
- Li, L.; Zhao, X.; Lu, W.; Tan, S. Deep learning for variational multimodality tumor segmentation in PET/CT. *Neurocomputing* 2020, 392, 277–295. [CrossRef]
- 10. Coccia, M. Deep learning technology for improving cancer care in society: New directions in cancer imaging driven by artificial intelligence. *Technol. Soc.* **2020**, *60*, 101198. [CrossRef]
- 11. Kadampur, M.A.; Al Riyaee, S. Skin cancer detection: Applying a deep learning based model driven architecture in the cloud for classifying dermal cell images. *Inform. Med. Unlock.* 2020, *18*, 100282. [CrossRef]
- 12. Saha, S.; Pagnozzi, A.; Bourgeat, P.; George, J.M.; Bradford, D.; Colditz, P.B.; Boyd, R.N.; Rose, S.E.; Fripp, J.; Pannek, K. Predicting motor outcome in preterm infants from very early brain diffusion MRI using a deep learning convolutional neural network (CNN) model. *NeuroImage* **2020**, *215*, 116807. [CrossRef] [PubMed]
- Thakur, S.; Doshi, J.; Pati, S.; Rathore, S.; Sako, C.; Bilello, M.; Ha, S.M.; Shukla, G.; Flanders, A.; Kotrotsou, A.; et al. Brain extraction on MRI scans in presence of diffuse glioma: Multi-institutional performance evaluation of deep learning methods and robust modality-agnostic training. *NeuroImage* 2020, 220, 117081. [CrossRef]
- 14. Bukenya, F.; Nerissa, C.; Serres, S.; Pardon, M.C.; Bai, L. An automated method for segmentation and quantification of blood vessels in histology images. *Microvasc. Res.* **2020**, *128*, 103928. [CrossRef] [PubMed]
- 15. Kucybała, I.; Tabor, Z.; Ciuk, S.; Chrzan, R.; Urbanik, A.; Wojciechowski, W. A fast graph-based algorithm for automated segmentation of subcutaneous and visceral adipose tissue in 3D abdominal computed tomography images. *Biocybern. Biomed. Eng.* **2020**, *40*, 729–739. [CrossRef]
- 16. Kumar, H.; DeSouza, S.V.; Petrov, M.S. Automated pancreas segmentation from computed tomography and magnetic resonance images: A systematic review. *Comput. Methods Programs Biomed.* **2019**, *178*, 319–328. [CrossRef]
- 17. Feng-Ping, A.; Zhi-Wen, L. Medical image segmentation algorithm based on feedback mechanism convolutional neural network. *Biomed. Signal Process. Control* 2019, 53, 101589. [CrossRef]
- 18. Khatami, A.; Nazari, A.; Khosravi, A.; Lim, C.P.; Nahavandi, S. A weight perturbation-based regularisation technique for convolutional neural networks and the application in medical imaging. *Expert Syst. Appl.* **2020**, *149*, 113196. [CrossRef]
- 19. Cang, Z.; Wei, G.W. TopologyNet: Topology based deep convolutional and multi-task neural networks for biomolecular property predictions. *PLoS Comput. Biol.* **2017**, *13*, e1005690. [CrossRef]
- Kasson, P.M.; Zomorodian, A.; Park, S.; Singhal, N.; Guibas, L.J.; Pande, V.S. Persistent voids: a new structural metric for membrane fusion. *Bioinformatics* 2007, 23, 1753–1759. [CrossRef]

- 21. Gameiro, M.; Hiraoka, Y.; Izumi, S.; Kramar, M.; Mischaikow, K.; Nanda, V. A topological measurement of protein compressibility. *Jpn. J. Ind. Appl. Math.* 2015, 32, 1–17. [CrossRef]
- 22. Takiyama, A.; Teramoto, T.; Suzuki, H.; Yamashiro, K.; Tanaka, S. Persistent homology index as a robust quantitative measure of immunohistochemical scoring. *Sci. Rep.* 2017, 7, 1–9.
- Mahbod, A.; Ecker, R.; Ellinger, I. Skin Lesion Classification Using Hybrid Deep Neural Networks. In Proceedings of the ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 1229–1233.
- 24. Yu, X.; Qiu, H.; Xiong, S. A novel hybrid deep neural network to predict pre-impact fall for older people based on wearable inertial sensors. *Front. Bioeng. Biotechnol.* 2020, *8*, 63. [CrossRef] [PubMed]
- 25. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015.
- Chung, Y.; Hull, M.; Lawson, A. Smooth Summaries of Persistence Diagrams and Texture Classification. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 14–19 June 2020; IEEE Computer Society: Los Alamitos, CA, USA, 2020; pp. 3667–3675. [CrossRef]
- 27. Adams, H.; Emerson, T.; Kirby, M.; Neville, R.; Peterson, C.; Shipman, P.; Chepushtanova, S.; Hanson, E.; Motta, F.; Ziegelmeier, L. Persistence images: A stable vector representation of persistent homology. *J. Mach. Learn. Res.* **2017**, *18*, 218–252.
- 28. Kusano, G.; Hiraoka, Y.; Fukumizu, K. Persistence weighted Gaussian kernel for topological data analysis. In Proceedings of the International Conference on Machine Learning, New York, NY, USA, 19–24 June 2016; pp. 2004–2013.
- 29. Bubenik, P. Statistical topological data analysis using persistence landscapes. J. Mach. Learn. Res. 2015, 16, 77–102.
- 30. Chung, Y.M.; Lawson, A. Persistence Curves: A canonical framework for summarizing persistence diagrams. *arXiv* 2019, arXiv:1904.07768.
- 31. Chung, Y.M.; Hu, C.S.; Lo, Y.L.; Wu, H.T. A Persistent Homology Approach to Heart Rate Variability Analysis With an Application to Sleep-Wake Classification. *Front. Physiol.* **2021**, *12*, 202. [CrossRef]
- 32. Chung, Y.M.; Hu, C.S.; Lawson, A.; Smyth, C. Topological approaches to skin disease image analysis. In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018.
- Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015.
- 34. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015.
- Edelsbrunner, H.; Letscher, D.; Zomorodian, A. Topological persistence and simplification. In Proceedings of the 41st Annual Symposium on Foundations of Computer Science, Redondo Beach, CA, USA, 12–14 November 2000; pp. 454–463.
- Bendich, P.; Marron, J.S.; Miller, E.; Pieloch, A.; Skwerer, S. Persistent homology analysis of brain artery trees. *Ann. Appl. Stat.* 2016, 10, 198. [CrossRef]
- Li, L.; Cheng, W.Y.; Glicksberg, B.S.; Gottesman, O.; Tamler, R.; Chen, R.; Bottinger, E.P.; Dudley, J.T. Identification of type 2 diabetes subgroups through topological analysis of patient similarity. *Sci. Transl. Med.* 2015, *7*, 311ra174. [CrossRef]
- 38. De Silva, V.; Ghrist, R. Coverage in sensor networks via persistent homology. Algebraic Geom. Topol. 2007, 7, 339–358. [CrossRef]
- 39. Carstens, C.J.; Horadam, K.J. Persistent homology of collaboration networks. Math. Probl. Eng. 2013, 2013, 815035. [CrossRef]
- 40. Kaczynski, T.; Mischaikow, K.; Mrozek, M. *Computational Homology*; Applied Mathematical Sciences; Springer: New York, NY, USA, 2004.
- 41. David S.; Dummit, R.M.F. Abstract Algebra; Wiley Publication: Hoboken, NJ, USA, 2003.
- 42. Edelsbrunner, H.; Harer, J. *Computational T Opology: An Introduction;* Miscellaneous Books, American Mathematical Society: Providence, RI, USA, 2010.
- 43. Atienza, N.; Gonzalez-Diaz, R.; Soriano-Trigueros, M. A new entropy based summary function for topological data analysis. *Electron. Notes Discret. Math.* **2018**, *68*, 113–118. [CrossRef]
- 44. Tierny, J. Topological Data Analysis for Scientific Visualization; Springer: Berlin/Heidelberg, Germany, 2017; Volume 3.
- 45. Atienza, N.; González-Díaz, R.; Soriano-Trigueros, M. On the stability of persistent entropy and new summary functions for TDA. *arXiv* **2018**, arxiv:1803.08304. Available online: https://arxiv.org/abs/1803.08304 (accessed on 30 September 2021).
- 46. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [CrossRef]
- 47. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv 2014, arXiv:1409.1556.
- 48. Nanda, V. Perseus, the Persistent Homology Software. 2013. Available online: http://www.sas.upenn.edu/~vnanda/perseus (accessed on 30 September 2021).
- 49. Sudo, T.; Ahara, K. CubicalRipser: Calculator of Persistence Pair for 2 Dimensional Pixel Data. 2018. Available online: https://github.com/CubicalRipser/CubicalRipser_2dim (accessed on 30 September 2021).
- 50. NCI. SEER Stat Fact Sheets: Melanoma of the Skin. Available online: https://seer.cancer.gov/statfacts/html/melan.html (accessed on 19 August 2018).
- 51. ISIC2018. Available online: https://challenge2018.isic-archive.com/ (accessed on 19 August 2018).
- 52. Tschandl, P.; Rosendahl, C.; Kittler, H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci. Data* 2018, *5*, 161–180. [CrossRef] [PubMed]

- 53. ISIC. Available online: https://www.isic-archive.com/ (accessed on 19 August 2018).
- 54. ISIC Challenges 2016–2018. Available online: https://challenge.isic-archive.com/ (accessed on 19 August 2018).
- 55. Codella, N.C.F.; Gutman, D.; Celebi, M.E.; Helba, B.; Marchetti, M.A.; Dusza, S.W.; Kalloo, A.; Liopyris, K.; Mishra, N.K.; Kittler, H.; Halpern, A. Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), Hosted by the International Skin Imaging Collaboration (ISIC). *arXiv* 2017, arXiv:1710.05006. Available online: https://arxiv.org/abs/1710.05006 (accessed on 30 September 2021).
- Hu, J.; Li, Y.; Zhao, X.; Xie, W. A spatial constraint and deep learning based hyperspectral image super-resolution method. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 23–28 July 2017; pp. 5129–5132.
- 57. Li, J.; You, S.; Robles-Kelly, A. A frequency domain neural network for fast image super-resolution. In Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 8–13 July 2018; pp. 1–8.
- Venkatraman, S.; Alazab, M.; Vinayakumar, R. A hybrid deep learning image-based analysis for effective malware detection. J. Inf. Secur. Appl. 2019, 47, 377–389. [CrossRef]
- 59. Salur, M.U.; Aydin, I. A Novel Hybrid Deep Learning Model for Sentiment Classification. *IEEE Access* 2020, *8*, 58080–58093. [CrossRef]
- 60. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
- Zheng, S.; Song, Y.; Leung, T.; Goodfellow, I.J. Improving the Robustness of Deep Neural Networks via Stability Training. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 4480–4488.
- 62. Liu, X.; Cheng, M.; Zhang, H.; Hsieh, C.J. Towards Robust Neural Networks via Random Self-ensemble. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.