nature portfolio

Peer Review File

Manifold Topological Deep Learning for Biomedical Data

Corresponding Author: Professor Guo-Wei Wei

Parts of this Peer Review File have been redacted as indicated to remove third-party material.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Attachments originally included by the reviewers as part of their assessment can be found at the end of this file.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The manuscript introduces Manifold Topological Deep Learning (MTDL) and demonstrates competitive performance on MedMNIST v2. The idea of extending topological deep learning from combinatorial data to differentiable manifolds via Hodge decomposition is interesting. However, the paper suffers from issues in methodological clarity, fairness of comparisons, rigor of evaluation, and substantiation of interpretability claims. The paper requires major revisions before it can be considered further.

Strengths:

1. The manuscript successfully proposes a novel framework, MTDL, extending TDL from combinatorial data (like graphs and point clouds) to differentiable manifold data, such as images.

2. The use of Hodge decomposition to factor an image's vector field into three orthogonal components (curl-free, divergence-free, and harmonic parts) is mathematically elegant. This method captures distinct geometric and topological features, offering the potential for enhanced interpretability.

3. MTDL achieves competitive or superior results across a diverse range of tasks categorized by dimensionality (2D vs. 3D), data modality, data scale, and task type. Moreover, the model is highly efficient, with a lightweight architecture containing only 0.56M and 0.75M parameters for 2D and 3D tasks, respectively.

Major Concerns:

1. Figure 1a contradicts the definitions of boundary conditions. According to the definition, the normal boundary condition (NBC) should generate a larger support set than the tangential boundary condition (TBC), but Figure 1a shows the opposite. While supplementary Figure 2 is consistent with the definition.

2. The authors state that topological deep learning models “offer greater interpretability.” However, the paper does not provide any analysis or demonstration of interpretability. The new representation of images as vector fields is introduced, but its meaning and diagnostic value are not justified.

3. In the ablation study, the authors replace the decomposed images (curl-free, divergence-free, and harmonic parts) with the original images. It remains unclear which of the three modalities contributes most to performance. The study does not clarify whether each component alone is weaker than the original image but the combination yields a gain.

4. According to Figure 4c, curl-free and divergence-free components appear very similar to each other and to the raw vector field. For images, especially medical images where most pixels are far from the boundary, NBC and TBC differ only at edge regions. Consequently, curl-free and divergence-free maps are nearly indistinguishable. This raises doubts about the necessity of including all three channels. Maybe simply augmenting the original image with the harmonic part would suffice?

An ablation explicitly comparing the role of these three channels is needed.

5. MTDL concatenates three components, thereby increasing the number of input channels and the effective model capacity. The ImgCNN ablation lacks sufficient description and parameter counts are not reported. It is essential to ensure the total number of parameters is equal. Otherwise, the performance gain cannot be attributed specifically to Hodge decomposition.

6. The BIG Laplacian is introduced without its full name or a clear derivation. The rationale for simplifying the Laplacian into this form, and its numerical and geometric consequences, is insufficiently discussed. It remains unclear whether the BIG Laplacian preserves orthogonality, which is a central property of Hodge decomposition. The conditions under which it is equivalent or a good approximation to the full Laplacian should be explicitly analyzed.

7. The paper emphasizes that MTDL is a “lightweight” model based on small parameter counts. However, the Hodge decomposition requires additional computation. The extra computational cost is not reported.

8. MTDL shows competitive results but is not consistently superior. The statistical significance of improvements is not demonstrated. In addition, the comparison set relies on some outdated baselines. More recent state-of-the-art methods from the past two years should be included for the comparison to have empirical relevance.

9. The so-called “Transformer-encoder-induced convolution layers” replace multi-head attention and the feed-forward network with group convolution and 1×1 convolution. This is essentially a CNN residual block, not a Transformer. The terminology is misleading. Either rename the block to reflect its actual nature, or implement true attention for comparison.

10. Operator construction lacks transparency. Equations (8)–(11) define several operators, but their explicit matrix construction is not shown. Providing a concrete example using a small toy data would improve reproducibility and make the work accessible to readers unfamiliar with TDL.

(Remarks on code availability)
I have not run the code myself but going through the code it seems to be a reasonable barebones version. If polished, it can be a useable, useful resources, though not at this point.

Reviewer #2

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
Co-Review of NCOMMS-25-20703-T; Nature Communications: by Zakeri A & Frangi A F
Title: Manifold Topological Deep Learning for Biomedical Data

Summary:
The manuscript introduces Manifold Topological Deep Learning (MTDL), which represents images as discrete manifolds, constructs a vector field on that manifold, applies a discrete Hodge decomposition into curl-free / divergence-free / harmonic components, concatenates those channels, and feeds the result into a CNN. The model is evaluated on MedMNIST v2 (11 2D + 6 3D datasets) and reported to outperform prior methods.

The idea of converting images to vector fields and apply a discrete Hodge decomposition before feeding the components to a CNN model for classification is promising. However, the paper needs further clarification on methodological details, additional ablation studies, and rigorous evaluation. Detailed comments below outline specific improvements needed before publication.

Our view is that the manuscript needs a Major Revision:

Methodology:
1. Explain how the harmonic component h is computed.
2. Provide Pseudo-code for the proposed methodological pipeline.
3. Clarify how the manifold M and its boundary ∂M are defined for each dataset. In practical settings, computing normal and tangential boundary conditions requires a segmentation mask or at least a binary indicator of the region of interest. In Section 4.3.1, the segmentation step is described only as thresholding but the choice of threshold (fixed, adaptive, or dataset-specific), preprocessing (e.g., normalization or filtering), and handling of ambiguous or noisy background regions can significantly influence the resulting manifold boundary ∂M, which in turn affects the computed boundary conditions and the subsequent Hodge decomposition. I recommend the authors (1) explicitly describe the segmentation procedure per dataset, (2) discuss its influence on the topological representation, and (3) provide an ablation or sensitivity analysis to

assess robustness to segmentation variations in the results.
4. The paper would strongly benefit from an explicit pseudocode summarizing the full implementation workflow, including how boundary conditions, discrete operators, and vector-field decompositions are computed. Although the Supplementary Material gives substantial mathematical detail, a procedural description (with solver specifications and preprocessing steps) is essential for reproducibility.

Results:
5. Vector-field construction is under-specified in the main text. The paper mentions several methods and refers to the Supplementary Information, but the main text does not make clear which construction is used for each experiment, what hyperparameters were used, and how sensitive results are to that choice. Since the vector field is the core representation, an ablation is necessary: different vector-field constructions (flow-based, gradient-based, others) and their effect on final performance.
6. Discretization errors and orthogonality checks:
Provide results showing the discrete decomposition is numerically orthogonal.

Report: [see formula in attachment]

Provide these for representative images or in a table for a dataset with mean ± std.
7. The paper concatenates components as extra channels. Provide an ablation showing which components contribute most to performance (curl-free / divergence-free / harmonic).
8. Comparisons / fairness of baselines:
9. The paper reports substantial improvements across multiple MedMNIST tasks; however, it is unclear whether the baseline models were trained and tuned under identical experimental conditions (e.g., data augmentation, optimizer settings, and hyperparameter search strategies). Please clarify whether the baseline results were all adopted from the original publications or reproduced by the authors. If any were re-implemented, details on the hyperparameter parity and training cost should be provided. Without this, it's hard to know whether gains derive from topology or from training differences.
10. Statistical analysis: While the reported results on MedMNIST v2 are promising, the paper lacks statistical analysis. Since many datasets are small and stochastic training introduces non-negligible variance, it is important to report mean ± standard deviation across multiple runs and perform statistical significance testing. A paired t-test or Wilcoxon test within each dataset, and a Friedman test with Nemenyi post-hoc across datasets, would clarify whether the observed improvements are statistically significant rather than due to random variation. Reporting confidence intervals and effect sizes (e.g., Cohen's d or Cliff's δ) would further strengthen the empirical evidence.
11. In Figure 3, the range of the AUC or ACC values on the x-axis should be consistent across all plots to facilitate clearer and quicker visual comparison between results.
12. Provide preprocessing time per sample (vector field + decomposition), GPU memory/CPU usage, and training time per epoch for representative datasets (2D and 3D).

Discussions:
13. Invariance/equivariance: discuss which geometric transforms of the input (translation, rotation, scaling) preserve each decomposed channel and whether the CNN architecture exploits or breaks these invariances.
14. Physical/semantic interpretation: Can we link the decomposed components to expected image structure per modality (e.g., curl-free ~ edge/gradient information; divergence-free ~ texture swirl)? so readers can interpret classifier reliance on components.
15. How the method might generalize to real-world biomedical images, e.g., high-resolution CT/MRI, anisotropic voxels, or dynamic sequences? MedMNIST is small and standardized; clinical data might pose challenges.
16. A discussion (or ideally an experiment) comparing MTDL to surface-based or mesh-based features could be helpful. This might help readers understand when grid-based manifold representation is preferred vs direct mesh representation or potential ways these approaches could complement each other in biomedical imaging applications.
17. The paper could benefit from explicit discussion of scenarios where MTDL might fail.

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
I thank the authors for the detailed response and the extensive additional experiments. Overall, I'm satisfied with the response. I only have a few remaining concerns.

Regarding Comment 2 on interpretability. I noticed that another reviewer raised a related concern in Comment 15. While I accept the abstract explanations, I was hoping to see some concrete evidence in the specific application. For example, the authors mention that "the directions of maximal intensity change typically align with anatomical edges", "in mammography, malignant lesions often show sharper and more irregular boundaries compared with benign lesions", and "swirling patterns and heterogeneous internal textures arise due to necrosis and fibrosis". I do not expect all of these claims to be

demonstrated, but even one or two illustrative examples would significantly improve the clarity and persuasiveness of the interpretability claims.

Finally, after reading the rebuttal to another reviewer, I agree that the lack of rotation and scale equivariance appears to be an inherent limitation of the proposed method. It would be helpful if this limitation could be properly addressed, or at least stated more explicitly in the discussion.

Moreover, a few other things to certainly correct for:

- Introduction exemplifies that the 'DNA chains' as differentiable manifolds. I think this is incorrect and must be toned down - differentiable manifolds can only approximate these at best.
- Is the original image fully recoverable by the decomposition, e.g. by gradient reconstruction etc.? Which information do we lose particularly? For which data is this not acceptable? Given that only some medical data is used, such a discussion would be helpful.

(Remarks on code availability)
I have skimmed through the code but it requires a deeper review to ensure.

Reviewer #2

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
Thank you to the authors for the thoughtful and thorough review. I am pleased that all of my previous concerns have been fully addressed.

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Response to Reviewers

Answers to Reviewer 1' Comments

The manuscript introduces Manifold Topological Deep Learning (MTDL) and demonstrates competitive performance on MedMNIST v2. The idea of extending topological deep learning from combinatorial data to differentiable manifolds via Hodge decomposition is interesting. However, the paper suffers from issues in methodological clarity, fairness of comparisons, rigor of evaluation, and substantiation of interpretability claims. The paper requires major revisions before it can be considered further.

Strengths:

1. The manuscript successfully proposes a novel framework, MTDL, extending TDL from combinatorial data (like graphs and point clouds) to differentiable manifold data, such as images.

2. The use of Hodge decomposition to factor an image's vector field into three orthogonal components (curl-free, divergence-free, and harmonic parts) is mathematically elegant. This method captures distinct geometric and topological features, offering the potential for enhanced interpretability.

3. MTDL achieves competitive or superior results across a diverse range of tasks categorized by dimensionality (2D vs. 3D), data modality, data scale, and task type. Moreover, the model is highly efficient, with a lightweight architecture containing only 0.56M and 0.75M parameters for 2D and 3D tasks, respectively.

Answer: We thank the positive comments and insights of the reviewer.

Major Concerns:

Comment *: 1) Figure 1a contradicts the definitions of boundary conditions. According to the definition, the normal boundary condition (NBC) should generate a larger support set than the tangential boundary condition (TBC), but Figure 1a shows the opposite. While supplementary Figure 2 is consistent with the definition.*

Answer*: We thank the reviewer for the careful review. You are correct that the left part should correspond to the tangential boundary condition, while the right part should correspond to the normal boundary condition. We have revised the figure accordingly.*

Comment *: 2) The authors state that topological deep learning models "offer greater interpretability." However, the paper does not provide any analysis or demonstration of inter-*

pretability. The new representation of images as vector fields is introduced, but its meaning and diagnostic value are not justified.

***Answer**: We thank the reviewer for the insightful comments. The mechanistic interpretability of our model arises from the explicit decomposition of the image information into three mathematically meaningful components, rather than relying on the original images. In our model, each image is represented as a vector field, which is then decomposed into a curl-free component, divergence-free component, and a harmonic component. These three components have clear geometric and physical meanings. Specifically, a vector field is curl-free means it does not have the local rotational behavior. A divergence-free vector field has neither sources nor sinks. Source regions correspond to positive divergence, where local intensity or structural patterns spread outward, while sink regions correspond to negative divergence, where patterns converge inward. The harmonic component is both curl-free and divergence-free, representing the smoothest global information associated with topological traits.*

For the diagnostic value in medical image analysis, the curl-free component can be used to capture the lesion boundaries and margins, as the directions of the strongest intensity transition usually correspond to tissue boundaries. For example, in mammography, malignant lesions often show sharper and more irregular boundaries compared with the benign lesions. Therefore, the curl-free component can be used to distinguish the malignant from benign cases. The divergence-free component captures the vorticity and local rotating behavior in the vector field. Consequently, it can be used to capture the irregular textures, such as the swirling patterns and the heterogeneous internal textures due to necrosis and fibrosis. Especially, this will be useful for malignant tumors since the swirling textures are common in malignant tumors such as breast cancer, brain tumor, and hepatocellular carcinoma. The harmonic component reflects the global topological properties (cycles) of the image and provides robust descriptors insensitive to noises and small perturbations. For a more comprehensive model interpretability about how the decomposed components relate to visual patterns such as edges, textures, and global structures, additional interpretability analysis would be required. We regard this as an important direction for future work.

We have added these discussions to section 6.4 of the revised supplementary information.

***Comment** : 3) In the ablation study, the authors replace the decomposed images (curl-free, divergence-free, and harmonic parts) with the original images. It remains unclear which of the three modalities contributes most to performance. The study does not clarify whether each component alone is weaker than the original image but the combination yields a gain.*

***Answer**: We thank the reviewer for the constructive comments. We have added an ablation study in which we evaluate the model performance by using different combinations of the components of the decomposition, the results have been included in section 6.2 of the revised supplementary information.*

For the reviewer's convenience, we also summarize these results in Table S1. In this analysis, we consider five 2D datasets of varying scales to provide a comprehensive evaluation:

RetinaMNIST (1,600 samples), PneumoniaMNIST (5856 samples), DermaMNIST (10,015 samples), OrganAMNIST (58,830 samples), and PathMNIST (107,180 samples). We have

Table S1: Performance (ACC & AUC) Comparison between the original images and decomposition images. Curl: the curl-free component, Div: the divergence-free component, Har: the harmonic component. Img: the original image.

	Retina	Pneumonia	Derma	OrganA	Path
Curl	0.628	0.910	0.820	0.953	0.914
	0.867	0.986	0.957	0.998	0.995
Div	0.610	0.885	0.725	0.887	0.700
	0.854	0.967	0.890	0.990	0.938
Har	0.485	0.793	0.633	0.887	0.847
	0.685	0.921	0.788	0.989	0.976
Cur+Div	0.635	0.910	0.827	0.956	0.914
	0.868	0.982	0.963	0.998	0.995
Cur+Har	0.635	0.911	0.826	0.955	0.918
	0.867	0.984	0.962	0.998	0.995
Div+Har	0.615	0.829	0.731	0.906	0.784
	0.843	0.960	0.894	0.993	0.963
Cur+Div+Har	0.655	0.910	0.836	0.956	0.920
	0.874	0.986	0.962	0.998	0.996
Img	0.608	0.885	0.808	0.955	0.902
	0.838	0.978	0.957	0.998	0.987
Img+Curl	0.625	0.886	0.821	0.955	0.905
	0.848	0.977	0.958	0.998	0.993
Img+Div	0.630	0.872	0.810	0.956	0.902
	0.834	0.972	0.959	0.998	0.993
Img+Har	0.630	0.864	0.810	0.959	0.914
	0.835	0.970	0.958	0.998	0.995

the following observations:

1. *The curl-free component contributes most to the model performance compared with the divergence-free and harmonic parts.*

2. *Using only the curl-free component can achieve performance comparable to or better than using the original images.*

3. *The combination of all three components consistently gives the best overall performance for all tested datasets, which is reasonable since each component captures distinct and complementary information, and their combination provides a more complete representation of the image features.*

4. *Augmenting the original images with any single component of the decomposition (curl-free, divergence-free, or harmonic) improves performance compared with using only the original images on several datasets, such as RetinaMNIST, DermaMNIST, and PathMNIST. These findings demonstrate that each component captures specific features that cannot be directly extracted from the original images.*

***Comment** : 4) According to Figure 4c, curl-free and divergence-free components appear very similar to each other and to the raw vector field. For images, especially medical images where most pixels are far from the boundary, NBC and TBC differ only at edge regions. Consequently, curl-free and divergence-free maps are nearly indistinguishable. This raises doubts about the necessity of including all three channels. Maybe simply augmenting the original image with the harmonic part would suffice? An ablation explicitly comparing the role of these three channels is needed.*

***Answer**: We thank the reviewer for the constructive comments. In response to Comment 3), we have included an ablation study that evaluates the model's performance using different combinations of the components in the decomposition. As shown in Table S1, the curl-free component contributes the most to the model performance, and the combination of all three components yields the best overall performance across all evaluated datasets.*

***Comment** : 5) MTDL concatenates three components, thereby increasing the number of input channels and the effective model capacity. The ImgCNN ablation lacks sufficient description and parameter counts are not reported. It is essential to ensure the total number of parameters is equal. Otherwise, the performance gain cannot be attributed specifically to Hodge decomposition.*

***Answer**: We thank the reviewer for pointing out this. It is true that the MTDL model receives three input channels for each original channel, as the Hodge decomposition automatically divides an image into three components. Consequently, for an image with n channels, MTDL uses $3n$ channels while ImgCNN uses n channels. However, we want to emphasize that this difference affects only the initial convolution layer. As shown in Figure 1e of the main text, both MTDL and ImgCNN employ an initialization convolution layer with 72 output channels regardless of the number of input channels. Consequently, after this first layer, the rest of the network architecture is identical for both models, and all hyperparameters and training settings are exactly the same. Therefore, the downstream model capacity remains matched.*

Moreover, Table S1 shows that using only the curl-free component already yields consistently better performance than using the original images on RetinaMNIST, PneumoniaMNIST, DermaMNIST, and PathMNIST datasets. Note that the curl-free component has the same channel number with the original image. This observation further demonstrates that the performance improvement arises from the informative features introduced by the Hodge decomposition rather than from increased input channel.

To have a direct comparison, we conducted an additional ablation study in which the original images were augmented by either zero-padding or by adding blurred low- and high-

frequency components such that the total number of input channels matched that of the MTDL representation. Specifically, the low-frequency and high-frequency components were obtained using Gaussian blurring and subsequently combined with the original images. These augmented images were then fed into the same CNN architecture used for MTDL. The results are shown in Table S2. It can be seen that the original images actually achieve comparable, and in most cases better performance than the zero-padded augmented images. Although augmenting the inputs with low- and high- frequency components yields better performance than simply adding zero-valued channels, the improvement remains substantially below that achieved by the full MTDL representation. This confirms that simply increasing the number of input channels does not improve performance and that the gains achieved by MTDL is from the meaningful decomposition provided by the Hodge decomposition rather than from the increased channel count.

Table S2: Performance (ACC & AUC) Comparison between the original images and the augmented images. Img: the original image. Img+zero: augmenting the original images with zero-padded channels. Img+low+high: augmenting the original images with both low- and high-frequency components.

	Retina	Pneumonia	Derma	OrganA	Path
Img	0.608	0.885	0.808	0.955	0.902
	0.838	0.978	0.957	0.998	0.987
Img+zero	0.603	0.880	0.801	0.958	0.905
	0.823	0.979	0.955	0.998	0.978
Img+low+high	0.612	0.888	0.817	0.958	0.912
	0.837	0.977	0.957	0.998	0.986
MTDL	0.655	0.910	0.836	0.956	0.920
	0.874	0.986	0.962	0.998	0.996

Comment *: 6) The BIG Laplacian is introduced without its full name or a clear derivation. The rationale for simplifying the Laplacian into this form, and its numerical and geometric consequences, is insufficiently discussed. It remains unclear whether the BIG Laplacian preserves orthogonality, which is a central property of Hodge decomposition. The conditions under which it is equivalent or a good approximation to the full Laplacian should be explicitly analyzed.*

Answer*: We thank the reviewer for the insightful comments. We have added the full name of the Boundary-Induced Graph (BIG) Laplacian and cited the corresponding reference accordingly. The BIG Laplacian can be derived by replacing the discrete Hodge star with the identity matrix. In the discrete setting, the Hodge star is a diagonal matrix whose entries correspond to the ratio between the volumes of dual cells and primal cells. As a result, only boundary cells yield non-unit diagonal entries, whereas interior cells have value 1. As the grid resolution increases, the number of interior cells becomes dramatically larger than that of boundary cells. Therefore, the identity matrix is an efficient approximation to the Hodge star, and thus the BIG Laplacian serves as an efficient approximation to the Hodge Laplacian.*

In our implementation, we compute the Hodge decomposition using a direct approach. Specifically, for the decomposition

$$\omega = d\alpha + \delta\beta + h$$

we firstly compute the potentials α and β, and then obtain the harmonic component by

$$h = \omega - d\alpha - \delta\beta$$

Actually, this direct approach does not guarantee strict orthogonality of the three components at the discrete level, and orthogonality is recovered only in the continuous limit.

Despite the lack of exact mathematical orthogonality in the discrete case, the three components exhibit strong numerical orthogonality in practice. To quantify this, we computed the normalized pairwise inner products between the curl-free, divergence-free, and harmonic components across five representative datasets: RetinaMNIST (1,600 samples), PneumoniaMNIST (5,856 samples), DermaMNIST (10,015 samples), OrganAMNIST (58,830 samples), and PathMNIST (107,180 samples). The normalized inner product for two components α, β is defined as

$$(\alpha, \beta) = \frac{| < \alpha, \beta > |}{||\alpha||_2 ||\beta||_2}.$$

We report the mean and standard deviation of these values across all samples for each dataset, which is shown in Table S3. The results indicate that the three components exhibit strong numerical orthogonality across all evaluated datasets.

Exact orthogonality is not easily achievable in discrete Hodge decompositions because the discrete differential operators generally do not satisfy all properties of the continuous theory. However, practical applications usually focus on the numerical approximation of orthogonality. Since our numerical errors of orthogonality are near zero, the three components already provide independent and complementary features. For machine learning applications, this level of numerical orthogonality is entirely sufficient, as it ensures that the extracted features are non-redundant and do not interfere with each other.

We have added these discussions to section 1.3 of the revised supplementary information.

Table S3: Numerical analysis of the orthogonality among the three components in the Hodge decomposition across the RetinaMNIST, PneumoniaMNIST, DermaMNIST, OrganAMNIST, and PathMNIST datasets. The mean and standard deviation values of the normalized inner product of all samples within each dataset are reported.

Method	Retina		Pneumonia		Derma		OrganA		Path	
	mean	std	mean	std	mean	std	mean	std	mean	std
$(d\alpha, \delta\beta)$	7e-20	4e-17	2e-19	2e-17	4e-20	3e-17	8e-20	3e-17	7e-20	2e-17
$(d\alpha, h)$	2e-13	9e-14	3e-14	2e-14	6e-14	1e-14	3e-14	3e-14	2e-14	1e-14
$(\delta\beta, h)$	2e-17	5e-17	2e-18	8e-17	7e-18	2e-16	5e-18	8e-17	2e-17	6e-17

Comment *: 7) The paper emphasizes that MTDL is a "lightweight" model based on small parameter counts. However, the Hodge decomposition requires additional computation. The extra computational cost is not reported.*

Answer*: We thank the reviewer for pointing out this. In our experiments, all 2d images are* $n \times 224 \times 224$*, with* $n = 3$ *for color images and* $n = 1$ *for gray images. All 3d images are grayscale with a size* $64 \times 64 \times 64$*. The Hodge decomposition is computationally efficient: 50 grayscale 2d images can be completed within 2 minutes, 50 color 2d images can be completed within 4 minutes, and 50 3d images can be completed within 30 minutes. We have added the running time for the Hodge decomposition in section 7 of the revised supplementary information.*

Comment *: 8) MTDL shows competitive results but is not consistently superior. The statistical significance of improvements is not demonstrated. In addition, the comparison set relies on some outdated baselines. More recent state-of-the-art methods from the past two years should be included for the comparison to have empirical relevance.*

Answer*: We thank the reviewer for the constructive comments.*

For statistical significance analysis of improvement, the baseline methods report only single-run results. Consequently, we conducted the Friedman test followed by the Nemenyi post-hoc test. This procedure is standard in multi-dataset benchmark evaluations, particularly in settings where only single-run baseline results are available. Specifically, we used the scipy.stats.friedmanchisquare package to compute the p-values on 11 MedMNIST2D datasets and 6 MedMNIST3D datasets, respectively. For the 2D evaluation, we included only those baseline models that reported results on all 11 MedMNIST2D datasets. Similarly, for the 3D evaluation, we restricted the comparison to baselines that provided results on all 6 MedMNIST3D datasets. The resulting p-values are 1.38e-14 (2D) and 5.32e-10 (3D), indicating significant performance differences among the compared models. Therefore, we provide Critical Difference (CD) diagrams for both the 2D and 3D cases. The CD value is the minimum rank difference required for two models to be considered significantly different under the Nemenyi test. As shown in Figure S1, our MTDL model achieves the best overall performance with the lowest average rank among all compared methods. Under the computed CD value, MTDL is significantly better than most compared methods for both 2D and 3D cases. This analysis has been added to section 6.1 of the revised supplementary information.

For comparison with recent state-of-the-art methods, our manuscript was originally submitted early this year, and at that time we included all publicly available models evaluated on this database. To provide an updated and comprehensive comparison, we conducted a thorough literature search for works published in 2025 that also utilized this database. We summarize these newly identified results in Table S4 to present an up-to-date benchmark comparison. It can be seen that our model achieves the best overall performance among all these methods. Notably, many of the recently published models are evaluated on only a subset rather than the full set of datasets while our model is evaluated on all the datasets. These new results have been added into section 9 of the revised supplementary information.

[1] *Halder, A., Dalal, A., Gharami, S., Wozniak, M., Ijaz, M. F., & Singh, P. K.*

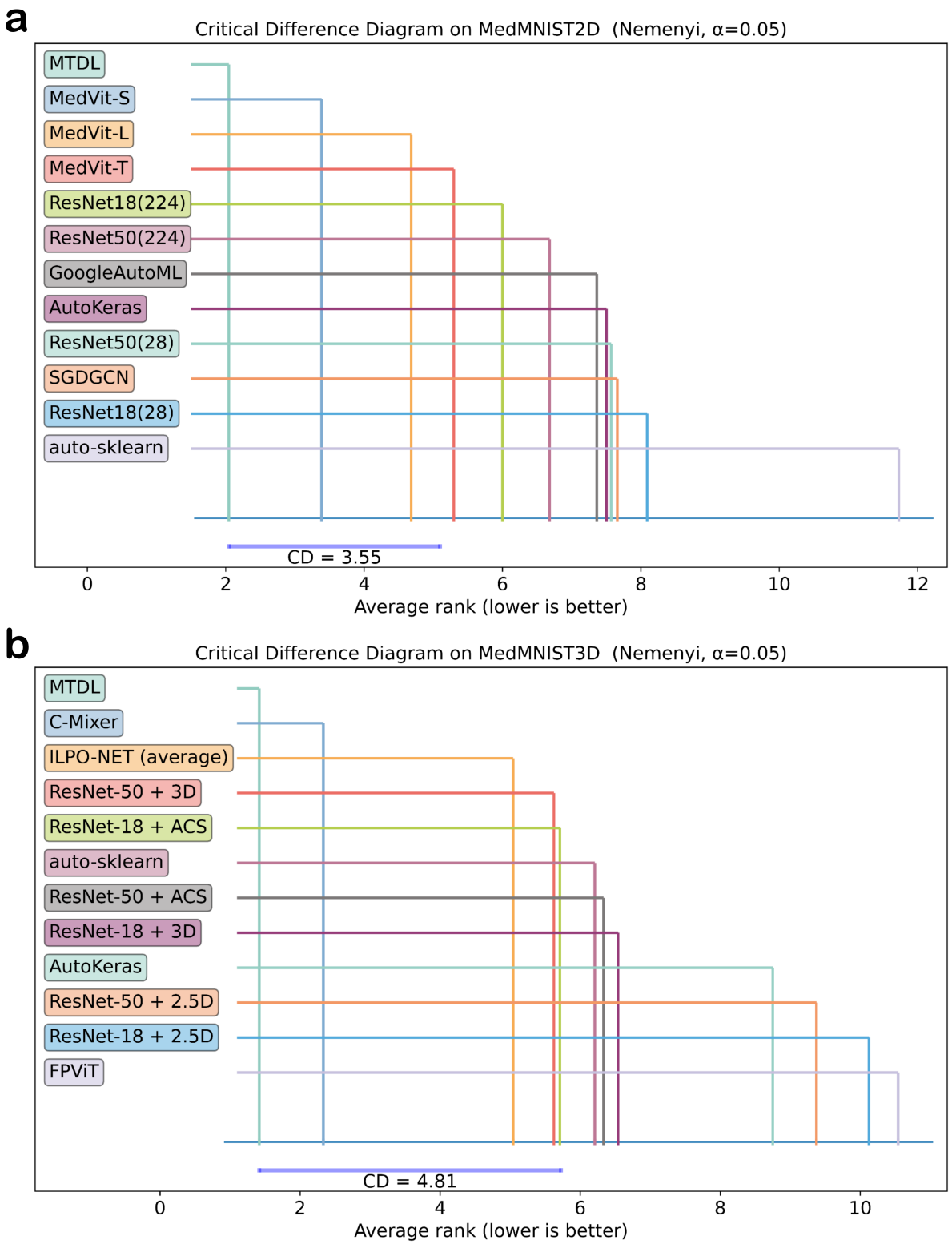

Figure S1: The Critical Difference (CD) diagrams illustrating the performance of our model across 11 MedMNIST2D datasets and 6 MedMNIST3D datasets.

Table S4: Performance comparison between our model and recent models on the MedMNIST2D dataset. The benchmark results are taken from the original papers. The best results are in bold. "-" indicates that the corresponding metric was not reported.

Methods	Path		Chest		Derma		OCT	
	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC
FRDEM [1]	-	-	-	-	-	0.783	-	-
AIGC [2]	-	-	-	0.83	-	-		
MedKAFormer [3]	0.985	0.882	-	-	0.919	0.748	0.974	0.810
LPT [4]	-	**0.941**	-	-	-	0.787	-	0.832
SGDGCN [5]	0.987	0.907	**0.928**	0.854	0.919	0.771	0.933	0.783
PSNAS-Net-ave [6]	-	-	-	-	-	0.772	-	-
SRE-CONV [7]	-	0.822	-	**0.948**	-	0.757	-	0.768
IMET [8]	-	-	-	-	-	-	0.960	0.803
NQNN [9]	-	0.752	-	-	-	0.827	-	0.591
MTDL	**0.996**	0.920	0.793	**0.948**	**0.962**	**0.836**	**0.989**	**0.888**

Methods	Pneumonia		Retina		Blood		Tissue	
	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC
MedKAFormer [3]	0.961	0.904	0.729	0.535	0.998	0.965	-	-
LPT [4]	-	-	-	0.575	-	0.967	-	-
SGDGCN [5]	0.963	**0.918**	0.782	0.550	0.994	0.940	0.801	0.613
PSNAS-Net-ave [6]	-	0.899	-	-	-	0.956	-	-
SRE-CONV [7]	-	0.906	-	0.523	-	0.960	-	0.712
IMET [8]	0.895	0.902	-	-	-	-	-	-
NQNN [9]	-	-	-	-	-	0.737	-	-
MTDL	**0.986**	0.910	**0.874**	**0.655**	**0.999**	**0.988**	**0.945**	**0.721**

Methods	OrganA		OrganC		OrganS	
	AUC	ACC	AUC	ACC	AUC	ACC
AIGC [2]	-	0.890	-	-	-	-
MedKAFormer [3]	0.995	0.922	0.991	0.899	**0.986**	0.788
LPT [4]	-	-	-	0.906	-	0.796
SGDGCN [5]	0.977	0.850	0.989	0.880	0.969	0.784
PSNAS-Net-ave [6]	-	-	-	0.913	-	-
SRE-CONV [7]	0.770	0.769	-	-	-	-
NQNN [9]	-	0.814	-	0.799	-	-
MTDL	**0.998**	**0.956**	**0.996**	**0.928**	0.978	**0.809**

(2025). A fuzzy rank-based deep ensemble methodology for multi-class skin cancer classification. Scientific Reports, 15(1), 6268.

[2] *Wu, B., Huang, J., & Duan, Q. (2025). Real-time intelligent healthcare enabled by federated digital twins with aoi optimization. IEEE Network.*

[3] *Wang, G., Zhu, Q., Song, C., Wei, B., & Li, S. (2025). MedKAFormer: When Kolmogorov-Arnold Theorem Meets Vision Transformer for Medical Image Representation. IEEE Journal of Biomedical and Health Informatics.*

[4] *Bai, Y., Bai, L., Yang, X., & Liang, J. (2025). Label-Semantic-Based Prompt Tuning for Vision Transformer Adaptation in Medical Image Analysis. IEEE Transactions on Circuits and Systems for Video Technology.*

[5] *Singh, A., Eising, C., Denny, P., & van de Ven, P. (2025). Improving GNNs for Image Classification: Addressing Homophily Challenges. IEEE Open Journal of the Computer Society.*

[6] *Zechen, Z., Xuelei, H., Fengjun, Z., & Xiaowei, H. (2025). PSNAS-Net: Hybrid gradient-physical optimizationfor efficient neural architecture search in customized medical imaging analysis. Expert Systems with Applications, 288, 128155.*

[7] *Du, Y., Zhang, J., Zeevi, T., Dvornek, N. C., & Onofrey, J. A. (2025, April). Sre-conv: Symmetric rotation equivariant convolution for biomedical image classification. In 2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI) (pp. 1-5). IEEE.*

[8] *Singh, R., & Guggilam, S. (2025). Iterative Misclassification Error Training (IMET): An Optimized Neural Network Training Technique for Image Classification. IEEE Access.*

[9] *Rahman, M., & Zhuang, J. (2025, September). NQNN: Noise-Aware Quantum Neural Networks for Medical Image Classification. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 433-442). Cham: Springer Nature Switzerland.*

***Comment** : 9) The so-called "Transformer-encoder-induced convolution layers" replace multi-head attention and the feed-forward network with group convolution and 1×1 convolution. This is essentially a CNN residual block, not a Transformer. The terminology is misleading. Either rename the block to reflect its actual nature, or implement true attention for comparison.*

***Answer**: We thank the reviewer for pointing this out. We have renamed the block as a simple-CNN layer in the revised manuscript.*

***Comment** : 10) Operator construction lacks transparency. Equations (8)–(11) define several operators, but their explicit matrix construction is not shown. Providing a concrete example using a small toy data would improve reproducibility and make the work accessible to readers unfamiliar with TDL.*

***Answer**: We thank the reviewer for pointing out this. In section 1.4 of the revised supplementary information, we have added pseudocode for computing the discrete differential operators and performing the Hodge decomposition. We have also included a simple example in section 1.5 as follows:*

Consider the manifold in Figure S2, it is a 2D disk in the Cartesian grid. The resulting

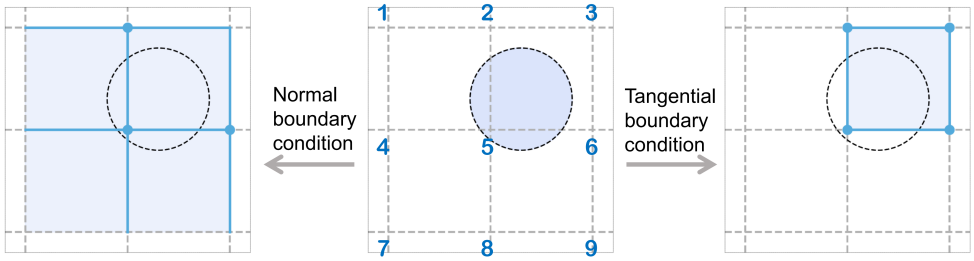

Figure S2: A 2d disk and its discrete manifold representation under normal and tangential boundary conditions.

discrete manifolds under the normal and tangential boundary conditions are displayed on the left and right sides, respectively. Specifically, the discrete manifold under the normal boundary condition consists of 3 vertices, 8 edges, and 4 faces, whereas the discrete manifold under the tangential boundary condition consists of 4 vertices, 4 edges, and 1 face.

We assign an ordering to the primal cells as follows:

- *vertices: 1, 2, 3, 4, 5, 6, 7, 8, 9*
- *edges: 12, 23, 45, 56, 78, 89, 14, 47, 25, 58, 36, 69*
- *faces: 1245, 2356, 4578, 5689*

We use the construction of the 0-th Laplacian matrix under the tangential boundary condition $L_{0,t}$ as an example to illustrate the computation. For this computation, we need to first compute the discrete differential D_0 on the entire grid. Then compute the projection matrix $P_{1,t}$ and $P_{0,t}$ for the tangential boundary condition. Using these projection matrices, the discrete differential under the tangential boundary condition $D_{0,t}$ can be derived by $D_{0,t} = P_{1,t}D_0P_{0,t}^T$. After this, we need to compute the discrete Hodge star $S'_{1,t}$ on the entire grid. Then the discrete Hodge star on the manifold can be derived by $S_{1,t} = P_{1,t}S'_{1,t}P_{1,t}^T$. Having these, the Laplacian can be computed by $L_{0,t} = D_{0,t}^T S_{1,t} D_{0,t}$. We give the details as follows:

1. *Let ∂_1 be the boundary matrix from edges to vertices on the entire grid, then $D_0 = \partial_1^T$*

$$\partial_1 = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad D_0 = \partial_1^T$$

2. *The projection matrices are obtained by selecting the rows of the identity matrix corresponding to the cells included in the discrete manifold. Thus, we have*

$$P_{0,t} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}, \quad P_{1,t} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

3. *The discrete differential $D_{0,t}$ can be computed as follows*

$$D_{0,t} = P_{1,t} D_0 P_{0,t}^T = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}$$

4. *The discrete Hodge star under tangential boundary condition on the entire grid $S'_{1,t}$ is a diagonal matrix in which only the entries associated with edges 23, 56, 25, and 36 are nonzero and all remaining diagonal entries are zero. Each nonzero diagonal entry represents the ratio between the volume of the dual cell and that of the corresponding primal cell. Under the tangential boundary condition, we follow the strategy described in paper "Hodge Decomposition of Vector Fields in Cartesian Grids", where the primal cell volumes are kept unchanged while the dual cell volumes are adjusted. That is, these nonzero entries are given by the volume of the intersection between each dual cell and the manifold. Assuming each grid edge has unit length, the intersection of the disk and the dual edge of 23 is a line segment of length 0.24, so is 36. The intersection of the disk and the dual edge of 56 is a line segment of length 0.64, so is 25. Hence, we have*

$$S'_{1,t} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.24 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.64 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.64 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.24 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{1}$$

5. *The discrete Hodge star under the tangential boundary condition on the manifold is as*

follows:

$$S_{1,t} = P_{1,t} S'_{1,t} P^T_{1,t} = \begin{pmatrix} 0.24 & 0 & 0 & 0 \\ 0 & 0.64 & 0 & 0 \\ 0 & 0 & 0.64 & 0 \\ 0 & 0 & 0 & 0.24 \end{pmatrix} \tag{2}$$

6. The Hodge Laplacian can be computed as

$$L_{0,t} = D^T_{0,t} S_{1,t} D_{0,t} = \begin{pmatrix} 0.24 & -0.24 & 0 & 0 \\ -0.24 & 0.48 & 0 & 0 \\ 0 & 0 & 0.64 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{3}$$

The eigenvalues of $L_{0,t}$ are 0, 0.09, 0.63, and 0.64. The single zero eigenvalue actually corresponds to the Betti number $\beta_0 = 0$ of the manifold. If we replace the Hodge star matrix with the identity matrix, we get the BIG Laplacian $L^B_{0,t}$. The BIG Laplaican $L^B_{0,t}$ also has exactly one zero eigenvalue since it preserves the topology of the manifold.

Answers to Reviewer 2 and Reviewer 3' Comments

Comment *: 1) Summary: The manuscript introduces Manifold Topological Deep Learning (MTDL), which represents images as discrete manifolds, constructs a vector field on that manifold, applies a discrete Hodge decomposition into curl-free / divergence-free / harmonic components, concatenates those channels, and feeds the result into a CNN. The model is evaluated on MedMNIST v2 (11 2D + 6 3D datasets) and reported to outperform prior methods.*

The idea of converting images to vector fields and apply a discrete Hodge decomposition before feeding the components to a CNN model for classification is promising. However, the paper needs further clarification on methodological details, additional ablation studies, and rigorous evaluation. Detailed comments below outline specific improvements needed before publication.

Our view is that the manuscript needs a Major Revision:

Answer*: We appreciate the summary from the reviewers of our work and thank the reviewers for the positive comments.*

Methodology:

Comment *: 2) Explain how the harmonic component h is computed.*

Answer*: We thank the reviewers for pointing out this. For the decomposition*

$$\omega = d\alpha_n + \delta\beta_t + h,$$

where $\omega \in \Omega^1, \alpha_n \in \Omega_n^0, \beta_t \in \Omega_t^2, h \in ker d \cap ker \delta$, we first determine the potentials α_n and β_t, and then compute h as

$$h = \omega - d\alpha_n - \delta\beta_t.$$

Note that α_n and β_t are not unique, as $\alpha_n + d\eta$ and $\beta_t + \delta\gamma$ with any $\eta \in \Omega_n^0$ and $\gamma \in \Omega_t^2$ can serve as valid potentials for the first two terms. We solve this issue by imposing the gauge conditions. Specifically, we restrict

$$\alpha_n \in ker\delta \cap \Omega_n^0, \quad \beta_t \in ker d \cap \Omega_t^2$$

Under these conditions, we have

$$\delta\omega = \delta d\alpha = (\delta d + d\delta)\alpha = \Delta\alpha, \quad d\omega = d\delta\beta_t = (d\delta + \delta d)\beta_t = \Delta\beta_t \tag{4}$$

Thus, α_n and β_t can be uniquely determined from the linear system (4) by resolving the rank deficiency of Δ.

***Comment** : 3) Provide Pseudo-code for the proposed methodological pipeline.*

***Answer**: We thank the reviewers for the constructive comments. In the discretization, three key things need to be considered: the discrete manifold, the discrete differential, and the discrete Hodge star. Once these three components are properly defined, all remaining discrete operators, such as the Laplacian and the Hodge decomposition, can be assembled or implemented following the same definitions of their continuous counterparts.*

We provide Algorithm 1 for constructing the discrete manifold under normal and tangential boundary conditions, Algorithm 2 for computing the projection matrices associated with each boundary condition, Algorithm 3 for building the k-th boundary matrix ∂_k on the entire Cartesian grid, and Algorithm 4 for computing the discrete Hodge star matrix $S'_{k,n}$ and $S'_{k,t}$. Using these four algorithms, we obtain the k-th boundary matrix ∂_k, the discrete Hodge star $S'_{k,n}, S'_{k,t}$, and the projection matrices $P_{k,n}$ and $P_{k,t}$ corresponding to the normal and tangential boundary conditions, respectively.

Based on these matrices, we can derive the k-th discrete differential operators $D_{k,n}$ and $D_{k,t}$, the discrete codifferentials $\delta_{k,n}$ and $\delta_{k,t}$, and the discrete Hodge Laplacians $L_{k,n}$ and $L_{k,t}$ for the two types of boundary conditions. Finally, using these discrete operators, we can compute the Hodge decomposition, which is presented in Algorithm 5.

We have added these algorithms to section 1.4 of the revised supplementary information.

***Comment** : 4) Clarify how the manifold M and its boundary ∂M are defined for each dataset. In practical settings, computing normal and tangential boundary conditions requires a segmentation mask or at least a binary indicator of the region of interest. In Section 4.3.1, the segmentation step is described only as thresholding but the choice of threshold (fixed, adaptive, or dataset-specific), preprocessing (e.g., normalization or filtering), and handling of ambiguous or noisy background regions can significantly influence the resulting manifold boundary ∂M, which in turn affects the computed boundary conditions and the subsequent Hodge decomposition. I recommend the authors (1) explicitly describe the segmentation procedure per dataset, (2) discuss its influence on the topological representation, and (3) provide*

Algorithm 1: Computing Discrete Manifold

Input: A manifold M on the Cartesian grid I.

Output: M_n: Discrete manifold under the normal boundary condition; M_t: discrete manifold under the tangential boundary condition.

Set $M_n = \emptyset$, $M_t = \emptyset$;

for *cell* σ *in the grid* I **do**

  if *one vertex of* σ *is in* M **then**

    add σ into M_n

  Let $\star\sigma$ be the dual cell of σ;

  if *one vertex of* $\star\sigma$ *is in* M **then**

    add σ into M_t

return (M_n, M_t)

Algorithm 2: Computing Projection Matrix

Input: A discrete manifold M_b with normal or tangential boundary condition on the Cartesian grid I.

Output: $P_{k,b}$: the k-th projection matrix

Let n be the number of k-cells of I;

Set $P_{k,b}$ be the $n \times n$ identity matrix;

for k*-cell* σ_k *in the grid* I **do**

  if σ_k *is not in* M_b **then**

    Remove the row corresponding to σ_k in $P_{k,b}$

return $P_{k,b}$

an ablation or sensitivity analysis to assess robustness to segmentation variations in the results.

***Answer**: We thank the reviewers for the insightful suggestions.*

For the segmentation strategy to determine the manifold region from medical images, we carefully examined all 17 datasets. As shown in Figure S3, the majority of the datasets, with the exceptions of PathMNIST, DermaMNIST, and BloodMNIST, contain images with a black background. Note that all datasets share a consistent pixel range from 0 to 255, where the black background corresponds to a pixel value of 0. Based on this observation, we adopt a unified and straightforward segmentation rule: pixels with intensity strictly greater than cutoff of 0 are treated as belonging to the manifold. This ensures consistent preprocessing across all datasets.

For the impact of segmentation parameter to the manifold representation, a very small cutoff can inadvertently include noise, while a large cutoff may discard meaningful information. Using a cutoff of 0 can guarantee that no potentially informative signal is removed. In addition, our vector-field representation is obtained using the gradient operator. Therefore, regions containing low-amplitude noise or approximately constant nonzero background

Algorithm 3: Computing Boundary Matrix

Input: A Cartesian grid I.

Output: ∂_k: the k-th boundary matrix from k-cells to $(k-1)$-cells

Let n_k be the number of k-cells of I;

Set ∂_k be the $n_{k-1} \times n_k$ zero matrix;

for $i = 1$ *to* n_k **do**

  Let σ be the i-th k-cell of I;

  for $j = 1$ *to* n_{k-1} **do**

    Let τ be the j-th $(k-1)$-cell of I;

    if τ *is a face of* σ **then**

      Compute the incidence sign $s(\tau, \sigma) \in \{-1, +1\}$ according to the orientation;

      Set $\partial_k(j, i) = s(\tau, \sigma)$;

return ∂_k

values produce near-zero gradients, contributing negligible vector-field information. As a result, selecting 0 as the segmentation threshold is both empirically effective and theoretically reasonable.

For the impact of segmentation parameter to model performance, we add an ablation study by varying the cutoff values across 0, 5, 10, 15, 20, 25, 30. Experiments were performed on four representative datasets: RetinaMNIST, PneumoniaMNIST, DermaMNIST, and OrganSMNIST. The results are summarized in Table S5. It can be seen that model performance remains stable across this range of thresholds. This indicates that the proposed method is robust to the choice of segmentation cutoff, and that the default setting of 0 is reliable.

We have added these discussions to section 6.3 of the revised supplementary information.

***Comment** : 5) The paper would strongly benefit from an explicit pseudocode summarizing the full implementation workflow, including how boundary conditions, discrete operators, and vector-field decompositions are computed. Although the Supplementary Material gives substantial mathematical detail, a procedural description (with solver specifications and preprocessing steps) is essential for reproducibility.*

***Answer**: We thank the reviewers for the constructive suggestions. We have added detailed pseudocode to section 1.4 of the revised supplementary information for computing the discrete manifold with normal and tangential boundary conditions, constructing the discrete differential operators, and performing the Hodge decomposition. For convenience, we also provide the pseudocode here as Algorithm 1, Algorithm 2, Algorithm 3, Algorithm 4, and Algorithm 5.*

Results:

***Comment** : 6) Vector-field construction is under-specified in the main text. The paper mentions several methods and refers to the Supplementary Information, but the main*

Algorithm 4: Computing Discrete Hodge Star

Input: A manifold M on a Cartesian grid I.

Output: $S'_{k,n}$: the k-th discrete Hodge star matrix under the normal boundary condition on the grid I; $S'_{k,t}$: the k-th discrete Hodge star matrix under the tangential boundary condition on the grid I.

Let n_k be the number of k-cells of I;
Set $S'_{k,n}$ be the $n_k \times n_k$ identity matrix;
Set $S'_{k,t}$ be the $n_k \times n_k$ identity matrix;
for $i = 1$ *to* n_k **do**
    Let σ be the i-th k-cell of I;
    Let $\star\sigma$ be the dual cell of σ;
    Set $S'_{k,n}(i,i) = \frac{volume(\star\sigma)}{volume(\sigma \cap M)}$;
    Set $S'_{k,t}(i,i) = \frac{volume(\star\sigma) \cap M}{volume(\sigma)}$;
return $(S'_{k,n}, S'_{k,t})$

text does not make clear which construction is used for each experiment, what hyperparameters were used, and how sensitive results are to that choice. Since the vector field is the core representation, an ablation is necessary: different vector-field constructions (flow-based, gradient-based, others) and their effect on final performance.

***Answer**: We thank the reviewers for the insightful comments. The equation (15) in the "Decomposed Image Generation" section of the main text is the formula we used to construct the vector field in our MTDL model for all 17 datasets. This formulation corresponds to a special case of the gradient-based method presented in the supplementary information with parameters $s = t = 1$.*

To further assess the impact of different vector field construction strategies, we have added an ablation study comparing three distinct methods: the gradient-based method, the flow-based method, and the RGB-based method, evaluated on the RetinaMNIST, PneumoniaMNIST, DermaMNIST, and OrganSMNIST datasets. For the gradient-based method, we examine three parameter settings: $(s,t) = (1,1)$, $(1,2)$, and $(2,1)$. For the RGB-based method, we construct vector fields using channel pairs (R,G), (R,B), and (G,B). The results of all methods are shown in Table S6. As shown in the table, the gradient-based method with $(s,t) = (1,1)$ achieves the best performance on the RetinaMNIST, PneumoniaMNIST, and DermaMNIST datasets, whereas the RGB-based method yields the best performance on the OrganSMNIST dataset. These results show that the optimal vector field construction strategy can vary with dataset characteristics. This could enhance the performance of our method and be further explored in the future study.

We have added these to section 2.4 of the revised supplementary information.

***Comment** : 7) Discretization errors and orthogonality checks: Provide results showing the discrete decomposition is numerically orthogonal. Report:*

Algorithm 5: Computing Hodge Decomposition

Input: Vector field ω

Output: The decomposition $(d\alpha_n, \delta\beta_t, h)$ where $\omega = d\alpha_n + \delta\beta_t + h$

Use Algorithm 1 to get the discrete manifold (M_n, M_t);

Use Algorithm 2 and (M_n, M_t) to compute the projection matrix $P_{k,n}, P_{k,t}$

Use Algorithm 3 to compute the boundary matrix ∂_k;

Use Algorithm 4 to compute the Hodge star $S'_{k,n}, S'_{k,t}$;

Set $D_k = \partial_{k+1}^T$;

Set the discrete differential $D_{0,n} = P_{1,n} D_0 P_{0,n}^T$, $D_{1,t} = P_{2,t} D_1 P_{1,t}^T$;

Set the discrete Hodge star $S_{k,n} = P_{k,n} S'_{k,n} P_{k,n}^T$, $S_{k,t} = P_{k,t} S'_{k,t} P_{k,t}^T$;

Set the discrete Laplacian $L_{0,n} = D_{0,n}^T S_{1,n} D_{0,n}$, $L_{2,t} = S_{2,t} D_{1,t} S_{1,t}^{-1} D_{1,t}^T S_{2,t}$;

Set $\delta_{1,n} = S_{0,n}^{-1} D_{0,n}^T S_{1,n}$;

Solve $L_{0,n}\alpha_n = \delta_{1,n}\omega$ to get α_n;

Solve $L_{2,t}\beta_t = D_{1,t}\omega$ to get β_t;

Set $\delta_{2,t} = S_{1,t}^{-1} D_{1,t}^T S_{2,t}$;

Set $h = \omega - D_{0,n}\alpha_n - \delta_{2,t}\beta_t$;

return $(D_{0,n}\alpha_n, \delta_{2,t}\beta_t, h)$

$$err_{recon} = \frac{||v - (d_o\phi + \delta\psi + h)||_2}{||v||_2}, err_{ortho} = \frac{| < d_0\phi, \delta\psi > |}{||d_0\phi||_2 ||\delta\psi||_2}$$

Provide these for representative images or in a table for a dataset with mean ± std.

***Answer**: We thank the reviewers for the constructive suggestions.*

For reconstruction error, in our computation, we first compute the curl-free component $d\phi$ and the divergence-free component $\delta\psi$, and then the harmonic component is obtained by

$$h = v - d\phi - \delta\psi$$

As a result, the reconstruction error is always 0.

For orthogonality error, we computed the normalized pairwise inner product between the curl-free, divergence-free, and harmonic parts on five representative datasets: RetinaMNIST (1,600 samples), PneumoniaMNIST (5856 samples), DermaMNIST (10,015 samples), OrganAMNIST (58,830 samples), and PathMNIST (107,180 samples). The normalized inner product for two components α, β is defined as

$$(\alpha, \beta) = \frac{| < \alpha, \beta > |}{||\alpha||_2 ||\beta||_2}.$$

We report the mean and standard deviation of these values across all samples for each dataset, which is shown in Table S7. The results indicate that the three components exhibit strong numerical orthogonality across all evaluated datasets.

We have added these discussions to section 1.3 of the revised supplementary information.

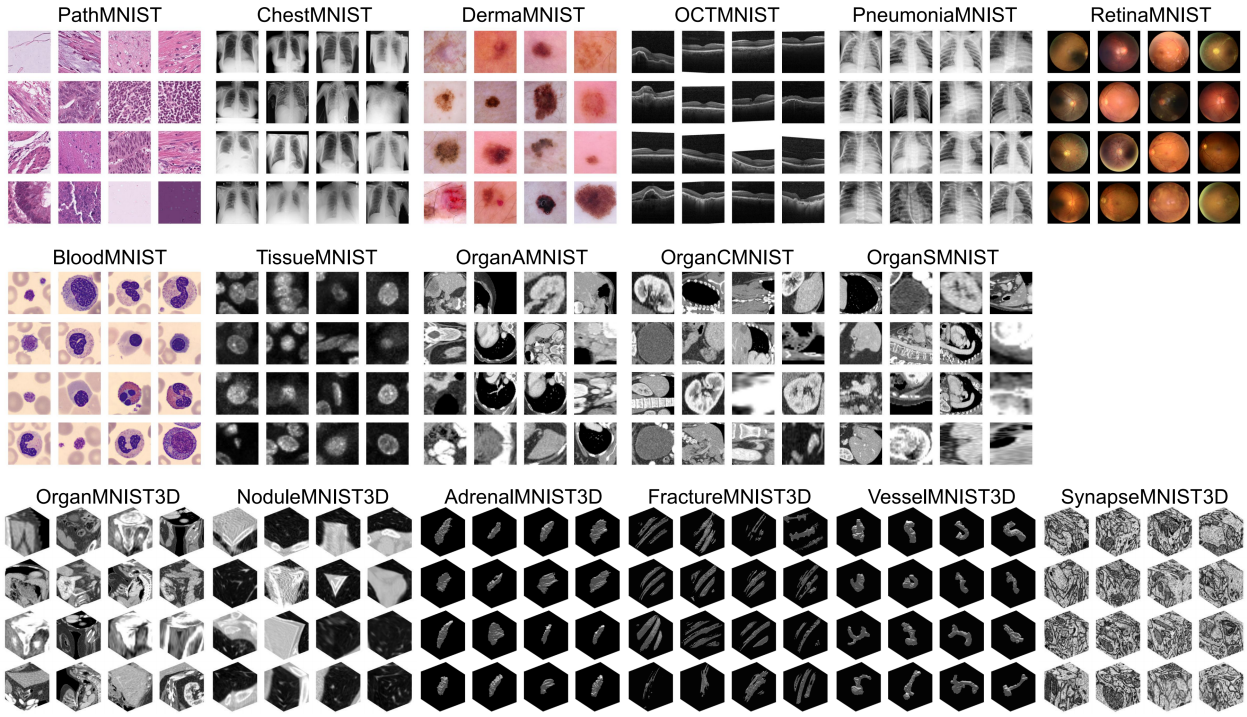

Figure S3: Illustration of the 11 2D datasets and 6 3D datasets in the MedMNIST database.

Comment *: 8) Comparisons / fairness of baselines:*

Answer*: We thank the reviewers for the comments. Our manuscript was originally submitted earlier this year, so the recently published related work has not been included. To provide an updated and comprehensive comparison, we conducted a thorough literature search for works published in 2025 that also utilized this database. We summarize these newly identified results in Table S8 to present an up-to-date benchmark comparison. It can be seen that our model achieves the best overall performance among all these methods. Notably, many of the recently published models are evaluated on only a subset rather than the full set of datasets while our model is evaluated on all the datasets. These new results have been added into section 9 of the revised supplementary information.*

[1] *Halder, A., Dalal, A., Gharami, S., Wozniak, M., Ijaz, M. F., & Singh, P. K. (2025). A fuzzy rank-based deep ensemble methodology for multi-class skin cancer classification. Scientific Reports, 15(1), 6268.*

[2] *Wu, B., Huang, J., & Duan, Q. (2025). Real-time intelligent healthcare enabled by federated digital twins with aoi optimization. IEEE Network.*

[3] *Wang, G., Zhu, Q., Song, C., Wei, B., & Li, S. (2025). MedKAFormer: When Kolmogorov-Arnold Theorem Meets Vision Transformer for Medical Image Representation. IEEE Journal of Biomedical and Health Informatics.*

[4] *Bai, Y., Bai, L., Yang, X., & Liang, J. (2025). Label-Semantic-Based Prompt Tuning for Vision Transformer Adaptation in Medical Image Analysis. IEEE Transactions on Circuits and Systems for Video Technology.*

Table S5: Performance comparison (AUC and ACC) across different segmentation cutoff parameters on representative datasets.

Cutoff	Retina		Pneumonia		Derma		OrganS	
	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC
0	0.874	0.655	0.986	0.910	0.962	0.836	0.978	0.809
5	0.872	0.667	0.986	0.922	0.962	0.843	0.978	0.809
10	0.870	0.655	0.985	0.915	0.964	0.832	0.978	0.802
15	0.878	0.643	0.983	0.914	0.962	0.842	0.977	0.806
20	0.874	0.673	0.985	0.917	0.963	0.838	0.978	0.808
25	0.875	0.658	0.986	0.914	0.961	0.831	0.977	0.802
30	0.870	0.643	0.985	0.910	0.965	0.839	0.978	0.803

Table S6: Performance (ACC & AUC) Comparison of different vector field generation methods, including the gradient-based method with parameter (s, t), the flow-based method, and RGB-based method.

Method	Retina		Pneumonia		Derma		OrganS	
	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC
Gradient(1,1)	0.874	0.655	0.986	0.910	0.962	0.836	0.978	0.809
Gradient(1,2)	0.864	0.643	0.982	0.904	0.958	0.815	0.979	0.815
Gradient(2,1)	0.854	0.613	0.983	0.899	0.953	0.821	0.979	0.818
Flow	0.815	0.563	0.973	0.901	0.931	0.785	0.976	0.792
RGB	0.803	0.585	0.976	0.867	0.952	0.816	0.981	0.818

[5] *Singh, A., Eising, C., Denny, P., & van de Ven, P. (2025). Improving GNNs for Image Classification: Addressing Homophily Challenges. IEEE Open Journal of the Computer Society.*

[6] *Zechen, Z., Xuelei, H., Fengjun, Z., & Xiaowei, H. (2025). PSNAS-Net: Hybrid gradient-physical optimizationfor efficient neural architecture search in customized medical imaging analysis. Expert Systems with Applications, 288, 128155.*

[7] *Du, Y., Zhang, J., Zeevi, T., Dvornek, N. C., & Onofrey, J. A. (2025, April). Sre-conv: Symmetric rotation equivariant convolution for biomedical image classification. In 2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI) (pp. 1-5). IEEE.*

[8] *Singh, R., & Guggilam, S. (2025). Iterative Misclassification Error Training (IMET): An Optimized Neural Network Training Technique for Image Classification. IEEE Access.*

[9] *Rahman, M., & Zhuang, J. (2025, September). NQNN: Noise-Aware Quantum Neural Networks for Medical Image Classification. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 433-442). Cham: Springer Nature Switzerland.*

***Comment** : 9) Provide Pseudo-code for the proposed methodological pipeline.*

***Answer**: We thank the reviewers for the constructive suggestions. We have added the pseudocode in the revised supplementary information.*

Table S7: Numerical analysis of the orthogonality of the three components in the Hodge decomposition on the RetinaMNIST, PneumoniaMNIST, DermaMNIST, OrganAMNIST, and PathMNIST datasets. The average and standard deviation values of the normalized inner product of all samples within each dataset are reported.

Method	Retina		Pneumonia		Derma		OrganA		Path	
	mean	std	mean	std	mean	std	mean	std	mean	std
$(d\phi, \delta\psi)$	7e-20	4e-17	2e-19	2e-17	4e-20	3e-17	8e-20	3e-17	7e-20	2e-17
$(d\phi, h)$	2e-13	9e-14	3e-14	2e-14	6e-14	1e-14	3e-14	3e-14	2e-14	1e-14
$(\delta\psi, h)$	2e-17	5e-17	2e-18	8e-17	7e-18	2e-16	5e-18	8e-17	2e-17	6e-17

Comment *: 10) The paper reports substantial improvements across multiple MedMNIST tasks; however, it is unclear whether the baseline models were trained and tuned under identical experimental conditions (e.g., data augmentation, optimizer settings, and hyperparameter search strategies). Please clarify whether the baseline results were all adopted from the original publications or reproduced by the authors. If any were re-implemented, details on the hyperparameter parity and training cost should be provided. Without this, it's hard to know whether gains derive from topology or from training differences.*

Answer*: We thank the reviewers for pointing this out. All baseline results reported in our manuscript are directly taken from their original publications. The MedMNIST v2 database provides clearly defined training, validation, and testing splits, and we have strictly followed these official splits to ensure a fair comparison with existing baseline models. We have emphasized this point in section 2.2.2 of the revised manuscript.*

Comment *: 11) Statistical analysis: While the reported results on MedMNIST v2 are promising, the paper lacks statistical analysis. Since many datasets are small and stochastic training introduces non-negligible variance, it is important to report mean +− standard deviation across multiple runs and perform statistical significance testing. A paired t-test or Wilcoxon test within each dataset, and a Friedman test with Nemenyi post-hoc across datasets, would clarify whether the observed improvements are statistically significant rather than due to random variation. Reporting confidence intervals and effect sizes (e.g., Cohen's d or Cliff's δ) would further strengthen the empirical evidence.*

Answer*: We thank the reviewers for the valuable suggestions.*

For mean and standard deviation, all reported results of our MTDL model are the averaged values across three independent runs with different random seeds. We have added the corresponding std value in section 9 of the revised supplementary information.

For the paired t-test and Cohen's d analysis, the existing baseline models only provide single performance values without reporting variance across multiple runs. Therefore, we cannot directly perform paired t-tests or compute Cohen's d between our model and these baselines, because both analyses require multiple runs for each method to estimate variance.

For the Friedman test, we use the scipy.stats.friedmanchisquare package to compute the p-values on 11 2D datasets and 6 3D datasets, respectively. For the 2D analysis, we only

Table S8: Performance comparison between our model and recent models on the MedMNIST2D dataset. The benchmark results are taken from the original papers. The best results are in bold. "-" indicates that the corresponding metric was not reported.

Methods	Path		Chest		Derma		OCT	
	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC
FRDEM [1]	-	-	-	-	-	0.783	-	-
AIGC [2]	-	-	-	0.83	-	-		
MedKAFormer [3]	0.985	0.882	-	-	0.919	0.748	0.974	0.810
LPT [4]	-	**0.941**	-	-	-	0.787	-	0.832
SGDGCN [5]	0.987	0.907	**0.928**	0.854	0.919	0.771	0.933	0.783
PSNAS-Net-ave [6]	-	-	-	-	-	0.772	-	-
SRE-CONV [7]	-	0.822	-	**0.948**	-	0.757	-	0.768
IMET [8]	-	-	-	-	-	-	0.960	0.803
NQNN [9]	-	0.752	-	-	-	0.827	-	0.591
MTDL	**0.996**	0.920	0.793	**0.948**	**0.962**	**0.836**	**0.989**	**0.888**

Methods	Pneumonia		Retina		Blood		Tissue	
	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC
MedKAFormer [3]	0.961	0.904	0.729	0.535	0.998	0.965	-	-
LPT [4]	-	-	-	0.575	-	0.967	-	-
SGDGCN [5]	0.963	**0.918**	0.782	0.550	0.994	0.940	0.801	0.613
PSNAS-Net-ave [6]	-	0.899	-	-	-	0.956	-	-
SRE-CONV [7]	-	0.906	-	0.523	-	0.960	-	0.712
IMET [8]	0.895	0.902	-	-	-	-	-	-
NQNN [9]	-	-	-	-	-	0.737	-	-
MTDL	**0.986**	0.910	**0.874**	**0.655**	**0.999**	**0.988**	**0.945**	**0.721**

Methods	OrganA		OrganC		OrganS	
	AUC	ACC	AUC	ACC	AUC	ACC
AIGC [2]	-	0.890	-	-	-	-
MedKAFormer [3]	0.995	0.922	0.991	0.899	**0.986**	0.788
LPT [4]	-	-	-	0.906	-	0.796
SGDGCN [5]	0.977	0.850	0.989	0.880	0.969	0.784
PSNAS-Net-ave [6]	-	-	-	0.913	-	-
SRE-CONV [7]	0.770	0.769	-	-	-	-
NQNN [9]	-	0.814	-	0.799	-	-
MTDL	**0.998**	**0.956**	**0.996**	**0.928**	0.978	**0.809**

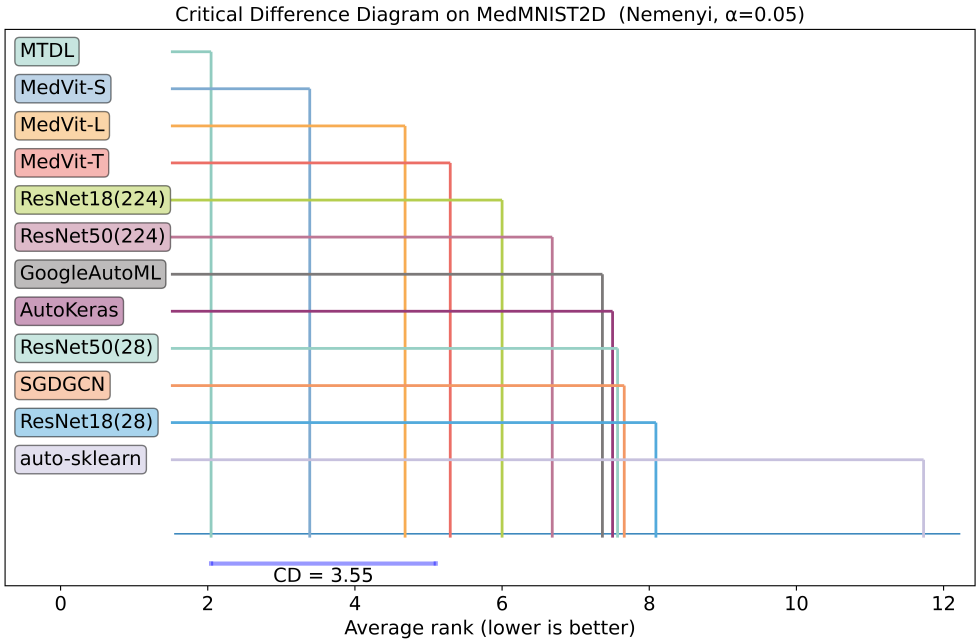

Figure S4: The Critical Difference (CD) diagram over 11 MedMNIST2D datasets.

include baseline models that report results on all 11 MedMNIST2D datasets. Similarly, for the 3D analysis, we only include baselines that report results on all 6 MedMNIST3D datasets. The resulting p-values are 1.38e-14 (2D) and 5.32e-10 (3D), indicating significant performance differences among the compared models. Therefore, we provide Critical Difference (CD) diagrams for both the 2D and 3D cases. Figure S4 shows the CD diagram comparing 12 models across the 11 MedMNIST2D datasets using the Nemenyi post-hoc test at the significance level $\alpha = 0.05$. Our MTDL model achieves the best overall performance with the lowest average rank among all compared methods. The CD = 3.55 shown at the bottom indicates the minimum rank difference required for two models to be considered significantly different under the Nemenyi test. Under this threshold, MTDL is significantly better than most compared methods. Similarly, Figure S5 shows the results for the 6 3D datasets, where our MTDL model also achieves the best overall performance with the lowest average rank and is significantly better than most other methods.

We have added these analysis into section 6.1 of the revised supplementary information.

***Comment** : 12) In Figure 3, the range of the AUC or ACC values on the x-axis should be consistent across all plots to facilitate clearer and quicker visual comparison between results.*

***Answer**: We thank the reviewers for pointing this out. We have revised the figure accordingly.*

***Comment** : 13) Provide preprocessing time per sample (vector field + decomposition), GPU memory/CPU usage, and training time per epoch for representative datasets (2D and 3D).*

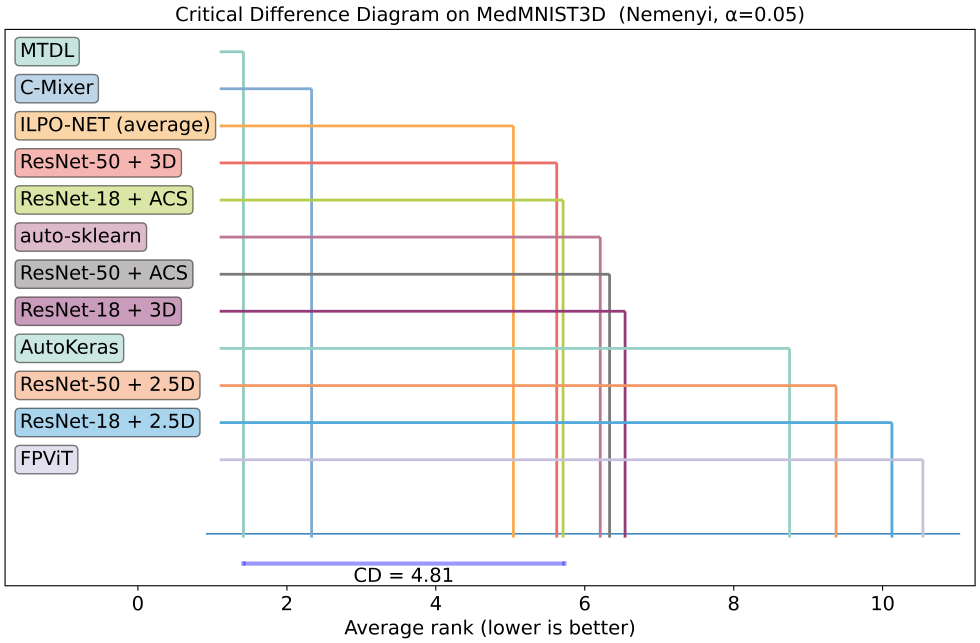

Figure S5: The Critical Difference (CD) diagram over 6 MedMNIST3D datasets.

Answer: *We thank the reviewers for the suggestions. In our experiments, all 2D images are of size $n \times 224 \times 224$, where $n = 3$ for RGB images and $n = 1$ for grayscale images. All 3D images are grayscale volumes with a resolution of $64 \times 64 \times 64$. The vector field generation and Hodge decomposition procedures are computationally efficient. When implemented on a single CPU core, processing 50 grayscale 2D images requires approximately 2 minutes, while processing 50 RGB 2D images requires around 4 minutes. For 3D images, the computation for 50 samples can be completed within 30 minutes. For model training, our MTDL framework is implemented in PyTorch and executed on an NVIDIA Tesla V100S GPU with 32 GB of memory. Representative training times for selected 2D and 3D datasets are reported in Table S9. As shown, the overall training process is also computationally efficient.*

These runtime analysis have been added into section 7 of the revised supplementary information.

Discussions:

Comment *: 14) Invariance/equivariance: discuss which geometric transforms of the input (translation, rotation, scaling) preserve each decomposed channel and whether the CNN architecture exploits or breaks these invariances.*

Answer: *We thank the reviewers for the constructive suggestions. The vector field generation method used in MTDL is translation-equivariant. However, it is not rotationally equivariant because the underlying Cartesian grid is not rotationally symmetric, nor is it scale-equivariant, since the grid spacing does not vary under geometric scaling. The Hodge decomposition exhibits the same property: it is translation-equivariant, but not rotationally*

Table S9: Training time per epoch of our MTDL model on representative 2D and 3D datasets.

Dataset	Sample Size	Image Type	Training Time Per Epoch (minutes)
RetinaMNIST	1,600	2D RGB	0.6
PneumoniaMNIST	5,856	2D gray	0.8
DermaMNIST	10,015	2D RGB	3.2
OrganAMNIST	58,830	2D gay	5.9
PathMNIST	107,180	2D RGB	35.1
OCTMNIST	109,309	2D gray	16.3
TissueMNIST	236,386	2D gray	32.1
OrganMNIST3D	1,742	3D gray	2.5
SynapseMNIST3D	1,759	3D gray	2.9
VesselMNIST3D	1,908	3D gray	3.1

or scale-equivariant.

For the CNN architecture employed in our model, the convolutional layers are translation-equivariant, and the use of global average pooling at the final stage guarantees translation-invariance at the classifier level. Therefore, the overall MTDL model achieves translation-invariance. This analysis suggests that our framework could be further extended to incorporate additional forms of equivariance or invariance, such as rotational or scale equivariance.

***Comment** : 15) Physical/semantic interpretation: Can we link the decomposed components to expected image structure per modality (e.g., curl-free edge/gradient information; divergence-free texture swirl)? so readers can interpret classifier reliance on components.*

***Answer**: We thank the reviewers for the valuable comments. In our model, each image is first represented as a vector field, which is subsequently decomposed into a curl-free component, a divergence-free component, and a harmonic component. These components have well-established geometric and physical interpretations. Specifically, a curl-free vector field exhibits no local rotational behavior. A divergence-free vector field has neither sources nor sinks. Source regions correspond to positive divergence, where local intensity or structural patterns spread outward, while sink regions correspond to negative divergence, where patterns converge inward. The harmonic component is both curl-free and divergence-free, capturing the smoothest global properties present in the vector field.*

In medical image analysis, the curl-free component can be used to capture the lesion boundaries and margins, as the directions of maximal intensity change typically align with anatomical edges. For example, in mammography, malignant lesions often show sharper and more irregular boundaries compared with the benign lesions. Therefore, the curl-free component can be used to distinguish the malignant from benign cases. The divergence-free component captures the vorticity and local rotating behavior in the vector field. Consequently, it can be used to capture the irregular textures, such as the swirling patterns and the heterogeneous internal textures due to necrosis and fibrosis. Especially, this will be useful for malignant tumors since the swirling textures are common in malignant tumors such as breast cancer, brain

tumor, and hepatocellular carcinoma. The harmonic component reflects the global topological properties of the image and provides robust descriptors insensitive to noises and small perturbations. For a more comprehensive understanding of how the decomposed components relate to visual patterns such as edges, textures, and global structures, additional interpretability analysis would be required. We regard this as an important direction for future work.

***Comment** : 16) How the method might generalize to real-world biomedical images, e.g., high-resolution CT/MRI, anisotropic voxels, or dynamic sequences? MedMNIST is small and standardized; clinical data might pose challenges.*

***Answer**: We thank the reviewers for pointing out this. For high-resolution real-world biomedical images, we have an ablation study in Section 2.2.5 of the main text. Specifically, we evaluate our model on the clinical HAM10000 dataset, which contains 10,015 dermatoscopic images of uniform size* $3 \times 600 \times 450$*. We center-crop each image to* $3 \times 450 \times 450$ *and then resize it to five different resolutions:* $3 \times 35 \times 35$*,* $3 \times 75 \times 75$*,* $3 \times 150 \times 150$*,* $3 \times 300 \times 300$*, and* $3 \times 450 \times 450$*. Our MTDL model is then evaluated on each resolution. Here we list the results in Table S10. It can be seen that the model performance improves consistently as the input resolution increases, indicating that MTDL is capable of extracting richer and more detailed features from higher-resolution inputs. Notably, even at the lowest resolution* $3 \times 35 \times 35$*, MTDL achieves an AUC (ACC) of 0.943 (0.797), demonstrating that the model maintains strong and stable performance across a wide range of image resolutions.*

Table S10: Performance of MTDL over different image reslutions on the clinical dataset HAM10000.

Resolution	$3 \times 35 \times 35$	$3 \times 75 \times 75$	$3 \times 150 \times 150$	$3 \times 300 \times 300$	$3 \times 450 \times 450$
AUC	0.943	0.947	0.958	0.970	0.973
ACC	0.797	0.803	0.822	0.856	0.863

For dynamic sequences, Hodge decomposition can be directly applied to analyze the temporal evolution of the three components and thereby characterize dynamic patterns. Alternatively, representative slices from the sequence can be selected, and Hodge decomposition can be performed independently on each slice. The resulting components can then be concatenated to form a multi-time, multi-channel image representation, which can be combined with a CNN to perform prediction tasks.

***Comment** : 17) A discussion (or ideally an experiment) comparing MTDL to surface-based or mesh-based features could be helpful. This might help readers understand when grid-based manifold representation is preferred vs direct mesh representation or potential ways these approaches could complement each other in biomedical imaging applications.*

***Answer**: We thank the reviewers for the thoughtful comments. The Hodge theory employed in this work is formulated on a Cartesian grid. This corresponds to the Eulerian discretization of the continuous Hodge theory. An alternative is the Lagrangian discretization, in which the Hodge operators are defined on triangular or tetrahedral meshes. The choice between Eulerian and Lagrangian approaches is usually determined by the nature of the underlying data. For example, if a well-defined surface segmentation is available, the*

surface can be naturally represented by a triangular mesh, making the Lagrangian method directly applicable. In contrast, when the data are given on a regular grid, such as 2D or 3D biomedical images, the Eulerian method can be applied without additional processing. It is important to note that Lagrangian methods typically require accurate surface extraction and mesh construction, which is not feasible when a clean and anatomically meaningful surface is not present or cannot be robustly segmented.

For biomedical image analysis, the Eulerian formulation is more suitable compared to the Lagrangian formulation for two reasons: (1) The primary data modality in clinical practice consists of images that include background regions, noise, and heterogeneous tissue boundaries. Applying a Lagrangian Hodge framework would require an additional segmentation step to extract surface meshes, which introduces methodological complexity and potential segmentation-induced variability. (2) Biomedical images are inherently defined on a Cartesian grid, providing a natural domain for Eulerian discrete differential operators. This natural grid structure makes Eulerian Hodge decomposition both computationally efficient and well-aligned with the data representation in medical image analysis.

***Comment** : 18) The paper could benefit from explicit discussion of scenarios where MTDL might fail.*

***Answer**: We thank the reviewers for pointing out this. (1) MTDL relies on the construction of image-derived vector fields and their Hodge decomposition. Therefore, when the image lacks meaningful intensity gradients or is dominated by noise, the curl-free, divergence-free, and harmonic components may not produce informative representations. For example, images with extremely low contrast or strong speckle noise may lead to unstable decompositions. (2) MTDL is based on Cartesian grids. Consequently, for tasks where the studied object is inherently a high-resolution surface manifold, such as the brain surface, mesh-based or surface-based Hodge decomposition may be more suitable. In such cases, applying MTDL directly to surface data may fail to capture important intrinsic geometric features.*

Response to Reviewers

Answers to Reviewer 1' Comments

I thank the authors for the detailed response and the extensive additional experiments. Overall, I'm satisfied with the response. I only have a few remaining concerns.

Answer: We thank the reviewer for the positive comments.

Comment *: 1) Regarding Comment 2 on interpretability. I noticed that another reviewer raised a related concern in Comment 15. While I accept the abstract explanations, I was hoping to see some concrete evidence in the specific application. For example, the authors mention that "the directions of maximal intensity change typically align with anatomical edges", "in mammography, malignant lesions often show sharper and more irregular boundaries compared with benign lesions", and "swirling patterns and heterogeneous internal textures arise due to necrosis and fibrosis". I do not expect all of these claims to be demonstrated, but even one or two illustrative examples would significantly improve the clarity and persuasiveness of the interpretability claims.*

Answer*: We thank the reviewer for the comments. Mammography refers to X-ray imaging of the breast, whereas ultrasound is a different imaging modality. Here we use the ultrasound image to illustrate the general observation that malignant breast lesions often exhibit more irregular boundaries compared with benign cases. As shown in Figure S1, the ground-truth annotation clearly delineates the lesion boundary, where the malignant case presents a more irregular contour.*

Comment *: 2) Finally, after reading the rebuttal to another reviewer, I agree that the lack of rotation and scale equivariance appears to be an inherent limitation of the proposed method. It would be helpful if this limitation could be properly addressed, or at least stated more explicitly in the discussion.*

Answer*: We thank the reviewer for the comments. Due to the page limit, a discussion of this issue has been given to Section 1.6 of the revised Supplementary Information.*

Comment *: 3) Moreover, a few other things to certainly correct for: Introduction exemplifies that the 'DNA chains' as differentiable manifolds. I think this is incorrect and must be toned down - differentiable manifolds can only approximate these at best.*

Answer*: We thank the reviewer for the comments. The term "DNA chains" has been removed in the revised manuscript.*

[editorial note: third party material redacted]

Figure S1: Comparison of benign and malignant breast lesions in ultrasound images. The images were obtained from Al-Dhabyani et al., Data in Brief (2020), 28:104863.

Comment *: 4) Is the original image fully recoverable by the decomposition, e.g. by gradient reconstruction etc.? Which information do we lose particularly? For which data is this not acceptable? Given that only some medical data is used, such a discussion would be helpful.*

Answer*: We thank the reviewer for pointing out this. In our model, the original image* $\mathcal{I}$ *is first transformed into a discrete vector field* $V(i,j) = (x_{i,j}, y_{i,j})$ *via*

$$\begin{aligned} x_{i,j} &= \frac{\mathcal{I}(i+1,j) - \mathcal{I}(i-1,j)}{2}, \\ y_{i,j} &= \frac{\mathcal{I}(i,j+1) - \mathcal{I}(i,j-1)}{2}. \end{aligned} \tag{1}$$

The Hodge decomposition is then applied to this vector field. By construction, the decomposition is algebraically exact for the discrete vector field, meaning that V *can be reconstructed from its components within numerical precision.*

Given a prescribed value at one reference point $\mathcal{I}(i_0, j_0)$*, the remaining values* $\mathcal{I}(i,j)$ *can be computed from the vector field* V *with formula (1). However, since the absolute value*

at the reference point is unknown, the reconstructed image may differ from the original by an additive constant. Therefore, the original image can be recovered only up to a constant offset.

It is true that there are situations in which the computation of Hodge decomposition may fail. In particular, when the imaged object contains extremely thin structures, accurate resolution of these features requires a substantially refined discretization grid. Although, in principle, a sufficiently fine grid allows for meaningful decomposition of a fixed-size object, in practice this significantly increases computational cost. Consequently, Hodge decomposition may become computationally impractical for data dominated by very thin structures.

Answers to Reviewer 3' Comments

Comment *: Thank you to the authors for the thoughtful and thorough review. I am pleased that all of my previous concerns have been fully addressed.*

Answer*: We are happy that the reviewer is satisfied with our revision.*

Co-Review of NCOMMS-25-20703-T; Nature Communications

Title: Manifold Topological Deep Learning for Biomedical Data

Summary:

The manuscript introduces Manifold Topological Deep Learning (MTDL), which represents images as discrete manifolds, constructs a vector field on that manifold, applies a discrete Hodge decomposition into curl-free / divergence-free / harmonic components, concatenates those channels, and feeds the result into a CNN. The model is evaluated on MedMNIST v2 (11 2D + 6 3D datasets) and reported to outperform prior methods.

The idea of converting images to vector fields and apply a discrete Hodge decomposition before feeding the components to a CNN model for classification is promising. However, the paper needs further clarification on methodological details, additional ablation studies, and rigorous evaluation. Detailed comments below outline specific improvements needed before publication.

Major revision:

Methodology:

1. Explain how the harmonic component h is computed.
2. Provide Pseudo-code for the proposed methodological pipeline.
3. Clarify how the manifold M and its boundary ∂M are defined for each dataset. In practical settings, computing normal and tangential boundary conditions requires a segmentation mask or at least a binary indicator of the region of interest. In Section 4.3.1, the segmentation step is described only as thresholding but the choice of threshold (fixed, adaptive, or dataset-specific), preprocessing (e.g., normalization or filtering), and handling of ambiguous or noisy background regions can significantly influence the resulting manifold boundary ∂M, which in turn affects the computed boundary conditions and the subsequent Hodge decomposition. I recommend the authors (1) explicitly describe the segmentation procedure per dataset, (2) discuss its influence on the topological representation, and (3) provide an ablation or sensitivity analysis to assess robustness to segmentation variations in the results.
4. The paper would strongly benefit from an explicit pseudocode summarizing the full implementation workflow, including how boundary conditions, discrete operators, and vector-field decompositions are computed. Although the Supplementary Material gives substantial mathematical detail, a procedural description (with solver specifications and preprocessing steps) is essential for reproducibility.

Results:

5. Vector-field construction is under-specified in the main text. The paper mentions several methods and refers to the Supplementary Information, but the main text does not make clear which construction is used for each experiment, what hyperparameters were used, and how sensitive results are to that choice. Since the vector field is the core representation, an ablation is necessary: different vector-field constructions (flow-based, gradient-based, others) and their effect on final performance.
6. Discretization errors and orthogonality checks:

   Provide results showing the discrete decomposition is numerically orthogonal.

   Report:

   $\mathrm{err}_{\mathrm{recon}} = \frac{\|v-(d_0\phi+\delta\psi+h)\|_2}{\|v\|_2}$ and $\mathrm{err}_{\mathrm{ortho}} = \frac{|\langle d_0\phi,\delta\psi\rangle|}{\|d_0\phi\|_2\|\delta\psi\|_2}$.

   Provide these for representative images or in a table for a dataset with mean ± std.
7. The paper concatenates components as extra channels. Provide an ablation showing which components contribute most to performance (curl-free / divergence-free / harmonic).
8. Comparisons / fairness of baselines:
9. The paper reports substantial improvements across multiple MedMNIST tasks; however, it is unclear whether the baseline models were trained and tuned under identical experimental conditions (e.g., data augmentation, optimizer settings, and hyperparameter search strategies). Please clarify whether the baseline results were all adopted from the original publications or reproduced by the authors. If any were re-implemented, details on the hyperparameter parity and training cost should be provided. Without this, it's hard to know whether gains derive from topology or from training differences.
10. Statistical analysis: While the reported results on MedMNIST v2 are promising, the paper lacks statistical analysis. Since many datasets are small and stochastic training introduces non-negligible variance, it is important to report mean ± standard deviation across multiple runs and perform statistical significance testing. A paired t-test or Wilcoxon test within each dataset, and a Friedman test with Nemenyi post-hoc across datasets, would clarify whether the observed improvements are statistically significant rather than due to random variation. Reporting confidence intervals and effect sizes (e.g., Cohen's d or Cliff's δ) would further strengthen the empirical evidence.
11. In Figure 3, the range of the AUC or ACC values on the x-axis should be consistent across all plots to facilitate clearer and quicker visual comparison between results.

12. Provide preprocessing time per sample (vector field + decomposition), GPU memory/CPU usage, and training time per epoch for representative datasets (2D and 3D).

Discussions:

13. Invariance/equivariance: discuss which geometric transforms of the input (translation, rotation, scaling) preserve each decomposed channel and whether the CNN architecture exploits or breaks these invariances.
14. Physical/semantic interpretation: Can we link the decomposed components to expected image structure per modality (e.g., curl-free ~ edge/gradient information; divergence-free ~ texture swirl)? so readers can interpret classifier reliance on components.
15. How the method might generalize to real-world biomedical images, e.g., high-resolution CT/MRI, anisotropic voxels, or dynamic sequences? MedMNIST is small and standardized; clinical data might pose challenges.
16. A discussion (or ideally an experiment) comparing MTDL to surface-based or mesh-based features could be helpful. This might help readers understand when grid-based manifold representation is preferred vs direct mesh representation or potential ways these approaches could complement each other in biomedical imaging applications.
17. The paper could benefit from explicit discussion of scenarios where MTDL might fail.