

Supplementary Information

Manifold Topological Deep Learning for Biomedical Data

Xiang Liu¹, Zhe Su², Yongyi Shi³, Yiyong Tong⁴, Ge Wang³, and Guo-Wei Wei ^{*1,5,6}

¹Department of Mathematics, Michigan State University, MI, 48824, USA

²Department of Mathematics and Statistics, Auburn University, AL, 36849, USA

³Biomedical Imaging Center, Rensselaer Polytechnic Institute, NY, 12180, USA

⁴Computer Science and Engineering, Michigan State University, MI 48824, USA

⁵Department of Mathematics, University of Georgia, Athens, GA 30602, USA

⁶Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30602, USA

Contents

1	Hodge theory	3
1.1	Hodge Theory in the Continuous Case	3
1.1.1	Hodge Laplacian	3
1.1.2	Hodge Decomposition for Closed Manifolds	4
1.1.3	Hodge Decomposition for Manifolds with Boundary	4
1.2	Discrete Topology-preserving Hodge Theory on Cartesian Grids	5
1.2.1	Discrete Manifolds with Boundary	6
1.2.2	Discrete Differential Forms	6
1.2.3	Discrete Exterior Derivative Operator	6
1.2.4	Discrete Hodge Star Operator	6
1.2.5	Discrete Hodge Laplacian	7
1.2.6	Discrete Hodge Decomposition	7
1.3	Numerical Analysis of the Orthogonality of the Decomposition	7
1.4	Algorithm	8
1.5	Example	9
1.6	Limitations	13
2	Vector Field Generation	13
2.1	Gradient-based Method	13
2.2	Flow-based Method	13
2.3	Other Methods	14
2.4	Evaluation of Different Methods	14

*Corresponding author: weig@msu.edu

3	Dataset Details	14
3.1	2D Datasets	15
3.2	3D Datasets	22
4	Data Separation by Task Type, Data Scale, and Data Modality	22
4.1	Data Modality	33
4.2	Data Scale	33
4.3	Task Type	33
5	Model Details	34
5.1	CNN Architecture	34
5.2	Data-distribution-based Model Layer Design	35
6	Performance Analysis	35
6.1	Statistical Significance Analysis	35
6.2	Ablation Study on the Components of the Hodge Decomposition	36
6.3	Ablation Study on the Segmentation Parameter	38
6.4	Interpretability Analysis	39
7	Runtime Analysis	39
8	Loss Function and Evaluation Metrics	40
8.1	Loss Function	40
8.2	Evaluation Metrics	40
9	Detailed Model Performance	41

1 Hodge theory

1.1 Hodge Theory in the Continuous Case

Here, we provide a comprehensive review of the Hodge decomposition in the continuous setting, as this work represents the first endeavor to apply the Hodge decomposition within the domain of medical image analysis.

1.1.1 Hodge Laplacian

Let M be an m -dimensional smooth, orientable, compact manifold with boundary, and let $\Omega^k(M)$ denote the space of differential k -forms on M , i.e., the space of all smooth, antisymmetric covariant tensor fields of degree k on M . A differential k -form can be integrated over any orientable k -dimensional submanifold of M . For any orientable $(k+1)$ -dimensional submanifold $S \subset M$ with boundary ∂S , Stokes' theorem states that the integral of a differential k -form ω over the boundary ∂S is equal to the integral of its differential over the manifold S . Explicitly, this is expressed as

$$\int_S d\omega = \int_{\partial S} \omega, \quad (1)$$

where the differential d (exterior derivative) is the unique $\mathbb{R}$ -linear mapping from $\Omega^k(M)$ to $\Omega^{k+1}(M)$ that satisfies the Leibniz rule with respect to the wedge product $\wedge$ and the property $dd = 0$. A differential k -form ω is called closed if $d\omega = 0$ and exact if there exists a $(k-1)$ -form ψ such that $d\psi = \omega$. The pair $(\Omega^*(M), d)$ forms a cochain complex known as the de Rham complex, and its k -th cohomology, denoted by $H_{DR}^k(M)$, is called the k -th de Rham cohomology of M .

Let g be a Riemannian metric on M and let $\langle \cdot, \cdot \rangle$ denote the inner product on $\Omega^k(M)$ induced by g . The Hodge star operator $\star$ is an isomorphism from the space of k -forms $\Omega^k(M)$ to the space of $(m-k)$ -forms $\Omega^{m-k}(M)$ satisfying the property

$$\omega \wedge \star \eta = \langle \omega, \eta \rangle_g \mu_g, \quad (2)$$

where ω and η are k -forms, and μ_g is the volume form induced by g on M . By taking the integral of formula (2), we obtain the Hodge L^2 -inner product on the space of differential forms $\Omega^k(M)$

$$(\omega, \eta) = \int_M \omega \wedge \star \eta. \quad (3)$$

The codifferential $\delta : \Omega^k(M) \rightarrow \Omega^{k-1}(M)$ is defined as

$$\delta = (-1)^{m(k-1)+1} \star d \star. \quad (4)$$

A differential k -form w is called coclosed if $\delta w = 0$, and coexact if there exists a $(k+1)$ -form ψ such that $\delta\psi = w$. The differential d and the codifferential δ satisfy the following relationship based on integration by parts

$$(d\omega, \eta) = (\omega, \delta\eta) + \int_{\partial M} \omega \wedge \star \eta, \quad (5)$$

where ω is a $(k-1)$ -form and η is a k -form. This shows that d and δ are adjoint if M is a closed manifold, i.e., $\partial M = \emptyset$.

The Hodge Laplacian for differential forms is defined as

$$\Delta = d\delta + \delta d. \quad (6)$$

The Laplacian operator maps k -forms to k -forms. The kernel of Δ is called the space of harmonic forms. We denote by $\mathcal{H}_{\Delta}^k(M)$ the space of harmonic k -forms and by $\mathcal{H}^k(M)$ the space of k -forms that are both closed and coclosed.

1.1.2 Hodge Decomposition for Closed Manifolds

Assume M is a closed manifold, i.e., a compact manifold without boundary. The standard Hodge decomposition [1] states that

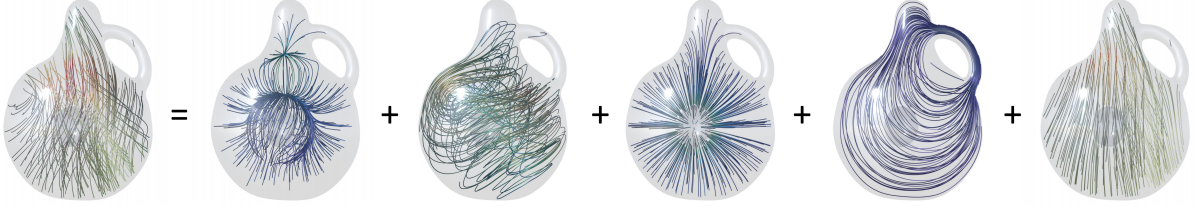
$$\Omega^k(M) = d\Omega^{k-1}(M) \oplus \delta\Omega^{k+1}(M) \oplus \mathcal{H}_\Delta^k(M), \quad (7)$$

where the adjointness of d and δ ensures that these three subspaces are orthogonal with respect to the inner product defined in (3). Consequently, any k -form can be uniquely decomposed as the sum of an exact form, a coexact form, and a harmonic form,

$$\omega = d\alpha + \delta\beta + h, \quad (8)$$

where $\omega \in \Omega^k(M)$, $\alpha \in \Omega^{k-1}(M)$, $\beta \in \Omega^{k+1}(M)$, $h \in \mathcal{H}_\Delta^k(M)$. The Hodge isomorphism theorem asserts that the space of harmonic k -forms is isomorphic to the k -th de Rham cohomology $H_{DR}^k(M)$ of M , this implies that the dimension of the harmonic space $\mathcal{H}_\Delta^k(M)$ is a topological invariant of the manifold, determined entirely by its topology.

1.1.3 Hodge Decomposition for Manifolds with Boundary



Supplementary Figure 1. Illustration of the 3D Hodge decomposition on a pear with a tunnel model. From left to right: the original vector field, the curl-free field, the divergence-free field, the normal harmonic field, the tangential harmonic field, and the curl gradient field.

When M is a manifold with non-empty boundary, the operators d and δ are generally not adjoint, as noted in (5). To ensure their adjointness and consequently achieve an orthogonal decomposition of differential forms, appropriate boundary conditions must be imposed.

The two most commonly used boundary conditions that ensure the adjointness of d and δ are the normal (Dirichlet) boundary condition and the tangential (Neumann) boundary condition. A form is called normal if it vanishes when applied to tangential vectors of the boundary and tangential if its dual vanishes when applied to tangential vectors of the boundary. These conditions define the following subspaces,

$$\Omega_n^k(M) = \{\omega \in \Omega^k(M) \mid \omega|_{\partial M} = 0\}, \quad \Omega_t^k(M) = \{\omega \in \Omega^k(M) \mid \star\omega|_{\partial M} = 0\}. \quad (9)$$

The Hodge star $\star$ provides an isomorphism between $\Omega_n^k(M)$ and $\Omega_t^{m-k}(M)$.

The Hodge-Morrey decomposition [2] states that

$$\Omega^k(M) = d\Omega_n^{k-1}(M) \oplus \delta\Omega_t^{k+1}(M) \oplus \mathcal{H}^k(M), \quad (10)$$

where the boundary conditions ensure the adjointness of d and δ , guaranteeing the orthogonality of the decomposition. The exterior derivative d preserves the normal boundary condition and the codifferential δ preserves the tangential boundary condition. As a result, any k -form can be decomposed as the sum of an exact normal form, a coexact tangential form, and a form that is both closed and coclosed.

$$\omega = d\alpha_n + \delta\gamma_t + \eta, \quad (11)$$

where $\omega \in \Omega^k(M)$, $\alpha_n \in \Omega_n^{k-1}(M)$, $\gamma_t \in \Omega_t^{k+1}(M)$, $\eta \in \mathcal{H}^k(M)$. To compute the components of this decomposition, we can firstly determine the potentials α_n and γ_t , and then compute η as $\eta = \omega - d\alpha_n - \delta\gamma_t$.

However, the potentials α_n and γ_t are not unique, as $\alpha_n + d\eta$ and $\gamma_t + \delta\gamma$, with any $\eta \in \Omega_n^{k-2}(M)$ and $\gamma \in \Omega_t^{k+2}(M)$, can serve as valid potentials for the first two terms. This issue can be solved by imposing gauge conditions. Specifically, we restrict

$$\alpha_n \in \ker \delta \cap \Omega_n^{k-1}(M), \quad \gamma_t \in \ker d \cap \Omega_t^{k+1}(M). \quad (12)$$

Under these conditions, the potentials satisfy the following equations,

$$\delta\omega = \delta d\alpha_n = (\delta d + d\delta)\alpha_n = \Delta\alpha_n, \quad d\gamma_t = d\delta\gamma_t = (d\delta + \delta d)\gamma_t = \Delta\gamma_t. \quad (13)$$

Up to a difference of harmonic forms in the kernel of Δ , the potentials α_n and γ_t can be uniquely determined by the equations in (13) by enforcing the boundary conditions $\delta\alpha_n|_{\partial M} = 0$, and $\star d\gamma_t|_{\partial M} = 0$, i.e., a nonsingular linear system by resolving the rank deficiency of Δ .

When we focus on the compact manifold in Euclidean spaces, the third term $\mathcal{H}^k(M)$ in (10) can be further decomposed into three orthogonal components [3],

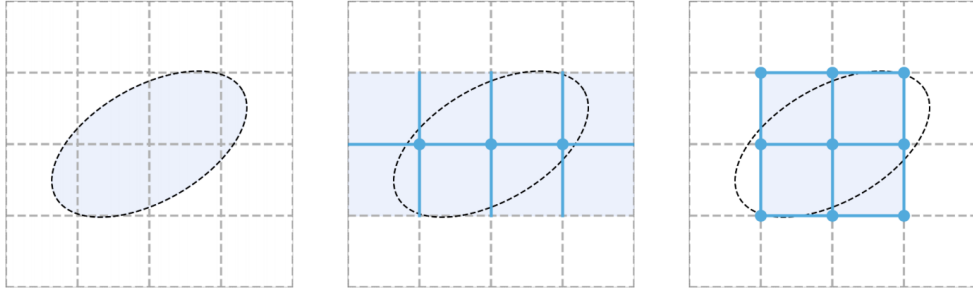
$$\mathcal{H}^k(M) = \mathcal{H}_n^k(M) \oplus \mathcal{H}_t^k(M) \oplus (d\Omega^{k-1}(M) \cap \delta\Omega^{k+1}(M)). \quad (14)$$

As a result, a five-component decomposition is obtained

$$\Omega^k(M) = d\Omega_n^{k-1}(M) \oplus \delta\Omega_t^{k+1}(M) \oplus \mathcal{H}_n^k(M) \oplus \mathcal{H}_t^k(M) \oplus (d\Omega^{k-1}(M) \cap \delta\Omega^{k+1}(M)), \quad (15)$$

where all five components are mutually orthogonal with respect to the inner product defined in (3). Supplementary Figure 1 gives an example of the five-component decomposition of a 3D pear with a tunnel model. The five components in the figure correspond to the five components in Equation 15. Particularly, the third term corresponds to a two-dimensional homology which is a void, and the forth term corresponds to a one-dimensional homology which is a loop.

1.2 Discrete Topology-preserving Hodge Theory on Cartesian Grids



Supplementary Figure 2. Illustration of discrete manifold representation for an image under normal and tangential boundary conditions. From left to right: the original image, the discrete manifold under normal boundary condition, and the discrete manifold on tangential boundary condition.

To obtain the discrete Hodge decomposition, it suffices to construct discrete versions of differential forms and differential operators, then replace the continuous forms and operators in (11) and (13) with their discrete counterparts. Here we focus on 2D/3D domains bounded by level set surfaces on Cartesian grids. The manifold is given as a sublevel set of a level set function defined on a Cartesian grid. Note that a medical image can be naturally seen as a level set function on a Cartesian grid, with its pixel values defining the scalar field. Therefore, the discrete Hodge decomposition on Cartesian grids can be directly used for medical image analysis.

1.2.1 Discrete Manifolds with Boundary

The discrete manifold M in the Cartesian grid can be given as a sublevel set of a level set function on the grid. The boundary of M is typically detected using a projection matrix. Note that in the grid representation, the boundary of M often intersects with boundary cells instead of being fully contained within them. To address this issue, we restrict computations to relevant cells by employing the inclusion or exclusion strategy proposed in [4]. For normal boundary condition, cells with at least one vertex inside M are included, while for tangential boundary condition, cells with at least one vertex of their dual cells inside M are included. The resulting set of cells is referred to as the normal support for the normal boundary condition and the tangential support for the tangential boundary condition. These supports can be seen as discrete versions of the manifolds with boundary. The projection matrices $P_{k,n}$ and $P_{k,t}$ for these boundary conditions are derived from the identity matrix by removing rows corresponding to cells outside the respective supports. Supplementary Figure 2 shows an example for the discrete manifold representation of an image under normal and tangential boundary conditions. It can be seen that the normal and tangential supports are different, and neither is a subset of the other.

1.2.2 Discrete Differential Forms

The discretization of a differential form can be achieved by the de Rham map, which maps a form to a cochain by integrating the form over cells [5]. For a Cartesian grid I_m with cells oriented according to the coordinate axes, let ω be a differential k -form on I_m , the discretization assigns to each k -cell σ_k the value $\int_{\sigma_k} \omega$, creating a cochain.

1.2.3 Discrete Exterior Derivative Operator

The discrete exterior derivative operator d on discrete k -forms can be derived by Stokes' theorem

$$\int_{\sigma} d\omega = \int_{\partial\sigma} \omega. \quad (16)$$

In matrix form, d corresponds to the transpose of the boundary matrix from $(k+1)$ -cells to k -cells. Let D_k denote the discrete exterior derivative on the entire grid, the discrete exterior derivative on the manifold M for normal and tangential boundary conditions, denoted by $D_{k,n}$ and $D_{k,t}$, are

$$D_{k,n} = P_{k+1,n} D_k P_{k,n}^T, \quad D_{k,t} = P_{k+1,t} D_k P_{k,t}^T. \quad (17)$$

We still have $D_{k+1,n} D_{k,n} = 0$ and $D_{k+1,t} D_{k,t} = 0$.

1.2.4 Discrete Hodge Star Operator

A dual grid with respect to the primal grid I_m can be constructed by treating the centers of m -cells of I_m as grid points of the dual grid. The discretization of Hodge star operator can be obtained by the following relationship in the continuous case

$$\frac{1}{|\sigma_k|} \int_{\sigma_k} \omega \approx \frac{1}{|\star\sigma_k|} \int_{\star\sigma_k} \star\omega, \quad (18)$$

where $|\sigma_k|$ is the volume of primal k -cell σ_k , $\star\sigma_k$ is the dual $(m-k)$ -cell of σ_k , and ω is a k -form. This gives a one-to-one correspondence between discrete k -forms on the primal grids and discrete $(m-k)$ -forms on its dual grids. And the correspondence leads to the discrete Hodge star as a diagonal matrix whose diagonal entries are the ratios of the volumes of dual $(m-k)$ -cells to primal k -cells. Let S_k denote the discrete Hodge star matrix on the entire grid, the discrete Hodge star on the manifold M for normal and tangential boundary conditions are $S_{k,n}$ and $S_{k,t}$ respectively as follows

$$S_{k,n} = P_{k,n} S_k P_{k,n}^T, \quad S_{k,t} = P_{k,t} S_k P_{k,t}^T. \quad (19)$$

The discrete Hodge L^2 inner product of two discrete k -forms V_k and W_k on I_m is

$$(V_k, W_k) = (V_k)^T S_k W_k. \quad (20)$$

1.2.5 Discrete Hodge Laplacian

Having the discrete Hodge star and discrete exterior derivative, the discrete codifferential can be expressed as $S_{k-1,n}^{-1} D_{k-1,n}^T S_{k,n}$ and $S_{k-1,t}^{-1} D_{k-1,t}^T S_{k,t}$ for normal and tangential boundary conditions respectively. Using these discrete differential and codifferential operators, the Laplacian $\Delta = \delta d + d\delta$ is not symmetric. Therefore the symmetric $\star\Delta$ matrix as used to define the discrete Laplacian. The discrete Laplacian for normal and tangential boundary conditions $L_{k,n}$ and $L_{k,t}$ are respectively defined as follows

$$\begin{aligned} L_{k,n} &= D_{k,n}^T S_{k+1,n} D_{k,n} + S_{k,n} D_{k-1,n} S_{k-1,n}^{-1} D_{k-1,n}^T S_{k,n}, \\ L_{k,t} &= D_{k,t}^T S_{k+1,t} D_{k,t} + S_{k,t} D_{k-1,t} S_{k-1,t}^{-1} D_{k-1,t}^T S_{k,t}. \end{aligned} \quad (21)$$

As in the continuous case, the Kernels of these discrete Laplacians are fully determined by the topology of M . Specifically, the dimension of $\ker L_{k,n}$ equals the Betti number β_{m-k} , while the dimension of $\ker L_{k,t}$ equals β_k .

When the Hodge star matrix is replaced by the identity matrix, the discrete Laplacians reduce to the Boundary-Induced Graph (BIG) Laplacians,

$$\begin{aligned} L_{k,n}^B &= D_{k,n}^T D_{k,n} + D_{k-1,n} D_{k-1,n}^T, \\ L_{k,t}^B &= D_{k,t}^T D_{k,t} + D_{k-1,t} D_{k-1,t}^T. \end{aligned} \quad (22)$$

The BIG Laplacians preserve the differential calculus properties of the Hodge Laplacian while retaining the combinatorial nature of the discrete Laplacian [4].

Note that the discrete Hodge Laplacian differs from the combinatorial Laplacian. For instance, when performing the spectral decomposition of a vector field on a point cloud, the use of combinatorial Laplacians defined on commonly employed simplicial complexes does not yield the same curl-free and divergence-free components as those obtained through the spectral decomposition of a vector field using discretized Hodge Laplacians. The latter are defined either on a point cloud with a boundary in the Eulerian representation or on a regular mesh in the Eulerian representation. A detailed comparison between the Hodge Laplacian and the combinatorial Laplacian can be found in the referenced literature [4].

1.2.6 Discrete Hodge Decomposition

With the discrete version of differential forms and differential operators, the discrete Hodge decomposition can be expressed as

$$V^k = D_{k-1,n} W_n + S_{k,t}^{-1} D_{k,t}^T S_{k+1,t} W_t + E, \quad (23)$$

where V^k , W_n , W_t , and E are the discrete version of ω , α_n , β_t , and η in (11) respectively. As in the continuous case, we can first find W_n and W_t , then compute E as $E = V^k - D_{k-1,n} W_n - S_{k,t}^{-1} D_{k,t}^T S_{k+1,t} W_t$. And W_n and W_t can be uniquely determined by the discrete version of equation (13)

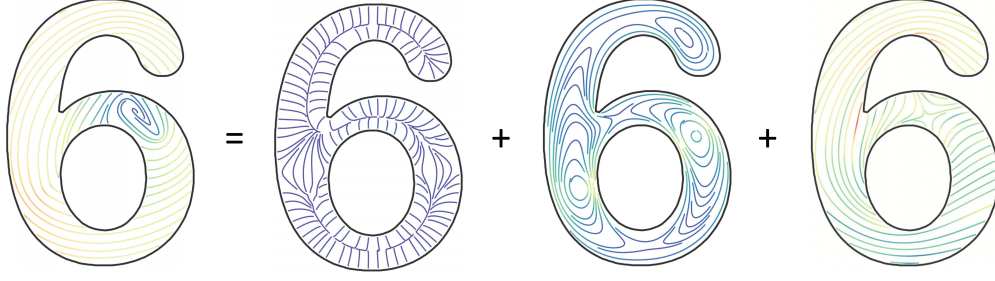
$$L_{k-1,n} W_n = D_{k-1,n}^T S_{k,n} V_n^k, \quad L_{k+1,t} W_t = S_{k+1,t} D_{k,t} V_t^k. \quad (24)$$

where V_n^k and V_t^k are the vectors of V^k under normal and tangential supports respectively. Supplementary Figure 3 gives an example for the three-component Hodge decomposition on a 2D 6 domain.

1.3 Numerical Analysis of the Orthogonality of the Decomposition

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

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



Supplementary Figure 3. Illustration of the three-component Hodge decomposition on a 2D 6 domain. From left to right: original vector field, the curl-free field, the divergence-free field, and the harmonic field.

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

$$(\alpha, \beta) = \frac{|\langle \alpha, \beta \rangle|}{\|\alpha\|_2 \|\beta\|_2}.$$

We report the mean and standard deviation of these values across all samples for each dataset, which is shown in Supplementary Table 1. 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. Therefore, practical applications usually focus on numerical orthogonality instead of strict theoretical orthogonality. Since our numerical errors 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.

Supplementary Table 1. 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

1.4 Algorithm

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 naturally constructed.

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.

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 $P_{k,b}$

1.5 Example

Consider the manifold in Supplementary Figure 4, it is a 2d disk in the Cartesian grid. The resulting 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

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

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 $S'_{k,n}(i, i) = \frac{\text{volume}(\star\sigma)}{\text{volume}(\sigma \cap M)}$;

 Set $S'_{k,t}(i, i) = \frac{\text{volume}(\star\sigma) \cap M}{\text{volume}(\sigma)}$;

return $(S'_{k,n}, S'_{k,t})$

- 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 $L_{0,t} = D_{0,t}^TS_{1,t}D_{0,t}$. We give the details as follows:

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

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 $D_{0,n} = P_{1,n}D_0P_{0,n}^T$, $D_{1,t} = P_{2,t}D_1P_{1,t}^T$;

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

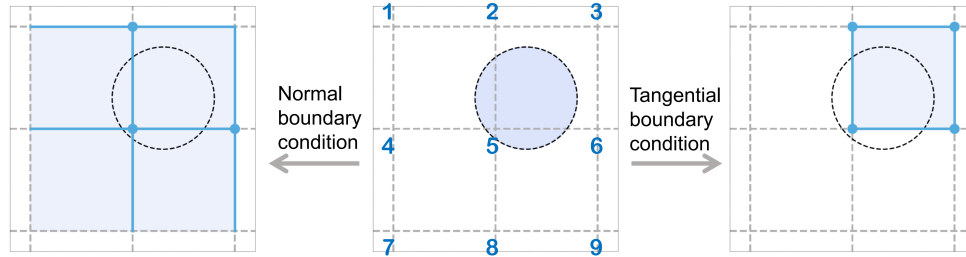
Solve $L_{0,n}\alpha_n = \delta_{1,n}\omega$ to get α_n ;

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



Supplementary Figure 4. A 2d disk and its discrete manifold representation under normal and tangential boundary conditions.

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

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

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

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

$$S'_{1,t} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.24 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.64 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.64 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.24 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad (25)$$

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

$$S_{1,t} = P_{1,t} S'_{1,t} P_{1,t}^T = \begin{pmatrix} 0.24 & 0 & 0 & 0 \\ 0 & 0.64 & 0 & 0 \\ 0 & 0 & 0.64 & 0 \\ 0 & 0 & 0 & 0.24 \end{pmatrix} \quad (26)$$

6. The Hodge Laplacian can be computed as

$$L_{0,t} = D_{0,t}^T S_{1,t} D_{0,t} = \begin{pmatrix} 0.24 & -0.24 & 0 & 0 \\ -0.24 & 0.48 & 0 & 0 \\ 0 & 0 & 0.64 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad (27)$$

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 $L_{0,t}^B$. The BIG Laplacian $L_{0,t}^B$ also has exactly one zero eigenvalue since it preserves the topology of the manifold.

1.6 Limitations

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.

In principle, the vector field generation method used in MTDL is translation- and rotation-equivariant. However, computationally, it is not rotationally equivariant because the underlying Cartesian grid is not rotationally symmetric. However, this imperfection occurs on the order of grid spacing and diminishes as the grid size goes to zero. Additionally, MTDL is not scale-equivariant, since the grid spacing does not vary under geometric scaling. 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 computationally further extended to incorporate additional forms of equivariance or invariance, such as rotational or scale equivariance.

2 Vector Field Generation

In this section, we introduce several methods for generating noise-resilient vector fields (1-forms) from images. Without loss of generality, we focus on 3D images, and these methods can be easily adapted to 2D images.

2.1 Gradient-based Method

The gradient-based method uses the discrete gradient operation to construct a vector field from the images. Formally, for a 3D image $\mathcal{I}$, where $\mathcal{I}(i, j, k)$ represents the pixel value at position (i, j, k) , a vector $(x_{i,j,k}, y_{i,j,k}, z_{i,j,k})$ is constructed for each pixel at (i, j, k) as follows

$$\begin{aligned} x_{i,j,k} &= \frac{\mathcal{I}(i+s, j, k) - \mathcal{I}(i-t, j, k)}{2} \\ y_{i,j,k} &= \frac{\mathcal{I}(i, j+s, k) - \mathcal{I}(i, j-t, k)}{2} \\ z_{i,j,k} &= \frac{\mathcal{I}(i, j, k+s) - \mathcal{I}(i, j, k-t)}{2} \end{aligned} \tag{28}$$

where s, t are parameters to control the forward and backward steps. The resulting vector field is called the gradient-based (s, t) -step vector field. This method is a generalization of the standard finite difference method for computing gradient. It allows different forward and backward steps for computing the difference of a point. The method described in section “Methods” is a special case of this gradient-based approach, corresponding to the standard case where $s = t = 1$.

2.2 Flow-based Method

The flow-based method constructs the vector field by analyzing the flow of pixel values, akin to unwind scheme. Formally, for a 3D image $\mathcal{I}$, where $\mathcal{I}(i, j, k)$ represents the pixel value at position (i, j, k) , a vector $(x_{i,j,k}, y_{i,j,k}, z_{i,j,k})$ for the pixel at (i, j, k) is constructed by the following steps

1. For the 26 voxel values adjacent to position (i, j, k) , let S be the set of positions whose pixel values are smaller than $\mathcal{I}(i, j, k)$.

2. If $S = \emptyset$, set $(x_{i,j,k}, y_{i,j,k}, z_{i,j,k}) = (0, 0, 0)$
3. If $S \neq \emptyset$, Identify the positions in S with the smallest pixel value. If there is a unique position, let $(x_{i,j,k}, y_{i,j,k}, z_{i,j,k})$ be the vector pointing from (i, j, k) to this position, with magnitude $\mathcal{I}(i, j, k)$. If multiple positions have the same smallest value, compute the average direction from (i, j, k) to these positions and assign the resulting vector a magnitude of $\mathcal{I}(i, j, k)$.

This vector field captures the flow of pixel values from regions of higher intensity to those of lower intensity. Alternatively, one can construct a vector field that represents the reverse flow, from lower-intensity to higher-intensity regions.

2.3 Other Methods

For color images, vector fields can be derived by selecting pairs of color channels. For instance, a (r, g, b) -channel image can yield three distinct vector fields based on the (r, g) , (r, b) and (g, b) channel pairs.

For images with large dimensions, the image can be divided into smaller patches, and vector fields can be computed for these patches. For example, consider an image of size 1024×1024 . By dividing it into 16×16 patches, a new 64×64 image can be formed, where each “pixel” represents a patch. Topological indices of the patches, such as (β_0, β_1) , can then be used to define vectors for the corresponding pixels in the new image.

2.4 Evaluation of Different Methods

We conducted a comparative analysis of three distinct vector field generation 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 Supplementary Table 2. 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 be further explored in the future study.

Supplementary Table 2. 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

3 Dataset Details

The MedMNIST v2 dataset is a standardized, MNIST-like collection of biomedical images. All images are preprocessed into a uniform size and labeled, eliminating the need for domain knowledge from users. The dataset includes twelve 2D datasets and six 3D datasets, covering a range of data modalities, data scales, and task types. In total, it includes 708,069 2D images and 9,998 3D images, with standard train-validation-test

splits provided for all datasets. The detailed data scale, data modality, and task type information for each dataset are shown in Supplementary Table 3. The available data resolutions are 28×28 , 64×64 , 128×128 ,

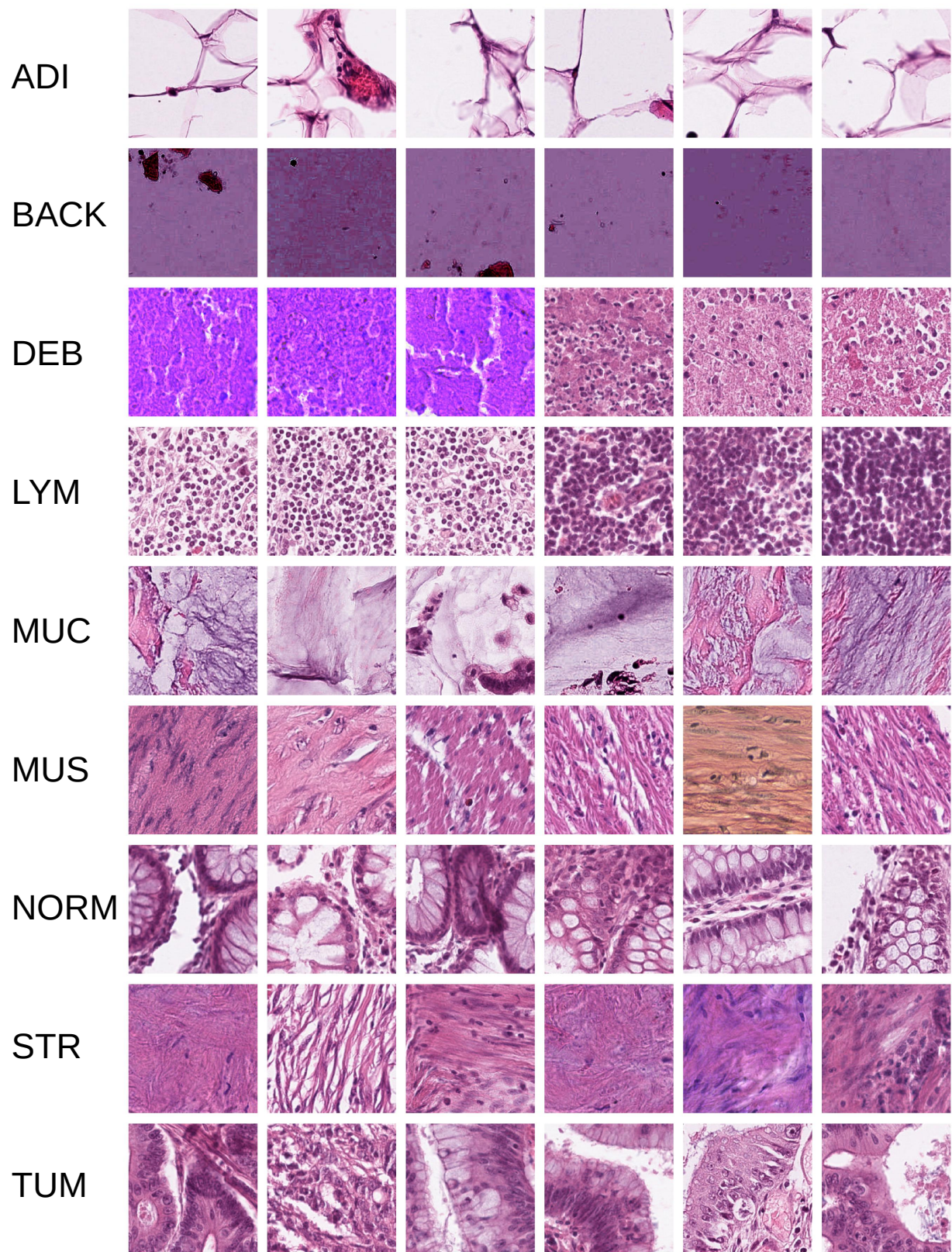
Supplementary Table 3. Overview of MedMNIST v2 dataset

MedMNIST2D	Data Modality	Task (# Classes / Labels)	# Samples	# Training / Validation / Test
PathMNIST	Colon Pathology	Multi-Class (9)	107,180	89,996 / 10,004 / 7,180
ChestMNIST	Chest X-Ray	Multi-Label (14) Binary-Class (2)	112,120	78,468 / 11,219 / 22,433
DermaMNIST	Dermatoscope	Multi-Class (7)	10,015	7,007 / 1,003 / 2,005
OCTMNIST	Retinal OCT	Multi-Class (4)	109,309	97,477 / 10,832 / 1,000
PneumoniaMNIST	Chest X-Ray	Binary-Class (2)	5,856	4,708 / 524 / 624
RetinaMNIST	Fundus Camera	Ordinal Regression (5)	1,600	1,080 / 120 / 400
BloodMNIST	Blood Cell Microscope	Multi-Class (8)	17,092	11,959 / 1,712 / 3,421
TissueMNIST	Kidney Cortex Microscope	Multi-Class (8)	236,386	165,466 / 23,640 / 47,280
OrganAMNIST	Abdominal CT	Multi-Class (11)	58,850	34,581 / 6,491 / 17,778
OrganCMNIST	Abdominal CT	Multi-Class (11)	23,660	13,000 / 2,392 / 8,268
OrganSMNIST	Abdominal CT	Multi-Class (11)	25,221	13,940 / 2,452 / 8,829
MedMNIST3D	Data Modality	Task (# Classes / Labels)	# Samples	# Training / Validation / Test
OrganMNIST3D	Abdominal CT	Multi-Class (11)	1,742	971 / 161 / 610
NoduleMNIST3D	Chest CT	Binary-Class (2)	1,633	1,158 / 165 / 310
AdrenalMNIST3D	Shape from Abdominal CT	Binary-Class (2)	1,584	1,188 / 98 / 298
FractureMNIST3D	Chest CT	Multi-Class (3)	1,370	1,027 / 103 / 240
VesselMNIST3D	Shape from Brain MRA	Binary-Class (2)	1,908	1,335 / 191 / 382
SynapseMNIST3D	Electron Microscope	Binary-Class (2)	1,759	1,230 / 177 / 352

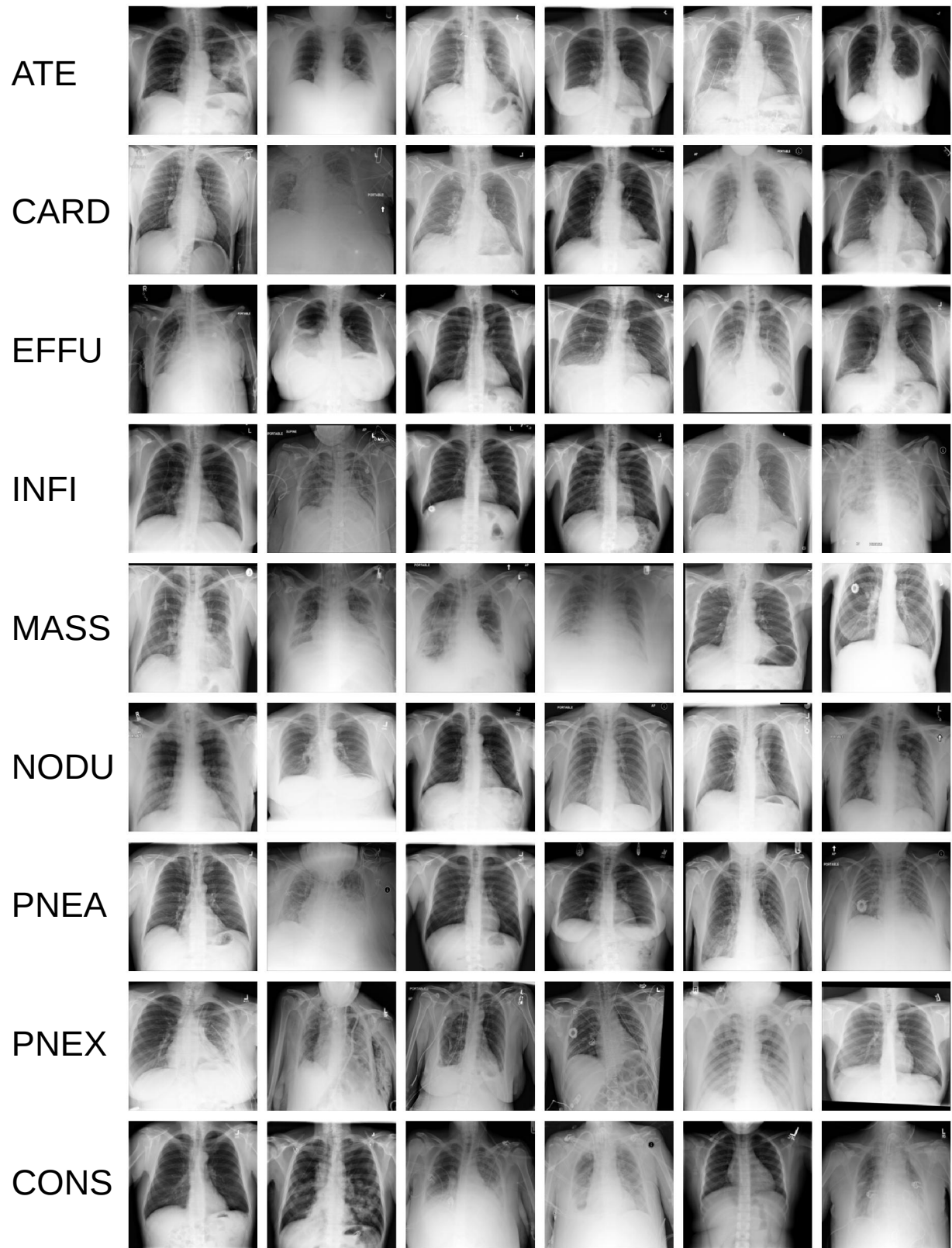
224×224 for 2D datasets, and $28\times 28\times 28$, $64\times 64\times 64$ for 3D datasets. Here we also give a brief introduction of all the 17 datasets. The example illustration of 2D and 3D datasets are shown in Fig. ?? and Fig. ?? respectively.

3.1 2D Datasets

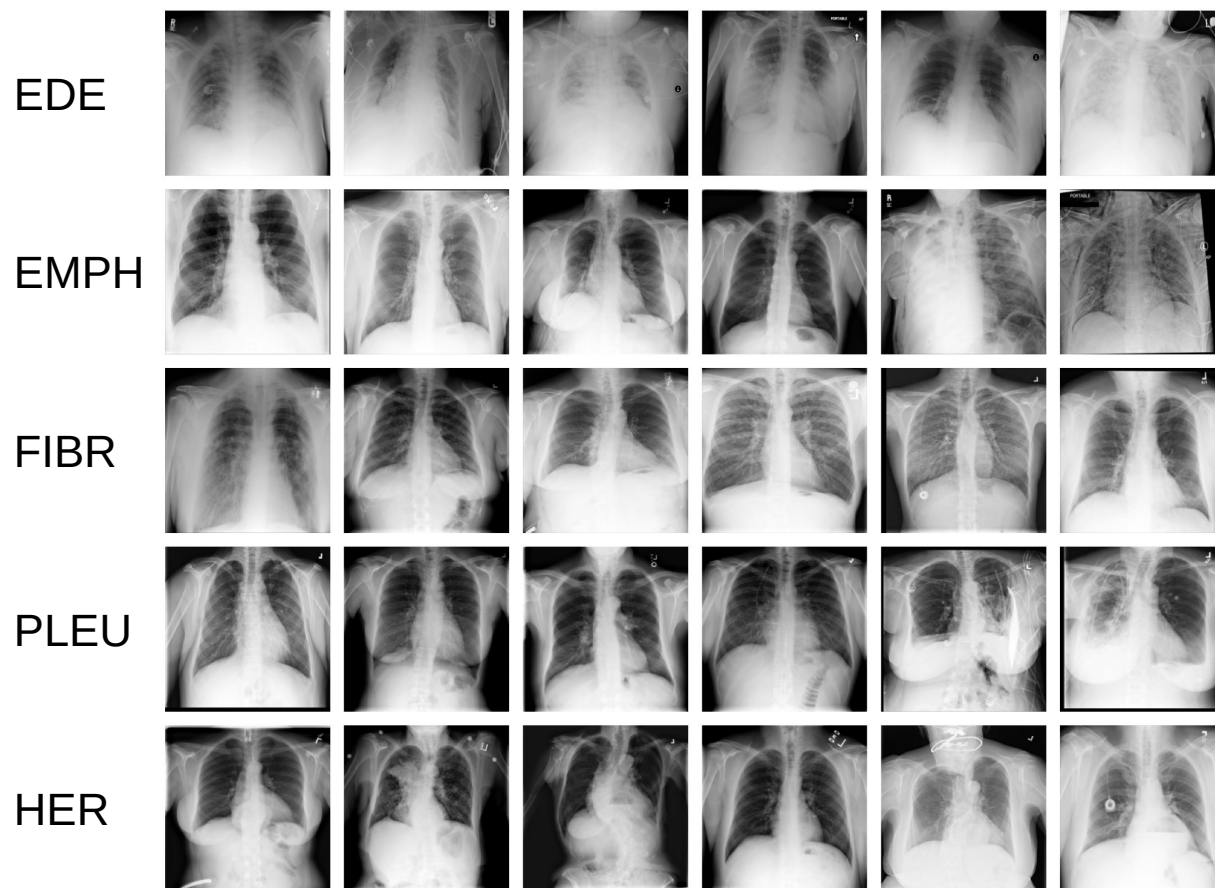
- PathMNIST: the PathMNIST is based on a prior study [6, 7] for predicting survival from colorectal cancer histology slides. The dataset is comprised of 9 types of tissues, resulting in a multi-class classification task. The labels are adipose, background, debris, lymphocytes, mucus, smooth muscle, normal colon mucosa, cancer-associated stroma, and colorectal adenocarcinoma epithelium. These images were manually extracted from N=86 H&E stained human cancer tissue slides from formalin-fixed paraffin-embedded (FFPE) samples from the NCT Biobank (National Center for Tumor Diseases, Heidelberg, Germany) and the UMM pathology archive (University Medical Center Mannheim, Mannheim, Germany). Example images are shown in Supplementary Figure 5.
- ChestMNIST: the ChestMNIST is based on the NIH-ChestXray14 dataset [8], consisting of frontal-view X-Ray images of 30,805 unique patients with the text-mined 14 disease labels, leads to a multi-label binary-class classification task. The labels are atelectasis, cardiomegaly, effusion, infiltration, mass, nodule, pneumonia, pneumothorax, consolidation, edema, emphysema, fibrosis, pleural, and hernia. Example images are shown in Supplementary Figure 6 and Supplementary Figure 7.
- DermaMNIST: the DermaMNIST is based on the HAM10000 dataset [9, 10], a collection of multi-source dermatoscopic images of common pigmented skin lesions, consisting of 7 diseases, formalizing as a multi-class classification task. The labels are actinic keratoses and intraepithelial carcinoma, basal cell carcinoma, benign keratosis-like lesions, dermatofibroma, melanoma, melanocytic nevi, and vascular lesions.
- OCTMNIST: the OCTMNIST is from a dataset [11] of valid optical coherence tomography (OCT) images for retinal diseases, consisting of 4 categories, leading to a multi-class classification task. The labels are choroidal neovascularization, diabetic macular edema, drusen, and normal. Example images are shown in Supplementary Figure 8.



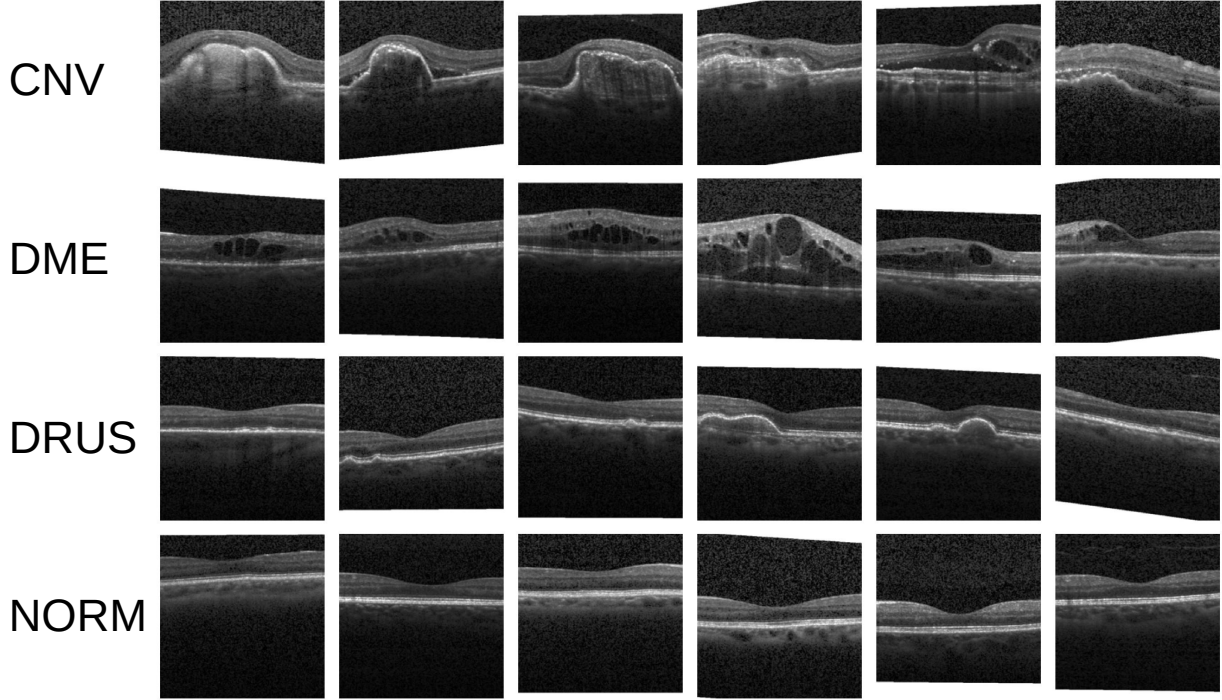
Supplementary Figure 5. Example images of the nine classes in the PathMNIST dataset. ADI: adipose tissue; BACK: background; DEB: debris; LYM: lymphocytes; MUC: mucus; MUS: smooth muscle; NORM: normal colon mucosa; STR: cancer-associated stroma; TUM: colorectal adenocarcinoma epithelium.



Supplementary Figure 6. Example images of the first nine classes in the ChestMNIST dataset. ATE: atelectasis; CARD: cardiomegaly; EFFU: effusion; INFI: infiltration; MASS: mass; NODU: nodule; PNEA: pneumonia; PNEX: pneumothorax; CONS: consolidation.

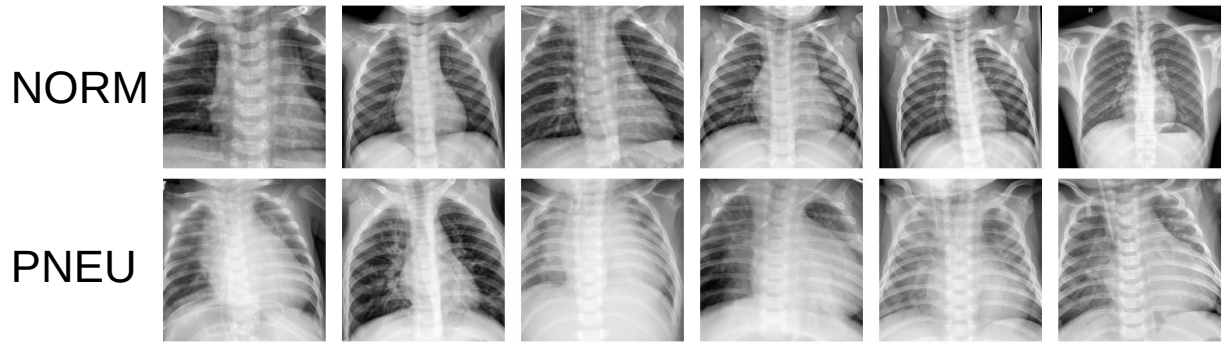


Supplementary Figure 7. Example images of the rest five classes in the ChestMNIST dataset. EDE: edema; EMPH: emphysema; FIBR: fibrosis; PLEU: pleural; HER: hernia.

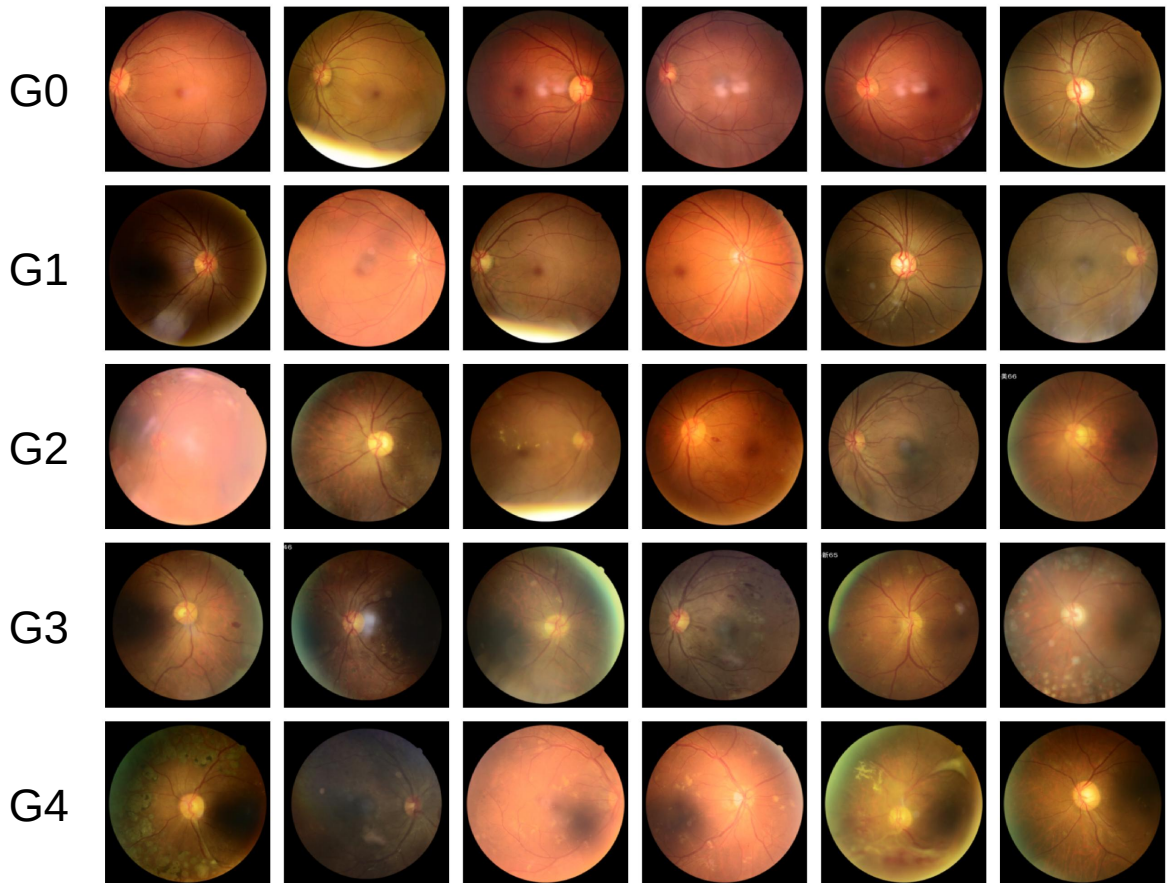


Supplementary Figure 8. Example images of the four classes in the OCTMNIST dataset. CNV: choroidal neovascularization; DME: diabetic macular edema; DRUS: drusen; NORM: normal.

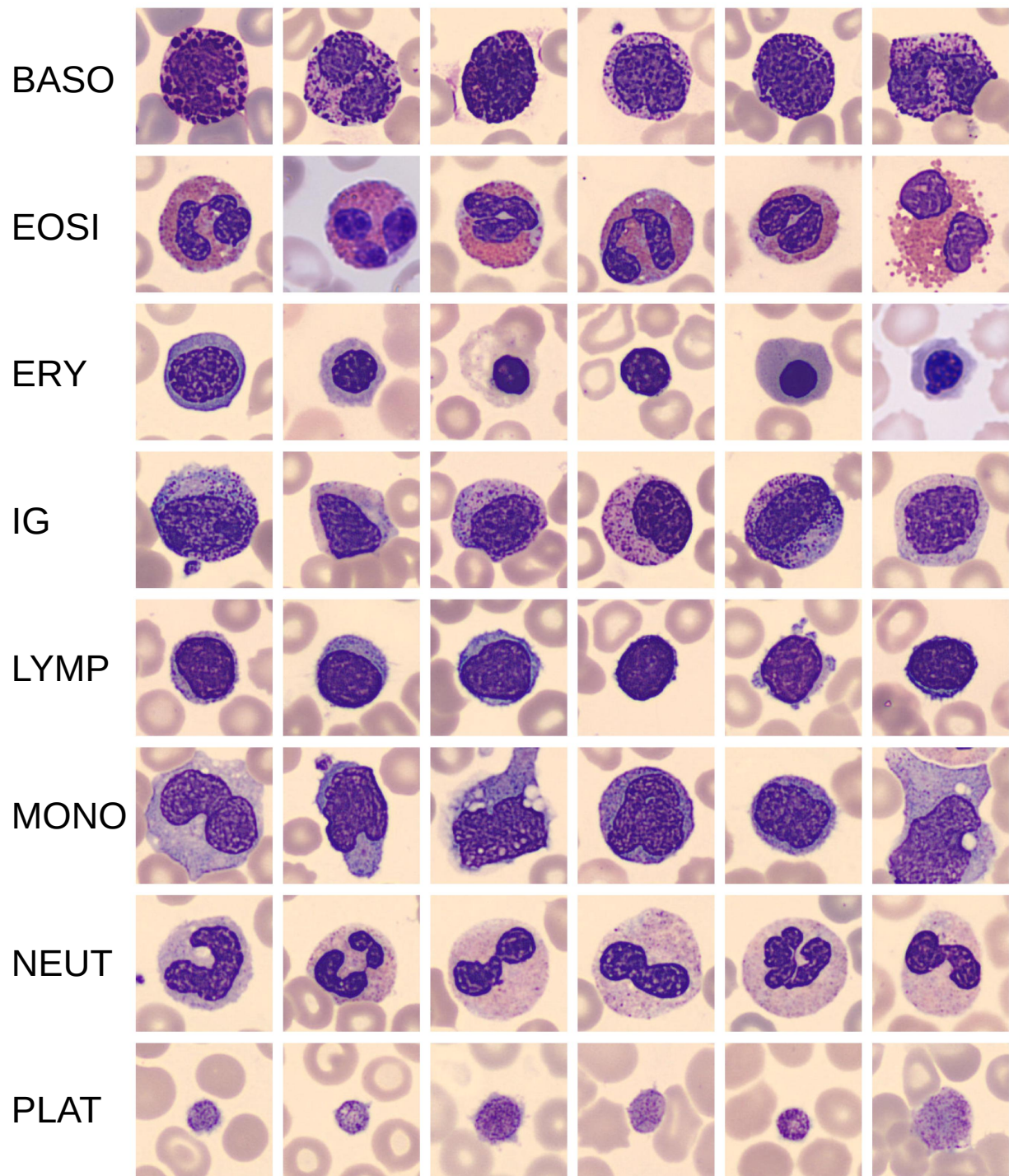
- **PneumoniaMNIST:** the PneumoniaMNIST is from a dataset [12] of 5856 pediatric chest X-Ray images, the task is binary-class classification of pneumonia against normal. These images were selected from retrospective cohorts of pediatric patients of one to five years old from Guangzhou Women and Children’s Medical Center, Guangzhou. All chest X-ray imaging was performed as part of patients’ routine clinical care. Example images are shown in Supplementary Figure 9.
- **RetinaMNIST:** the RetinaMNIST is based on the DeepDRiD24 challenge [13], which provides a collection of 1600 retina fundus images. The task is ordinal regression of 5-level grading of diabetic retinopathy severity. An internationally accepted method of grading the DR levels classifies DR into non-proliferative DR (NPDR) and proliferative DR (PDR) [14]. NPDR is the early stage of DR and is characterized by the presence of microaneurysms, whereas PDR is an advanced stage of DR and can lead to severe vision loss. Example images are shown in Supplementary Figure 10.
- **BloodMNIST:** the BloodMNIST is based on a dataset of individual normal cells [15], captured from individuals without infection, hematologic or oncologic disease and free of any pharmacologic treatment at the moment of blood collection, consisting of 8 classes. The labels are basophil, eosinophil, erythroblast, immature granulocytes(myelocytes, metamyelocytes and promyelocytes), lymphocyte, monocyte, neutrophil, and platelet. The images were acquired using the analyzer CellaVision DM96 in the Core Laboratory at the Hospital Clinic of Barcelona. Example images are shown in Supplementary Figure 11.
- **TissueMNIST:** the TissueMNIST is from the BBBC051 [16], available from the Broad Bioimage Benchmark Collection [17]. It contains human kidney cortex cells, segmented from 3 reference tissue specimens and organized into 8 categories. The labels are Collecting Duct, Connecting Tubule, Distal Convoluted Tubule, Glomerular endothelial cells, Interstitial endothelial cells, Leukocytes, Podocytes,



Supplementary Figure 9. Example images of the two classes in the PneumoniaMNIST dataset. NORM: normal; PNEU: pneumonia.



Supplementary Figure 10. Example images of the five classes in the RetinaMNIST dataset. G0: no apparent retinopathy; G1: mild NPDR; G2: moderate NPDR; G3: severe NPDR; G4: PDR.



Supplementary Figure 11. Example images of the eight classes in the BloodMNIST dataset. BASO: basophil; EOSI: eosinophil; ERY: erythroblast; IG: immature granulocytes (myelocytes, metamyelocytes and promyelocytes); LYMP: lymphocyte; MONO: monocyte; NEUT: neutrophil; PLAT: platelet.

Proximal Tubule Segments, and Thick Ascending Limb. Example images are shown in Supplementary Figure 12.

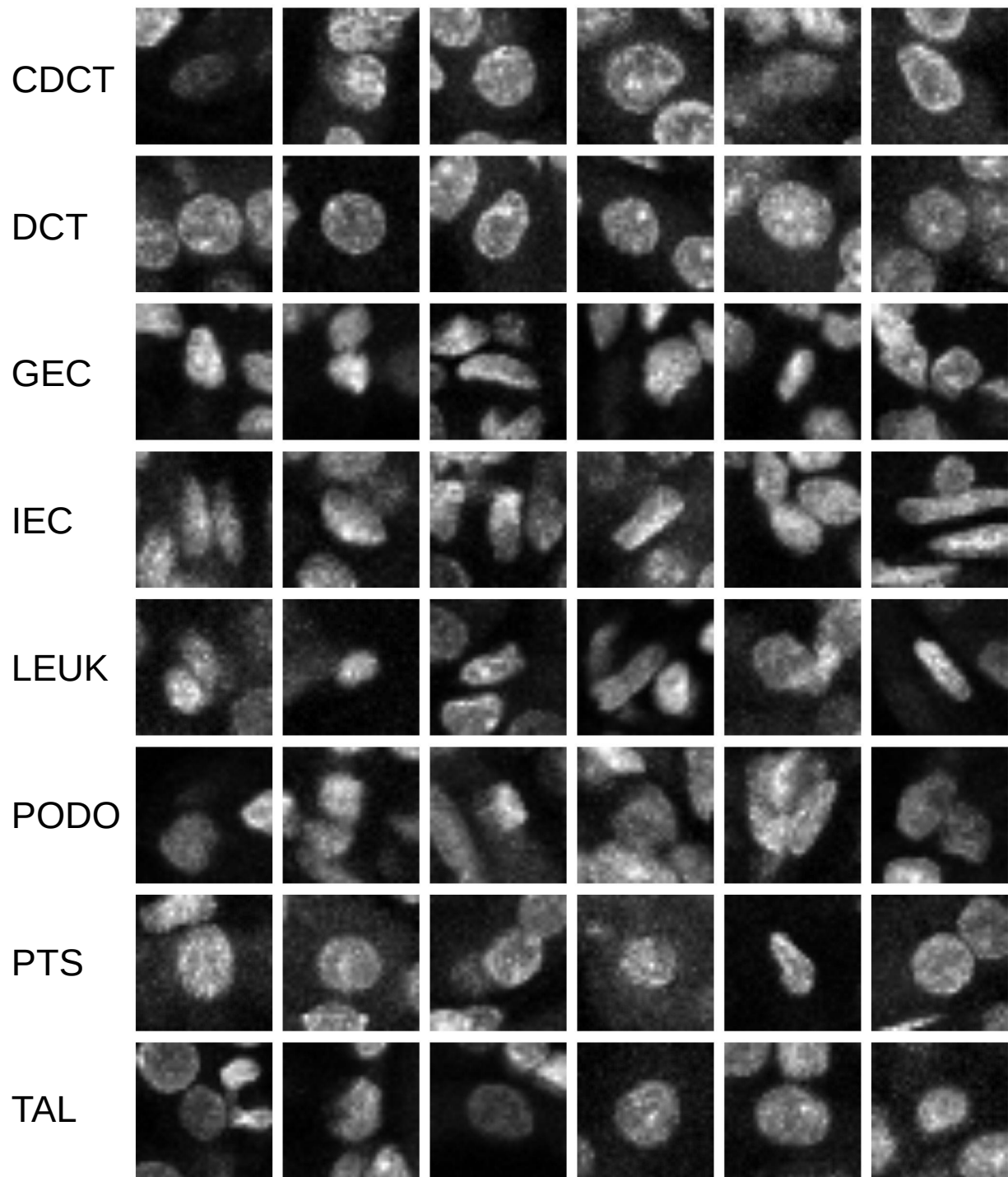
- **Organ{A,C,S}MNIST**: the **Organ{A,C,S}** datasets is based on the 3D computed tomography (CT) images from Liver Tumor Segmentation Benchmark [18], they are from the center slices of the 3D bounding boxes in axial/coronal/sagittal views respectively. The tasks are all multi-class classification with 11-classes. The labels are bladder, femur-left, femur-right, heart, kidney-left, kidney-right, liver, lung-left, lung-right, pancreas, and spleen. Example images for OrganAMNIST are shown in Supplementary Figure 13 and 14, for OrganCMNIST are shown in Supplementary Figure 15 and Fig. 16, and for OrganSMNIST are shown in Supplementary Figure 17 and 18.

3.2 3D Datasets

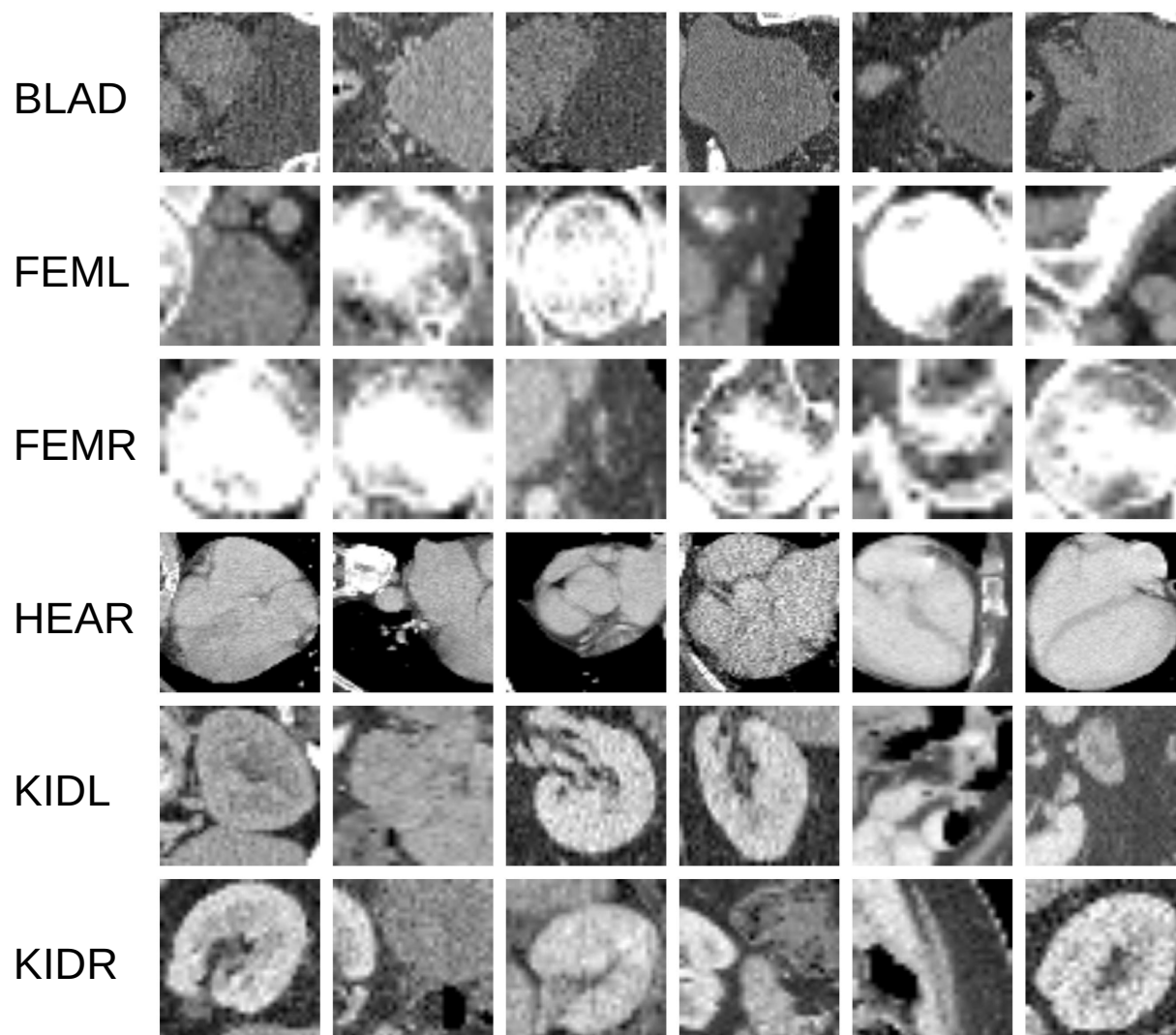
- **OrganMNIST3D**: the **OrganMNIST3D** is from the same source with 2D Organ datasets. The tasks are multi-class classification with 11-classes. The labels are bladder, femur-left, femur-right, heart, kidney-left, kidney-right, liver, lung-left, lung-right, pancreas, and spleen. Example images are shown in Supplementary Figure 19 and 20.
- **NoduleMNIST3D**: the **NoduleMNIST3D** is based on the LIDC-IDRI dataset [19] that contains images from thoracic CT scans. It contains 1018 cases, each of which includes images from a clinical thoracic CT scan and an associated XML file that records the results of a two-phase image annotation process performed by four experienced thoracic radiologists. The task is binary classification of benign against malignant. Example images are shown in Supplementary Figure 21.
- **AdrenalMNIST3D**: the **AdrenalMNIST3D** dataset consists of shape masks from 1,584 left and right adrenal glands, collected from Zhongshan Hospital Affiliated to Fudan University, each 3D shape of adrenal gland is annotated by an expert endocrinologist using abdominal computed tomography (CT), together with a binary classification label of normal adrenal gland or adrenal mass. The task is binary classification of normal against mass. Example images are shown in Supplementary Figure 22.
- **FractureMNIST3D**: the **FractureMNIST3D** is based on the RibFrac Dataset3 [20], containing around 5,000 rib fractures from 660 computed tomography (CT) scans, which were annotated with a human-in-the-loop labeling procedure. The task is a 3-class classification. The labels are buckle rib fracture, nondisplaced rib fracture, and displaced rib fracture. Example images are shown in Supplementary Figure 23.
- **VesselMNIST3D**: the **VesselMNIST3D** is based on an open-access 3D intracranial aneurysm dataset, IntraA3 [21], containing 103 3D models (meshes) of entire brain vessels collected by reconstructing MRA images. The task is a binary classification of vessel and aneurysm. Example images are shown in Supplementary Figure 24.
- **SynapseMNIST3D**: the **SynapseMNIST3D** dataset is to classify whether a synapse is excitatory or inhibitory. It uses a 3D image volume of an adult rat acquired by a multi-beam scanning electron microscope. Three neuroscience experts segment a pyramidal neuron within the whole volume and proofread all the synapses on this neuron with excitatory/inhibitory labels. Example images are shown in Supplementary Figure 25.

4 Data Separation by Task Type, Data Scale, and Data Modality

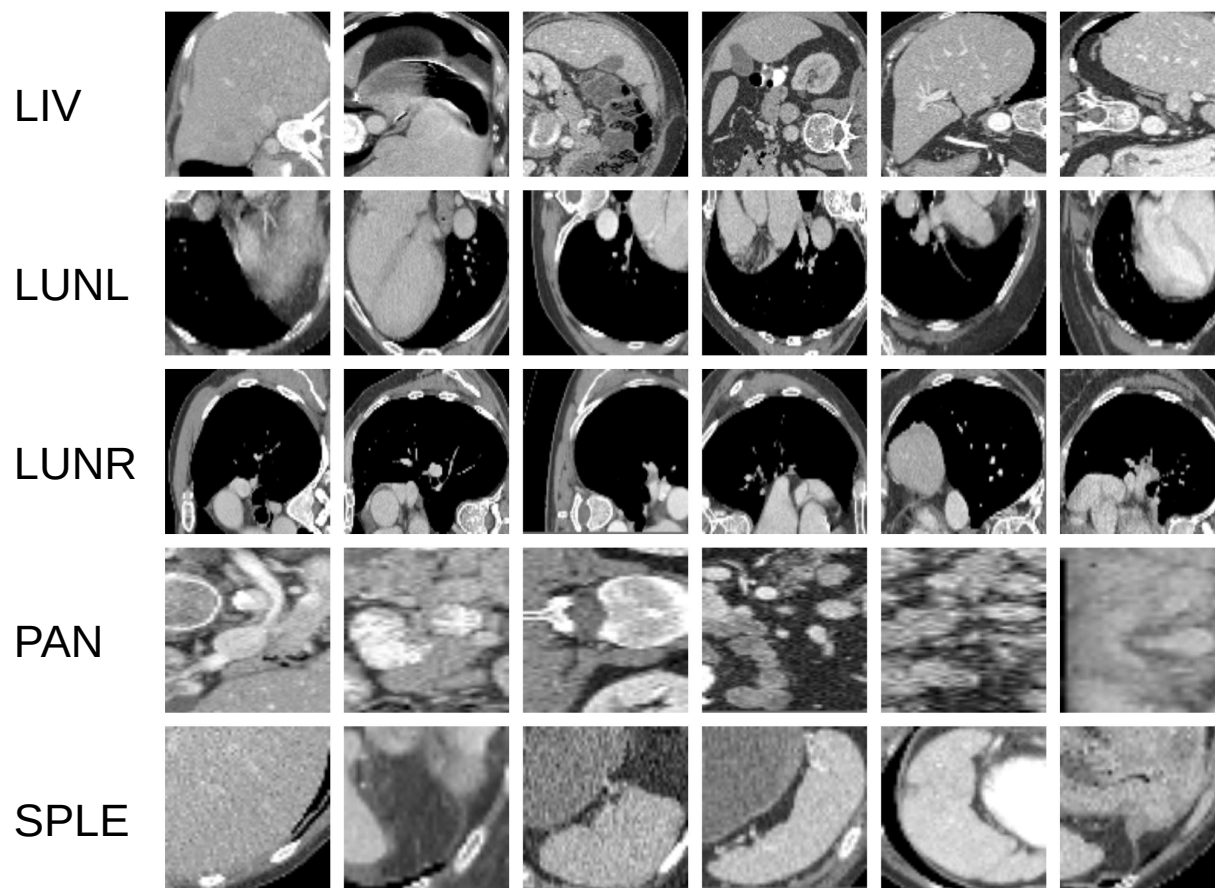
Here we give the detailed breakdown of the 17 datasets into different groups based on data modality, data scale, and task type.



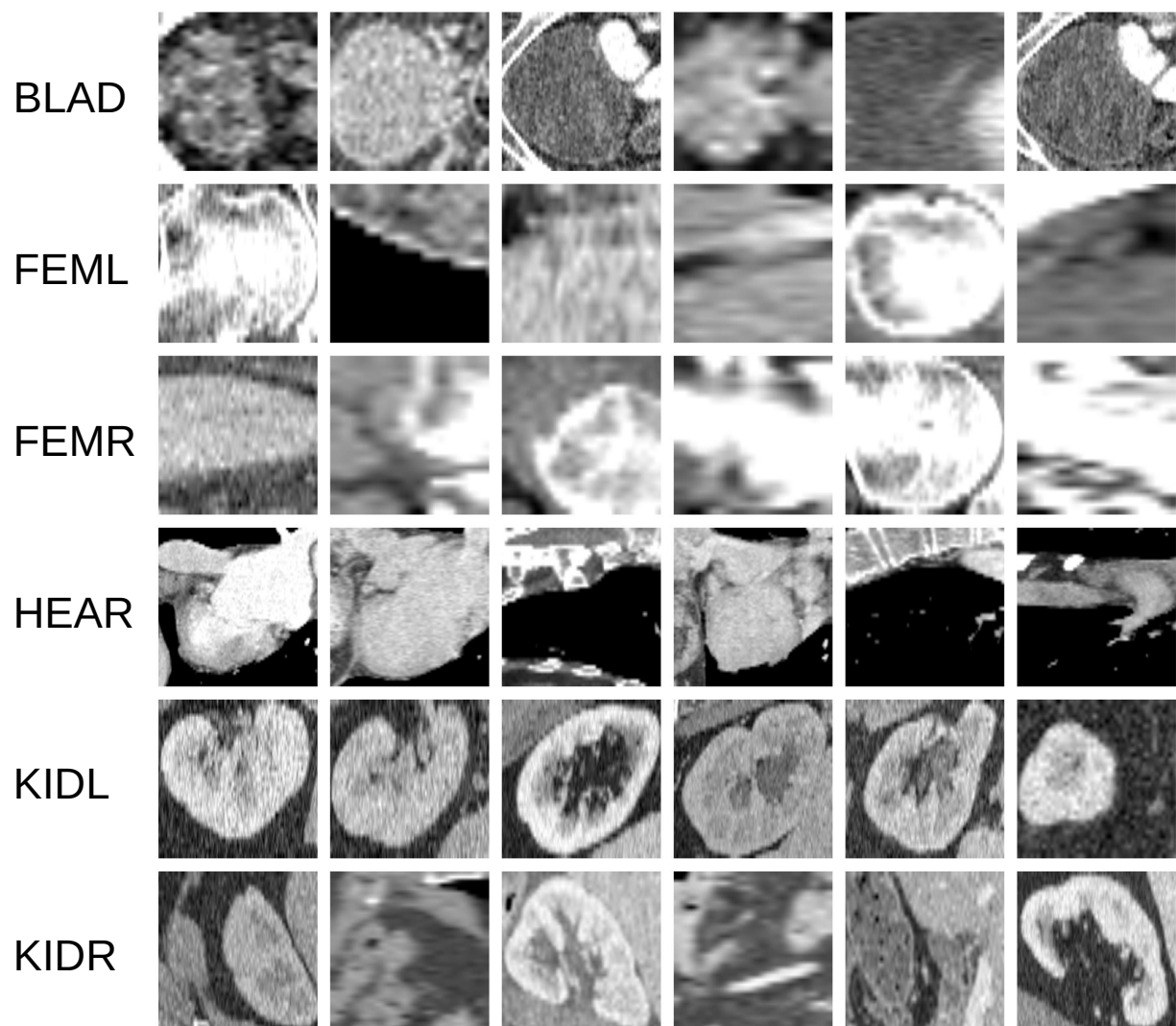
Supplementary Figure 12. Example images of the eight classes in the TissueMNIST dataset. CDCT: Collecting Duct, Connecting Tubule; DCT: Distal Convoluted Tubule; GEC: Glomerular Endothelial Cells; IEC: Interstitial Endothelial Cells; LEUK: Leukocytes; PODO: Podocytes; PTS: Proximal Tubule Segments; TAL: Thick Ascending Limb.



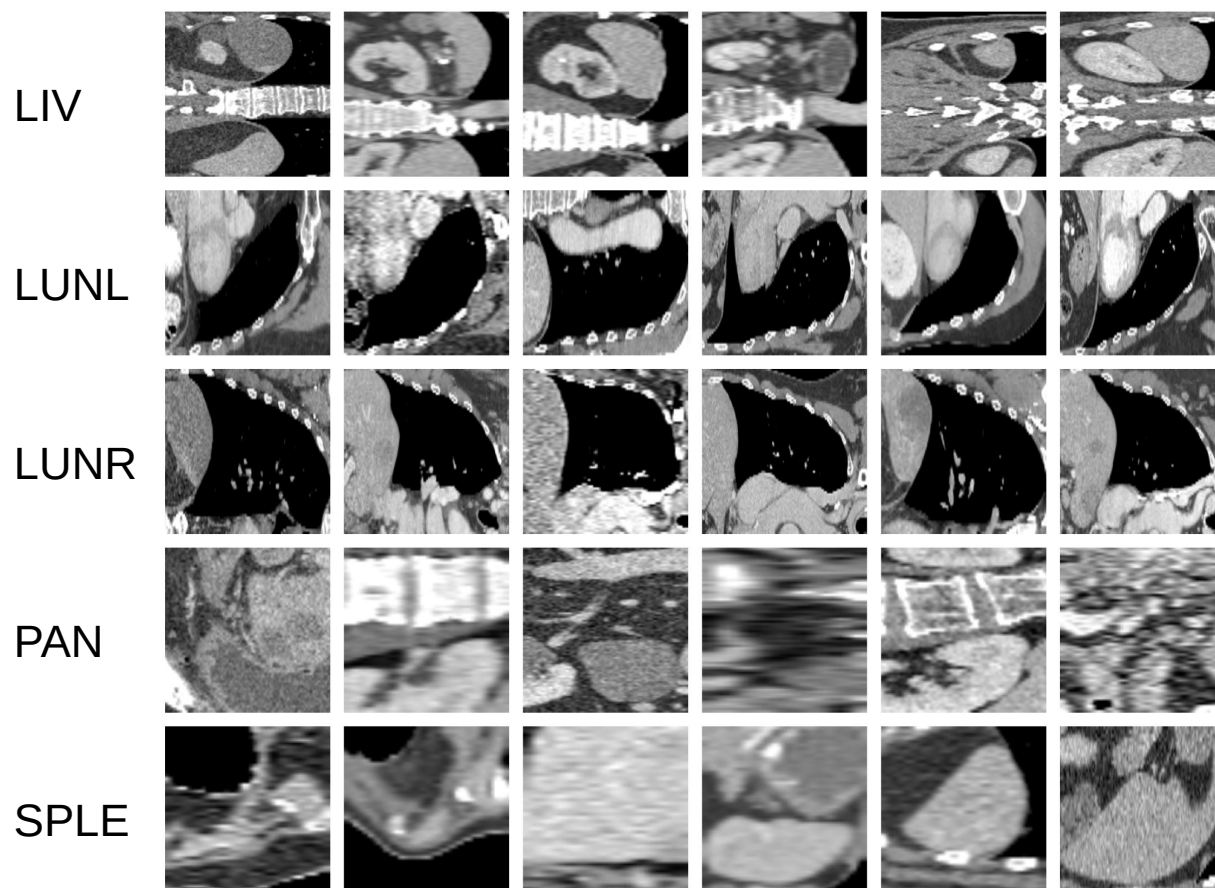
Supplementary Figure 13. Example images of the first six classes in the OrganAMNIST dataset. BLAD: bladder; FEML: femur-left; FEMR: femur-right; HEAR: heart; KIDL: kidney-left; KIDR: kidney-right.



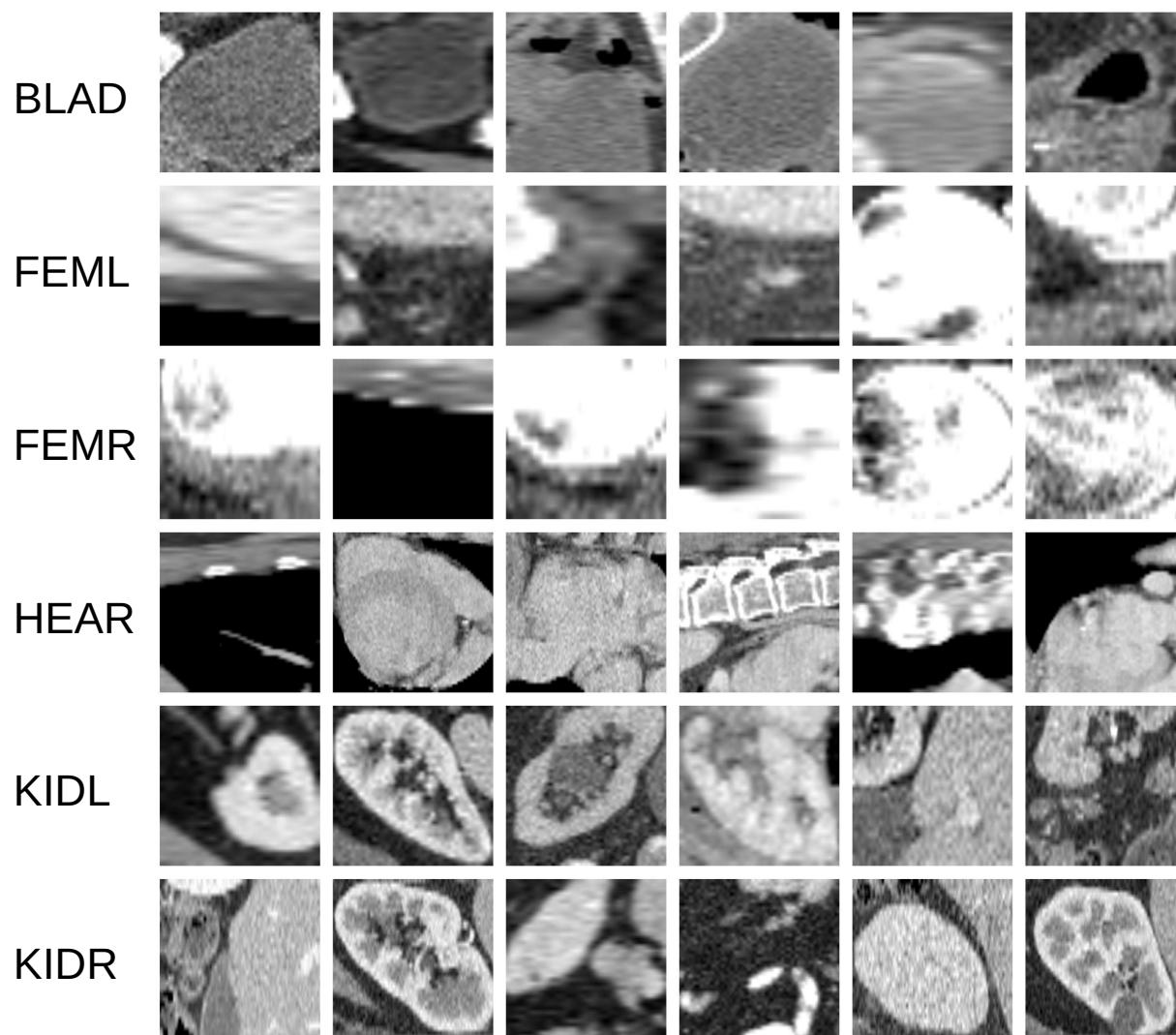
Supplementary Figure 14. Example images of the rest five classes in the OrganAMNIST dataset. LIV: liver; LUNL: lung-left; LUNR: lung-right; PAN: pancreas; SPLE: spleen.



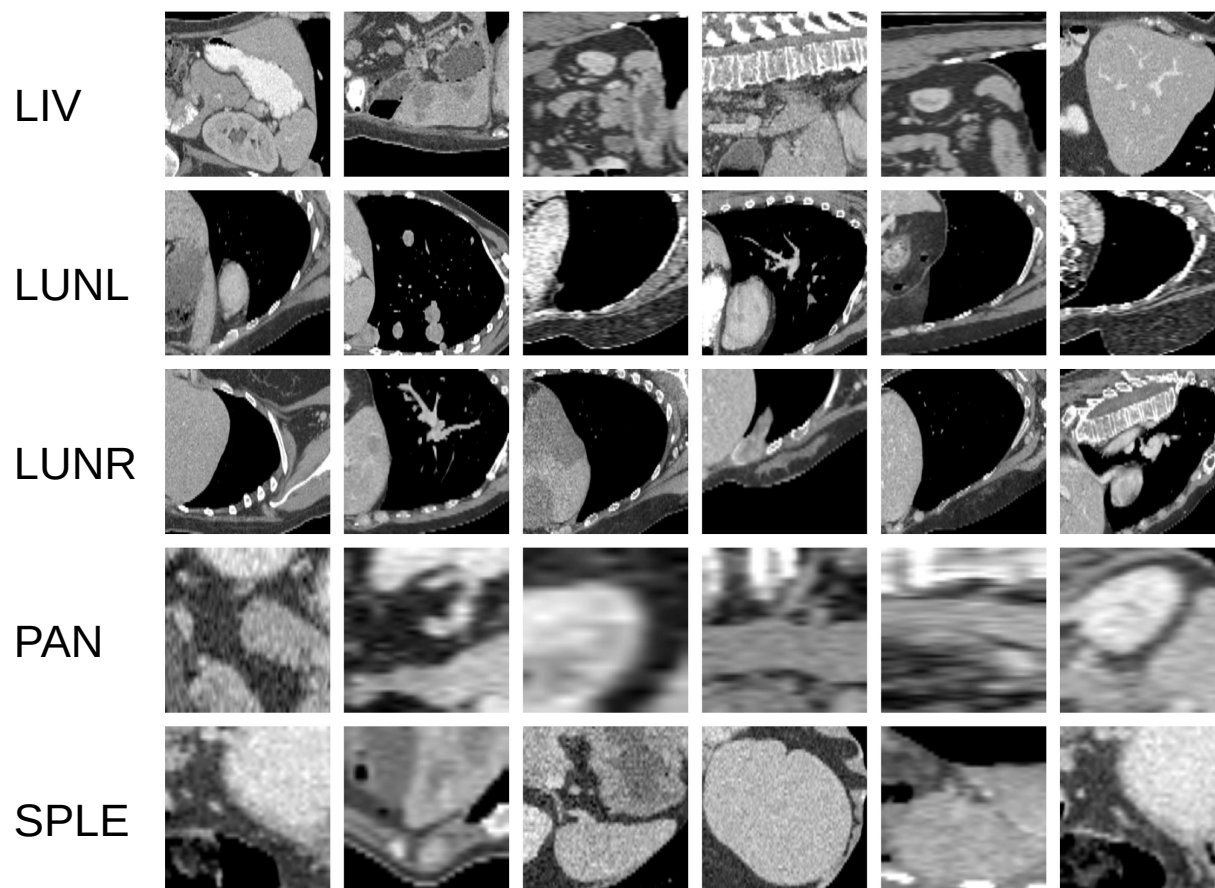
Supplementary Figure 15. Example images of the first six classes in the OrganCMNIST dataset. BLAD: bladder; FEML: femur-left; FEMR: femur-right; HEAR: heart; KIDL: kidney-left; KIDR: kidney-right.



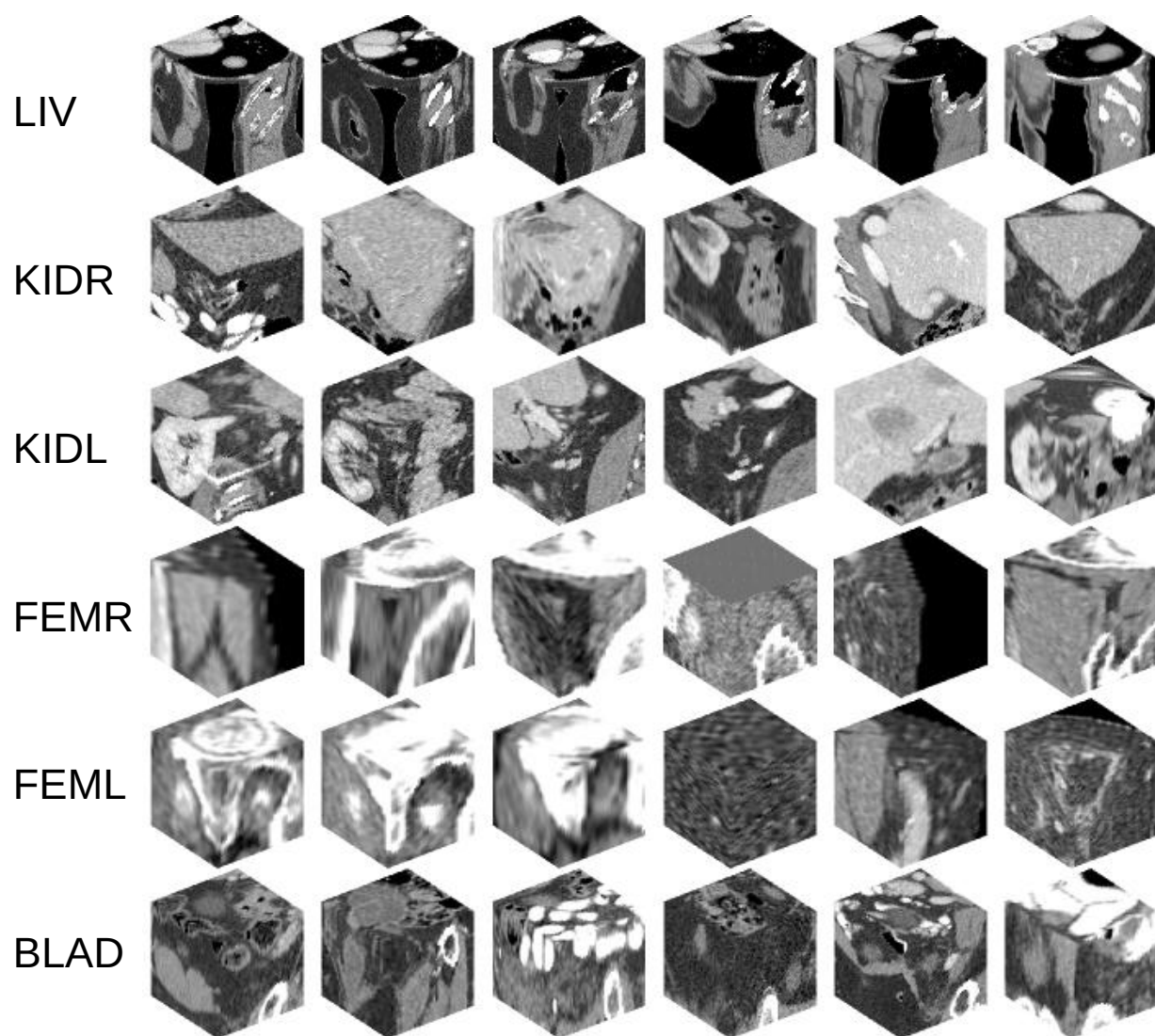
Supplementary Figure 16. Example images of the rest five classes in the OrganCMNIST dataset. LIV: liver; LUNL: lung-left; LUNR: lung-right; PAN: pancreas; SPLE: spleen.



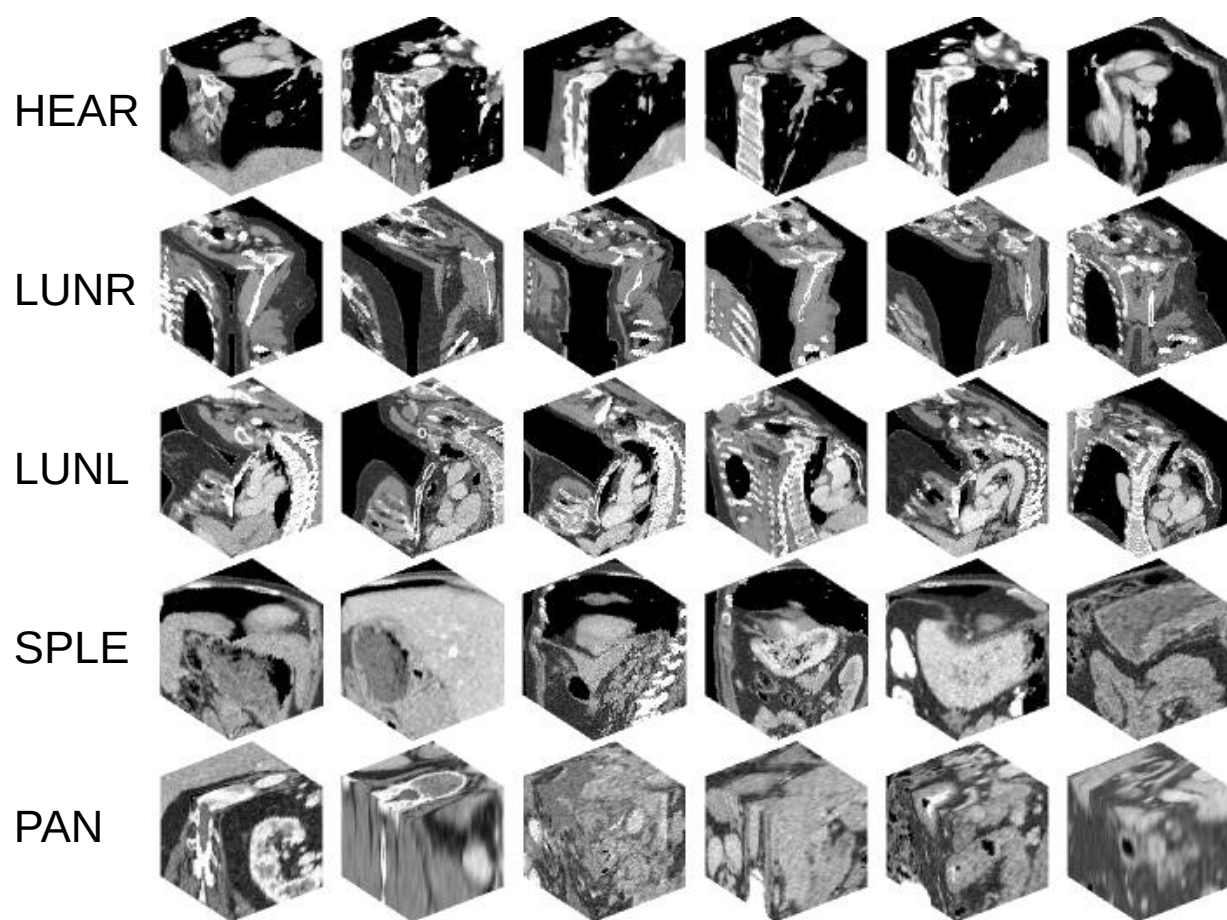
Supplementary Figure 17. Example images of the first six classes in the OrganSMNIST dataset. BLAD: bladder; FEML: femur-left; FEMR: femur-right; HEAR: heart; KIDL: kidney-left; KIDR: kidney-right.



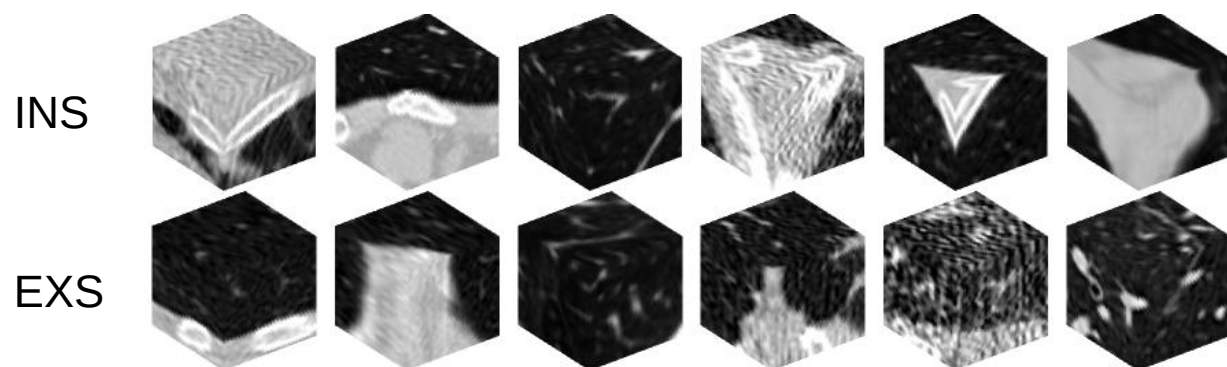
Supplementary Figure 18. Example images of the rest five classes in the OrganSMNIST dataset. LIV: liver; LUNL: lung-left; LUNR: lung-right; PAN: pancreas; SPLE: spleen.



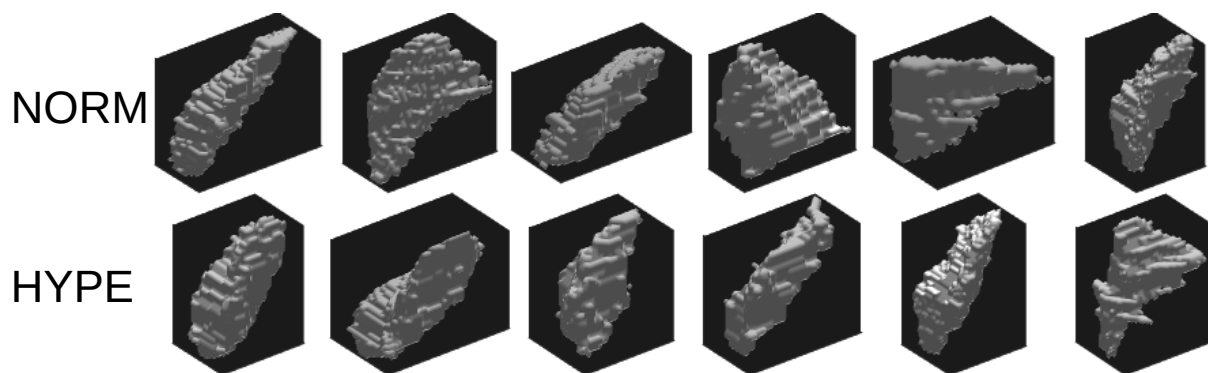
Supplementary Figure 19. Example images of the first six classes in the OrganMNIST3D dataset. LIV: liver; KIDR: kidney-right; KIDL: kidney-left; FEMR: femur-right; FEML: femur-left; BLAD: bladder.



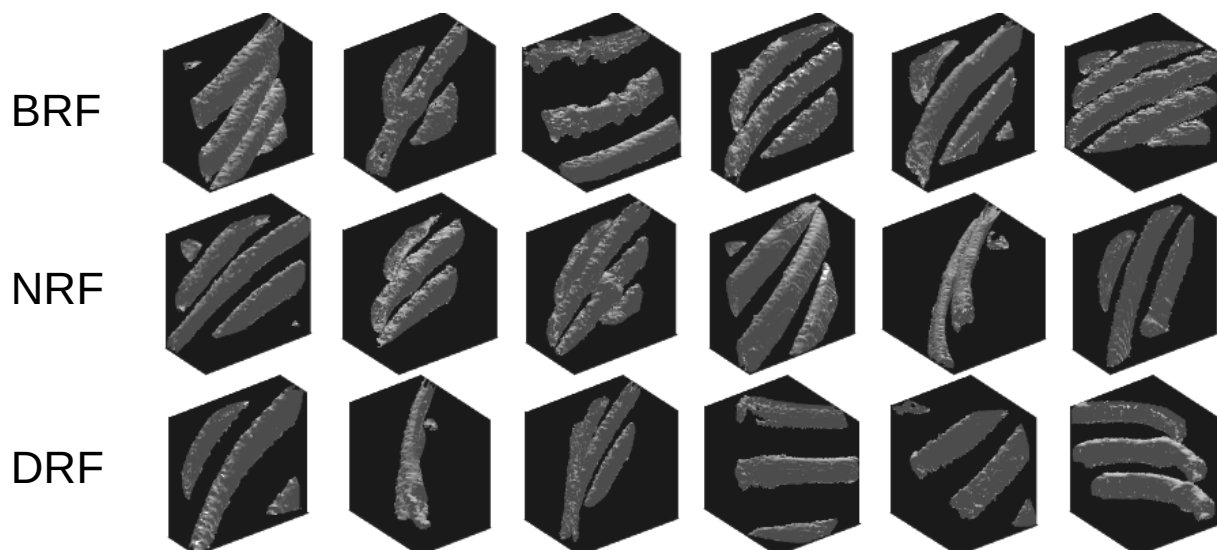
Supplementary Figure 20. Example images of the rest five classes in the OrganMNIST3D dataset. HEAR: heart; LUNR: lung-right; LUNL: lung-left; SPLE: spleen; PAN: pancreas.



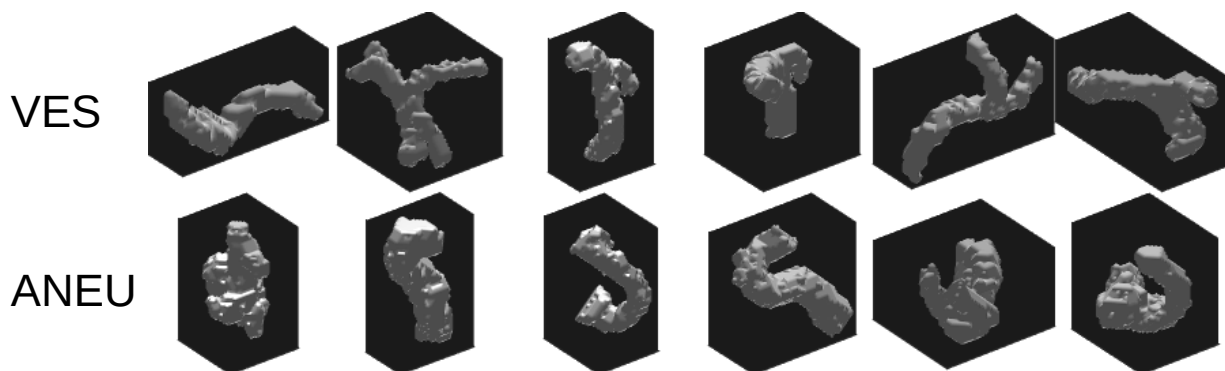
Supplementary Figure 21. Example images of the two classes in the NoduleMNIST3D dataset. BENI: benign; MALI: malignant.



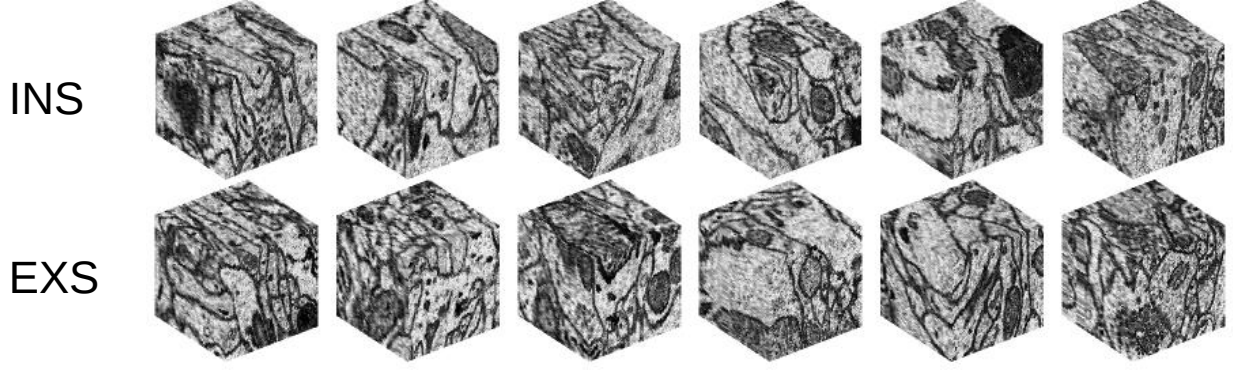
Supplementary Figure 22. Example images of the two classes in the adrenalMNIST3D dataset. NORM: norm; HYPE: hyperplasia.



Supplementary Figure 23. Example images of the three classes in the fractureMNIST3D dataset. BRF: buckle rib fracture; NRF: nondisplaced rib fracture; DRF: displaced rib fracture.



Supplementary Figure 24. Example images of the two classes in the VesselMNIST3D dataset. VES: vessel; ANEU: aneurysm.



Supplementary Figure 25. Example images of the two classes in the SynapseMNIST3D dataset. INS: inhibitory synapse; EXS: excitatory synapse.

4.1 Data Modality

For data modality, we have four groups: Radiology, Microscopy, Ophthalmology, and Dermatology.

- Radiology: ChestMNIST (Chest X-Ray), PneumoniaMNIST (Chest X-Ray), OrganAMNIST (Abdominal CT), OrganCMNIST (Abdominal CT), OrganSMNIST (Abdominal CT), OrganMNIST3D (Abdominal CT), NoduleMNIST3D (Chest CT), FractureMNIST3D (Chest CT), AdrenalMNIST3D (Abdominal CT), and VesselMNIST3D (Brain MRA).
- Microscopy: PathMNIST (Colon Pathology), BloodMNIST (Blood Cell Microscope), TissueMNIST (Kidney Cortex Microscope), and SynapseMNIST3D (Electron Microscope).
- Ophthalmology: OCTMNIST (Retinal OCT), and RetinaMNIST (Fundus Camera).
- Dermatology: DermaMNIST (Dermatoscope).

4.2 Data Scale

For data scale, we have four groups based on the sample size n of each dataset: G1 ($n \leq 10K$), G2 ($10K < n \leq 50K$), G3 ($50K < n \leq 100K$), and G4 ($100K < n$). Here we only use 2D datasets since all the six 3D datasets has a sample size about 1500.

- G1 ($n \leq 10K$): RetinaMNIST (1600 samples), PneumoniaMNIST (5856 samples)
- G2 ($10K < n \leq 50K$): DermaMNIST (10015 samples), BloodMNIST (17092 samples), OrganCMNIST (23583 samples), OrganSMNIST (25211 samples).
- G3 ($50K < n \leq 100K$): OrganAMNIST (58830 samples).
- G4 ($100K < n$): PathMNIST (107180 samples), OCTMNIST (109309 samples), ChestMNIST (112120 samples), TissueMNIST (236386 samples).

4.3 Task Type

For task type, we have four groups based on the number of class n for each classification task: G1 ($n=2$), G2 ($2 < n \leq 5$), G3 ($5 < n \leq 10$), and G4 ($10 < n$).

- G1 ($n=2$): ChestMNIST (Multi-Label (14) Binary-Class (2)), PneumoniaMNIST (Binary-Class), NoduleMNIST3D (Binary-Class), AdrenalMNIST3D (Binary-Class), VesselMNIST3D (Binary-Class), SynapseMNIST3D (Binary-Class).

- G2 ($2 < n \leq 5$): OCTMNIST (4-Class), RetinaMNIST (5-Class), FractureMNIST3D (3-Class).
- G3 ($5 < n \leq 10$): PathMNIST (9-Class), DermaMNIST (7-Class), BloodMNIST (8-Class), TissueMNIST (8-Class).
- G4 ($10 < n$): OrganAMNIST (11-Class), OrganCMNIST (11-Class), OrganSMNIST (11-Class), OrganMNIST3D (11-Class).

5 Model Details

5.1 CNN Architecture

In our MTDL model, the decomposed image x is first fed into an initialization block, which consists of a convolutional layer, a batch normalization operation, and a nonlinear ReLU activation function. The initialized image is then passed through a sequence of simple-CNN layers to extract hierarchical features. Finally, the extracted features are spatially averaged and fed into a multilayer perceptron (MLP) for classification.

Formally, for an input image x , the simple-CNN layer is defined as:

$$\begin{aligned} x' &= \text{Norm}(x + \text{MultiHeadConv}(x)), \\ x'' &= \text{Norm}(x' + \text{FeedForwardConv}(x')), \\ x''' &= \text{MaxPool}(x''), \end{aligned} \tag{29}$$

where Norm is the batch normalization operation, MaxPool represents the max pooling operation, and x''' is the output for x after a TransConv layer. If the input x is a 2D image of dimensions (W, H, C) , where W , H and C correspond to the width, height, and number of channels, respectively, then the output x''' will have dimensions $(\frac{W}{2}, \frac{H}{2}, C)$ due to the pooling operation.

The MultiHeadConv block is designed to mimic the multi-head attention mechanism in the Transformer encoder. It consists of a group convolution, a ReLU activation, and a 1×1 convolution operation. Let $\text{Conv}_{C_{in}, C_{out}, k, g}$ denote a convolution operation with a kernel size k , group number g , input channel C_{in} , and output channel C_{out} . For an input image x with C channels, the MultiHeadConv block is expressed as

$$\begin{aligned} x' &= \text{Conv}_{C, C \times h, 3, h}(x) \\ x'' &= \text{ReLU}(x') \\ x''' &= \text{Conv}_{C \times h, C, 1, 1}(x'') \end{aligned} \tag{30}$$

where x''' is the output of x after the MultiHeadConv block, h is a hyperparameter corresponding to the number of heads in the multi-head attention mechanism. The first convolution emulates the multi-head attention operation, while the second 1×1 convolution serves as a linear layer for feature fusion. Importantly, the MultiHeadConv block preserves the input image dimensions.

The FeedForwardConv block imitates the feedforward neural network layers typically found in the Transformer's encoder. It consists of two group convolutions separated by a ReLU activation function. Formally, for an input image x with C channels, the FeedForwardConv block is defined as:

$$\begin{aligned} x' &= \text{Conv}_{C, 2 \times C, 1, g}(x), \\ x'' &= \text{ReLU}(x'), \\ x''' &= \text{Conv}_{2 \times C, C, 1, g}(x''), \end{aligned} \tag{31}$$

where the two 1×1 convolutions mimic the linear layers in a standard feedforward neural network. Similar to the MultiHeadConv block, the FeedForwardConv block maintains the input image dimensions.

5.2 Data-distribution-based Model Layer Design

The design of model layers is determined based on the data distribution for specific tasks. Formally, for a k -class classification dataset, we compute the average pixel values of all images in the training set for each class and normalize these values to the range $[0, 1]$. This results in k values, and the variance of these values serves as an index for determining the model architecture. The variance and the corresponding model layer configurations for eleven 2D datasets and six 3D datasets are presented in Supplementary Table 4. The

Supplementary Table 4. Data distribution and model layer for different datasets.

MedMNIST2D	Variance	Layer
RetinaMNIST	0.11	5
PneumoniaMNIST	0.25	4
DermaMNIST	0.13	5
BloodMNIST	0.10	5
OrganAMNIST	0.09	5
OrganCMNIST	0.09	5
OrganSMNIST	0.09	5
PathMNIST	0.07	6
OCTMNIST	0.15	5
ChestMNIST	0.25	4
TissueMNIST	0.10	5
MedMNIST3D	Variance	Layer
OrganMNIST3D	0.09	6
NoduleMNIST3D	0.25	5
FractureMNIST3D	0.18	5
AdrenalMNIST3D	0.25	5
VesselMNIST3D	0.25	5
SynapseMNIST3D	0.25	5

variance reflects the degree of dispersion among samples from different classes within the dataset. A larger variance indicates greater separability between classes, making the classification task comparatively easier. Consequently, datasets with larger variances require fewer model layers, while those with smaller variances require deeper models.

For 2D datasets, a 5-layer architecture serves as the baseline for datasets with variances around 0.10. For datasets with significantly larger variances, such as PneumoniaMNIST2D (0.25) and ChestMNIST2D (0.25), the number of layers is reduced to 4. Conversely, for datasets with smaller variances, such as PathMNIST2D (0.07), the number of layers is increased to 6.

Similarly, for 3D datasets, a 5-layer architecture is used as the baseline for datasets with variances around 0.25. For datasets with greatly smaller variances, such as OrganMNIST3D (0.09), the model depth is increased to 6 to accommodate the greater complexity of the classification task.

6 Performance Analysis

6.1 Statistical Significance Analysis

We conducted the Friedman test followed by the Nemenyi post-hoc test to perform the statistical significance analysis of the performance improvement. 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 $1.38\text{e-}14$ (2D) and $5.32\text{e-}10$ (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 Supplementary Figure 26, 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.

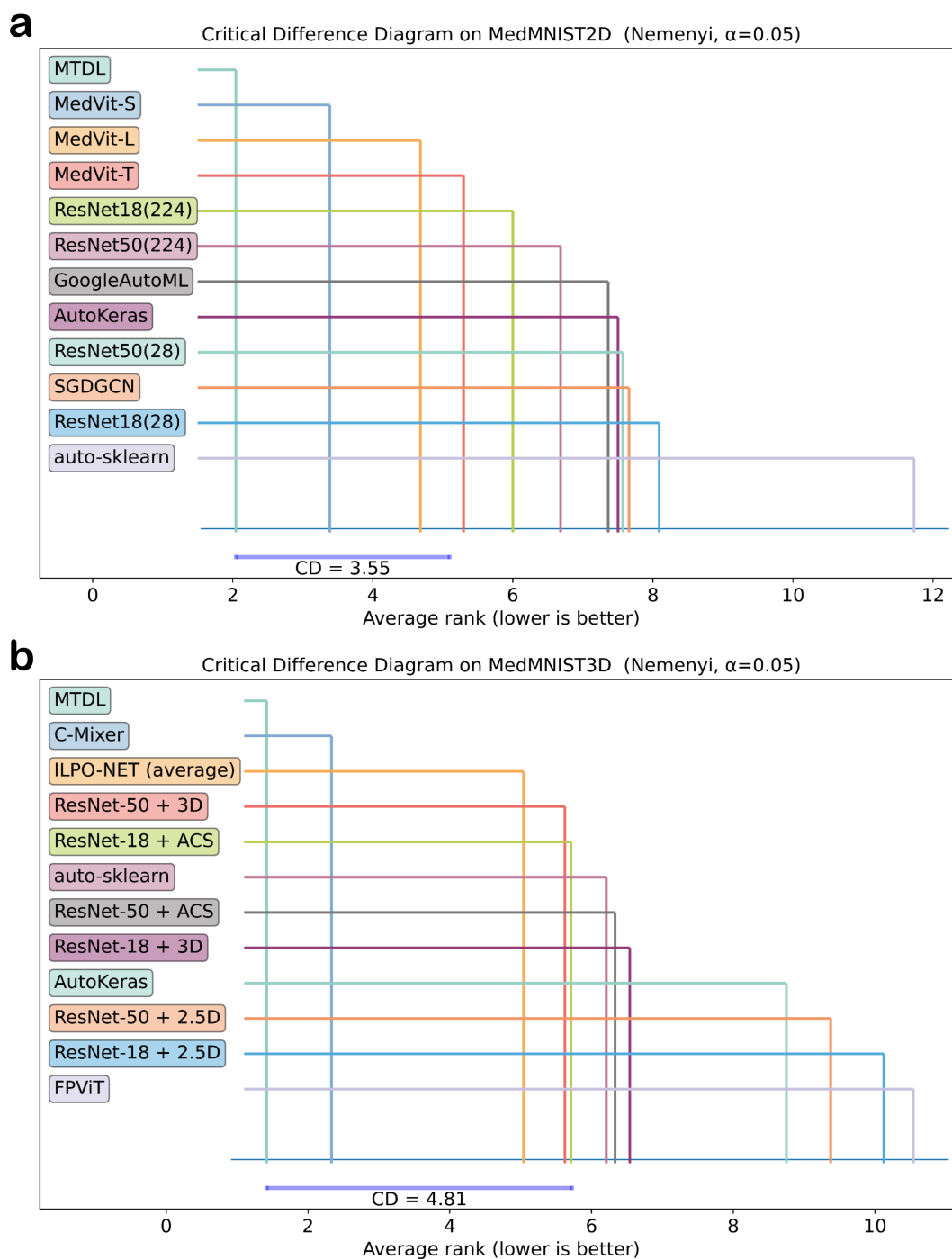
6.2 Ablation Study on the Components of the Hodge Decomposition

In our MTDL model, the three components obtained from the Hodge decomposition are concatenated to construct an enhanced representation of the original images, which is then processed by a CNN architecture for prediction tasks. To better understand the contribution of each component to the overall model performance, We perform an ablation study in which we evaluate the model performance by using different combinations of the components of the decomposition. 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). The results are shown in Supplementary Table 5. We have the following observations:

Supplementary Table 5. 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

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

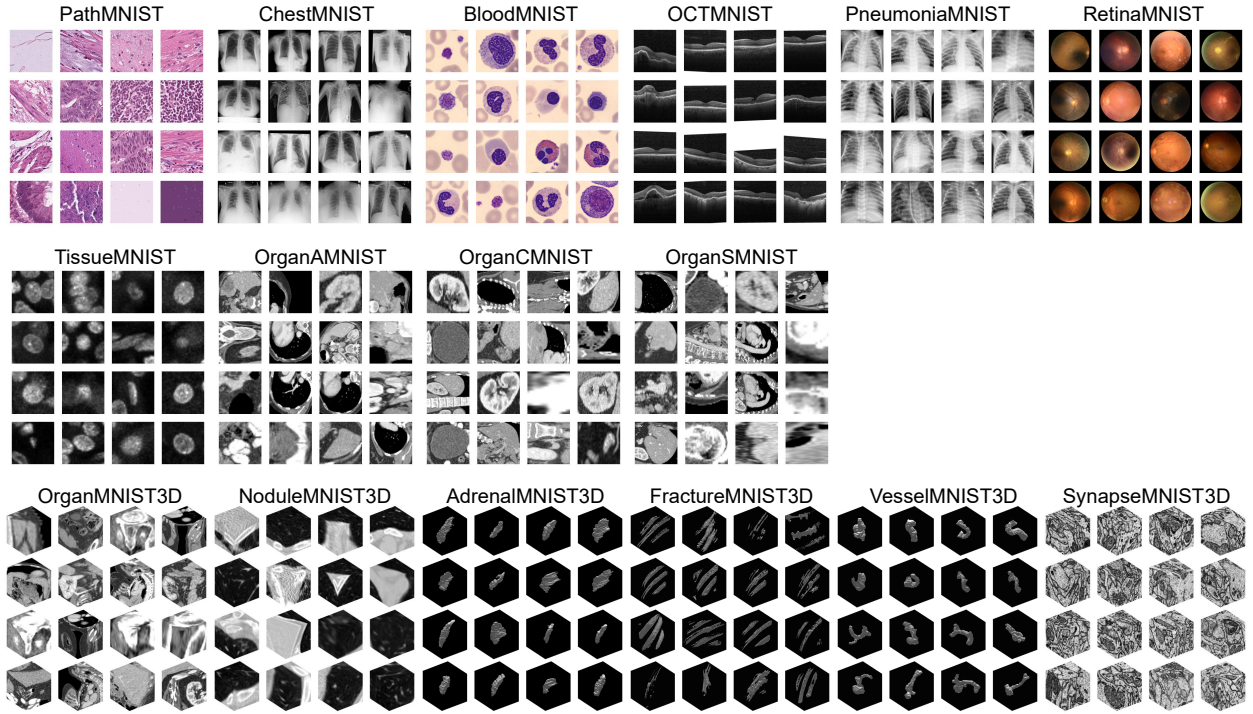


Supplementary Figure 26. The Critical Difference (CD) diagrams illustrating the performance of our model across 11 MedMNIST2D datasets and 6 MedMNIST3D datasets.

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, this 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. This demonstrates that each component captures specific features that cannot be directly extracted from the original images.

6.3 Ablation Study on the Segmentation Parameter

As shown in Supplementary Figure 27, 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.



Supplementary Figure 27. Illustration of the 10 2D datasets and 6 3D datasets in the MedMNIST database. DermaMNIST samples are not shown due to license restrictions.

For the choice of the cutoff parameter, 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 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.

To analyze the impact of segmentation parameter to model performance, we perform 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 Supplementary Table 6. 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.

Supplementary Table 6. 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

6.4 Interpretability Analysis

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.

7 Runtime Analysis

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. Since all the 2D images are 224×224 and all the 3D images are $64 \times 64 \times 64$, we can further speed up the computation using prefactorization of the operators on the entire grid. 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 Supplementary Table 7. As shown, the overall training process is also computationally efficient.

Supplementary Table 7. 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 gray	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

8 Loss Function and Evaluation Metrics

8.1 Loss Function

There are mainly two different tasks in MedMNIST v2 ($\{\text{multi-label, binary-class}\}, \{\text{binary/multi-class, ordinal regression}\}$), we use different loss functions $\mathcal{L}$ for them respectively. For multi-label, binary-class, we employ the BCEWithLogitsLoss $\mathcal{L} = \{L_1, L_2, \dots, L_n\}$ where n is the label number and L_i is as follows

$$\mathcal{L} = -\frac{1}{N} \sum_{n=1}^N \sum_{k=1}^C (z_{n,k} \times \log \sigma(y_{n,k}) + (1 - z_{n,k}) \times \log (1 - \sigma(y_{n,k}))) \quad (32)$$

where N is the number of samples in the current batch, C is the label number, $z_{n,k}$ is the binary label for sample n and label k , $y_{n,k}$ is the logit for sample n and class label k , σ is the sigmoid function. For binary/multi-class, ordinal regression, we use the CrossEntropyLoss

$$\mathcal{L} = -\frac{1}{N} \sum_{n=1}^N \log \frac{\exp(y_{n,l_n})}{\sum_{t=1}^C \exp(y_{n,t})} \quad (33)$$

where N is the number of samples in the current batch, C is the total number of classes, $y_{n,t}$ is the logit for class c of the n -th sample, and l_n is the logit of the target class of the n -th sample.

8.2 Evaluation Metrics

The evaluation metrics employed are the Area under the ROC curve (AUC) and Accuracy (ACC). AUC evaluates the continuous prediction scores without relying on a threshold, whereas ACC assesses the discrete prediction labels based on a given threshold. AUC is less susceptible to class imbalance compared to ACC. There is no severe class imbalance in the MeMNIST v2 dataset, so ACC can effectively serve as a reliable

metric. While numerous other metrics exist, we simply use AUC and ACC have ensure a fair comparison with existing benchmark methods.

9 Detailed Model Performance

The detailed model performance between MTDL and other models over MedMNIST2D and MedMNIST3D are shown in Supplementary Table 8 and Supplementary Table 9 respectively. We also report our results in the form of mean $\pm$ std for all 17 datasets in Supplementary Table 10. The average performance of all models over MedMNIST2D and MedMNIST3D are illustrated in Supplementary Table 11 and Supplementary Table 12 respectively.

References

- [1] William Vallance Douglas Hodge. *The theory and applications of harmonic integrals*. CUP Archive, 1989.
- [2] Charles B Morrey. A variational method in the theory of harmonic integrals, ii. *American Journal of Mathematics*, 78(1):137–170, 1956.
- [3] Clayton Shonkwiler. *Poincaré duality angles on Riemannian manifolds with boundary*. PhD thesis, University of Pennsylvania, 2009.
- [4] Emily Ribando-Gros, Rui Wang, Jiahui Chen, Yiyong Tong, and Guo-Wei Wei. Combinatorial and hodge laplacians: Similarities and differences. *SIAM Review*, 66(3):575–601, 2024.
- [5] Mathieu Desbrun, Eva Kanso, and Yiyong Tong. Discrete differential forms for computational modeling. In *ACM SIGGRAPH 2006 Courses*, pages 39–54. 2006.
- [6] Jakob Nikolas Kather, Johannes Krisam, Pornpimol Charoentong, Tom Luedde, Esther Herpel, Cleo-Aron Weis, Timo Gaiser, Alexander Marx, Nektarios A Valous, Dyke Ferber, et al. Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. *PLoS medicine*, 16(1):e1002730, 2019.
- [7] Jakob Nikolas Kather, Niels Halama, and Alexander Marx. 100,000 histological images of human colorectal cancer and healthy tissue, April 2018. URL <https://doi.org/10.5281/zenodo.1214456>.
- [8] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M Summers. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2097–2106, 2017.
- [9] Philipp Tschandl, Cliff Rosendahl, and Harald Kittler. The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific data*, 5(1):1–9, 2018.
- [10] Noel Codella, Veronica Rotemberg, Philipp Tschandl, M Emre Celebi, Stephen Dusza, David Gutman, Brian Helba, Aadi Kalloo, Konstantinos Liopyris, Michael Marchetti, et al. Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). *arXiv preprint arXiv:1902.03368*, 2019.
- [11] Daniel S Kermany, Michael Goldbaum, Wenjia Cai, Carolina CS Valentim, Huiying Liang, Sally L Baxter, Alex McKeown, Ge Yang, Xiaokang Wu, Fangbing Yan, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *cell*, 172(5):1122–1131, 2018.

Supplementary Table 8. Performance comparison between our model and other benchmarks in terms of Accuracy (ACC) and Area Under the ROC Curve (AUC) on the MedMNIST2D dataset. The benchmark results are taken from the original papers [22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34]. The best results are in bold.

Methods	Path		Chest		Derma		OCT		Pneumonia		Retina	
	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC
ResNet-18 (28)	0.983	0.907	0.768	0.947	0.917	0.735	0.943	0.743	0.944	0.854	0.717	0.524
ResNet-18 (224)	0.989	0.909	0.773	0.947	0.920	0.754	0.958	0.763	0.956	0.864	0.710	0.493
ResNet-50 (28)	0.990	0.911	0.769	0.947	0.913	0.735	0.952	0.762	0.948	0.854	0.726	0.528
ResNet-50 (224)	0.989	0.892	0.773	0.948	0.912	0.731	0.958	0.776	0.962	0.884	0.716	0.511
auto-sklearn	0.934	0.716	0.649	0.779	0.902	0.719	0.887	0.601	0.942	0.855	0.690	0.515
AutoKeras	0.959	0.834	0.742	0.937	0.915	0.749	0.955	0.763	0.947	0.878	0.719	0.503
Google AutoML	0.944	0.728	0.778	0.948	0.914	0.768	0.963	0.771	0.991	0.946	0.750	0.531
MedViT-T (224)	0.994	0.938	0.786	0.956	0.914	0.768	0.961	0.767	0.993	0.949	0.752	0.534
MedViT-S (224)	0.993	0.942	0.791	0.954	0.937	0.780	0.960	0.782	0.995	0.961	0.773	0.561
MedViT-L (224)	0.984	0.933	0.805	0.959	0.920	0.773	0.945	0.761	0.991	0.921	0.754	0.552
ViT	0.962	0.785	0.724	0.947	0.914	0.745	0.905	0.679	0.958	0.885	0.750	0.565
Resnet+ViT	0.991	0.915	0.703	0.947	0.906	0.748	0.968	0.807	0.972	0.897	0.740	0.548
unORANIC	-	-	-	-	0.776	0.699	-	-	0.961	0.862	0.691	0.530
FPViT	0.994	0.918	0.725	0.948	0.923	0.766	0.968	0.813	0.973	0.896	0.753	0.568
BSDA	0.992	0.919	-	-	0.931	0.764	0.989	0.888	0.957	0.888	0.750	0.533
FRDEM	-	-	-	-	-	0.783	-	-	-	-	-	-
AIGC	-	-	-	0.830	-	-	-	-	-	-	-	-
MedKAFormer	0.985	0.882	-	-	0.919	0.748	0.974	0.810	0.961	0.904	0.729	0.535
LPT	-	0.941	-	-	-	0.787	-	0.832	-	-	-	0.575
SGDGCN	0.987	0.907	0.928	0.854	0.919	0.771	0.933	0.783	0.963	0.918	0.782	0.550
PSNAS-Net-ave	-	-	-	-	-	0.772	-	-	-	0.899	-	-
SRE-CONV	-	0.822	-	0.948	-	0.757	-	0.768	-	0.906	-	0.523
IMET	-	-	-	-	-	-	0.960	0.803	0.895	0.902	-	-
NQNN	-	0.752	-	-	-	0.827	-	0.591	-	-	-	-
MTDL	0.996	0.920	0.793	0.948	0.962	0.836	0.989	0.888	0.986	0.910	0.874	0.655

Methods	Blood		Tissue		OrganA		OrganC		OrganS	
	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC
ResNet-18 (28)	0.998	0.958	0.930	0.676	0.997	0.935	0.992	0.900	0.972	0.782
ResNet-18 (224)	0.998	0.963	0.933	0.681	0.998	0.951	0.994	0.920	0.974	0.778
ResNet-50 (28)	0.997	0.956	0.931	0.680	0.997	0.935	0.992	0.905	0.972	0.770
ResNet-50 (224)	0.997	0.950	0.932	0.680	0.998	0.947	0.993	0.911	0.975	0.785
auto-sklearn	0.984	0.878	0.828	0.532	0.963	0.762	0.976	0.829	0.945	0.672
AutoKeras	0.998	0.961	0.941	0.703	0.994	0.905	0.990	0.879	0.974	0.813
Google AutoML	0.998	0.966	0.924	0.673	0.990	0.886	0.988	0.877	0.964	0.749
MedViT-T (224)	0.996	0.950	0.943	0.703	0.995	0.931	0.991	0.901	0.972	0.789
MedViT-S (224)	0.997	0.951	0.952	0.731	0.996	0.928	0.993	0.916	0.987	0.805
MedViT-L (224)	0.996	0.954	0.935	0.699	0.997	0.943	0.994	0.922	0.973	0.806
ViT	-	-	-	-	0.978	0.830	0.976	0.835	0.939	0.657
Resnet+ViT	-	-	-	-	0.995	0.929	0.991	0.900	0.971	0.783
unORANIC	0.977	0.848	-	-	-	-	-	-	-	-
FPViT	-	-	-	-	0.997	0.935	0.993	0.903	0.976	0.785
BSDA	0.999	0.988	0.937	0.704	-	-	-	-	-	-
AIGC	-	-	-	-	-	0.890	-	-	-	-
MedKAFormer	0.998	0.965	-	-	0.995	0.922	0.991	0.899	0.986	0.788
LPT	-	0.967	-	-	-	-	-	0.906	-	0.796
SGDGCN	0.994	0.940	0.801	0.613	0.977	0.850	0.989	0.880	0.969	0.784
PSNAS-Net-ave	-	0.956	-	-	-	-	-	0.913	-	-
SRE-CONV	-	0.960	-	0.712	0.770	0.769	-	-	-	-
NQNN	-	0.737	-	-	-	0.814	-	0.799	-	-
MTDL	0.999	0.988	0.945	0.721	0.998	0.956	0.996	0.928	0.978	0.809

Supplementary Table 9. Performance comparison between our model and other benchmarks in terms of Accuracy (ACC) and Area Under the ROC Curve (AUC) on the MedMNIST3D dataset. The benchmark results are taken from the original papers [22, 35, 23, 36, 37]. The results are in bold.

Methods	Organ		Nodule		Fracture		Adrenal		Vessel		Synapse	
	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC
ResNet-18 + 2.5D	0.977	0.788	0.838	0.835	0.587	0.451	0.718	0.772	0.748	0.846	0.634	0.696
ResNet-18 + 3D	0.996	0.907	0.863	0.844	0.712	0.508	0.827	0.721	0.874	0.877	0.820	0.745
ResNet-18 + ACS	0.994	0.900	0.873	0.847	0.714	0.497	0.839	0.754	0.930	0.928	0.705	0.722
ResNet-50 + 2.5D	0.974	0.769	0.835	0.848	0.552	0.397	0.732	0.763	0.751	0.877	0.669	0.735
ResNet-50 + 3D	0.994	0.883	0.875	0.847	0.725	0.494	0.828	0.745	0.907	0.918	0.851	0.795
ResNet-50 + ACS	0.994	0.889	0.886	0.841	0.750	0.517	0.828	0.758	0.912	0.858	0.719	0.709
auto-sklearn	0.977	0.814	0.914	0.874	0.628	0.453	0.828	0.802	0.910	0.915	0.631	0.730
AutoKeras	0.979	0.804	0.844	0.834	0.642	0.458	0.804	0.705	0.773	0.894	0.538	0.724
FPViT (224)	0.923	0.800	0.814	0.822	0.640	0.438	0.801	0.704	0.770	0.888	0.530	0.712
BSDA	0.994	0.887	0.892	0.861	0.731	0.569	0.892	0.838	0.917	0.932	-	-
ILPO-NET (average)	0.972	0.728	0.900	0.861	0.776	0.577	0.880	0.811	0.888	0.888	0.854	0.782
C-Mixer	0.995	0.912	0.915	0.860	0.729	0.660	0.969	0.801	0.932	0.940	0.866	0.820
MTDL	0.999	0.952	0.916	0.865	0.753	0.583	0.903	0.862	0.938	0.937	0.951	0.931

Supplementary Table 10. Model performance (mean $\pm$ std) of our model over all the 17 datasets.

	Path	Chest	Derma	OCT	Pneumonia	Retina
AUC	0.996 $\pm$ 0.0012	0.793 $\pm$ 0.0029	0.962 $\pm$ 0.0038	0.989 $\pm$ 0.0029	0.986 $\pm$ 0.0008	0.874 $\pm$ 0.0017
ACC	0.920 $\pm$ 0.0075	0.948 $\pm$ 0.0005	0.836 $\pm$ 0.0025	0.888 $\pm$ 0.0017	0.910 $\pm$ 0.0095	0.655 $\pm$ 0.0071
	Blood	Tissue	OrganA	OrganC	OrganS	
AUC	0.999 $\pm$ 0.0	0.945 $\pm$ 0.0037	0.998 $\pm$ 0.0	0.996 $\pm$ 0.0005	0.978 $\pm$ 0.0012	
ACC	0.988 $\pm$ 0.0012	0.721 $\pm$ 0.0071	0.956 $\pm$ 0.0026	0.928 $\pm$ 0.0041	0.809 $\pm$ 0.0012	
	Organ	Nodule	Fracture	Adrenal	Vessel	Synapse
AUC	0.999 $\pm$ 0.0005	0.916 $\pm$ 0.0045	0.753 $\pm$ 0.0042	0.903 $\pm$ 0.0060	0.938 $\pm$ 0.0065	0.951 $\pm$ 0.0075
ACC	0.952 $\pm$ 0.0053	0.865 $\pm$ 0.0098	0.583 $\pm$ 0.0177	0.862 $\pm$ 0.0029	0.937 $\pm$ 0.0098	0.931 $\pm$ 0.0089

Supplementary Table 11. Average Performance Comparison in ACC and AUC over all datasets in MedMNIST2D. The best result is in bold.

Methods	Average	
	AUC	ACC
ResNet-18 (28)	0.924	0.815
ResNet-18 (224)	0.928	0.820
ResNet-50 (28)	0.926	0.817
ResNet-50 (224)	0.928	0.820
auto-sklearn	0.882	0.714
AutoKeras	0.921	0.811
Google AutoML	0.928	0.804
MedVIT-T	0.936	0.835
MedVIT-S	0.943	0.846
MedVIT-L	0.936	0.838
ViT	0.901	0.770
Resnet+ViT	0.915	0.830
unORANIC	0.851	0.735
FPViT	0.922	0.837
SGDGCN	0.931	0.805
BSDA	0.936	0.812
MTDL	0.956	0.868

Supplementary Table 12. Average Performance Comparison in ACC and AUC over all datasets in MedMNIST3D, the best result is in bold.

Methods	Average	
	AUC	ACC
ResNet-18 + 2.5D	0.750	0.731
ResNet-18 + 3D	0.849	0.767
ResNet-18 + ACS	0.842	0.775
ResNet-50 + 2.5D	0.752	0.732
ResNet-50 + 3D	0.863	0.780
ResNet-50 + ACS	0.848	0.762
auto-sklearn	0.815	0.765
AutoKeras	0.763	0.737
FPVT	0.746	0.727
BSDA	0.885	0.817
ILPO-NET(average)	0.878	0.775
C-Mixer	0.901	0.832
MTDL	0.910	0.855

- [12] Daniel Kermany, Kang Zhang, and Michael Goldbaum. Large dataset of labeled optical coherence tomography (oct) and chest x-ray images. *Mendeley Data*, 3(10.17632), 2018.
- [13] Ruhan Liu, Xiangning Wang, Qiang Wu, Ling Dai, Xi Fang, Tao Yan, Jaemin Son, Shiqi Tang, Jiang Li, Zijian Gao, et al. Deepdrid: Diabetic retinopathy—grading and image quality estimation challenge. *Patterns*, 3(6), 2022.
- [14] World Health Organization. *Prevention of blindness from diabetes mellitus: report of a WHO consultation in Geneva, Switzerland, 9-11 November 2005*. World Health Organization, 2006.
- [15] Andrea Acevedo, Anna Merino, Santiago Alf  rez,   ngel Molina, Laura Bold  , and Jos   Rodellar. A dataset of microscopic peripheral blood cell images for development of automatic recognition systems. *Data in brief*, 30:105474, 2020.
- [16] Andre Woloshuk, Suraj Khochare, Aljohara F Almulhim, Andrew T McNutt, Dawson Dean, Daria Barwinska, Michael J Ferkowicz, Michael T Eadon, Katherine J Kelly, Kenneth W Dunn, et al. In situ classification of cell types in human kidney tissue using 3d nuclear staining. *Cytometry Part A*, 99(7): 707–721, 2021.
- [17] Vebjorn Ljosa, Katherine L Sokolnicki, and Anne E Carpenter. Annotated high-throughput microscopy image sets for validation. *Nature methods*, 9(7):637–637, 2012.
- [18] Patrick Bilic, Patrick Christ, Hongwei Bran Li, Eugene Vorontsov, Avi Ben-Cohen, Georgios Kaissis, Adi Szeskin, Colin Jacobs, Gabriel Efrain Humpire Mamani, Gabriel Chartrand, et al. The liver tumor segmentation benchmark (lits). *Medical Image Analysis*, 84:102680, 2023.
- [19] Samuel G Armato III, Geoffrey McLennan, Luc Bidaut, Michael F McNitt-Gray, Charles R Meyer, Anthony P Reeves, Binsheng Zhao, Denise R Aberle, Claudia I Henschke, Eric A Hoffman, et al. The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans. *Medical physics*, 38(2):915–931, 2011.
- [20] Liang Jin, Jiancheng Yang, Kaiming Kuang, Bingbing Ni, Yiyi Gao, Yingli Sun, Pan Gao, Weiling Ma, Mingyu Tan, Hui Kang, et al. Deep-learning-assisted detection and segmentation of rib fractures from ct scans: Development and validation of fracnet. *EBioMedicine*, 62, 2020.
- [21] Xi Yang, Ding Xia, Taichi Kin, and Takeo Igarashi. Intra: 3d intracranial aneurysm dataset for deep learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2656–2666, 2020.
- [22] Jinwei Liu, Yan Li, Guitao Cao, Yong Liu, and Wenming Cao. Feature pyramid vision transformer for medmnist classification decathlon. In *2022 International joint conference on neural networks (IJCNN)*, pages 1–8. IEEE, 2022.
- [23] Jiancheng Yang, Rui Shi, Donglai Wei, Zequan Liu, Lin Zhao, Bilian Ke, Hanspeter Pfister, and Bingbing Ni. Medmnist v2-a large-scale lightweight benchmark for 2d and 3d biomedical image classification. *Scientific Data*, 10(1):41, 2023.
- [24] Omid Nejati Manzari, Hamid Ahmadabadi, Hossein Kashiani, Shahriar B Shokouhi, and Ahmad Aya-tollahi. Medvit: a robust vision transformer for generalized medical image classification. *Computers in Biology and Medicine*, 157:106791, 2023.
- [25] Sebastian Doerrich, Francesco Di Salvo, and Christian Ledig. unoranic: Unsupervised orthogonalization of anatomy and image-characteristic features. In *International Workshop on Machine Learning in Medical Imaging*, pages 62–71. Springer, 2023.

- [26] Arindam Halder, Anogh Dalal, Sanghita Gharami, Marcin Wozniak, Muhammad Fazal Ijaz, and Pawan Kumar Singh. A fuzzy rank-based deep ensemble methodology for multi-class skin cancer classification. *Scientific Reports*, 15(1):6268, 2025.
- [27] Beining Wu, Jun Huang, and Qiang Duan. Real-time intelligent healthcare enabled by federated digital twins with aoi optimization. *IEEE Network*, 2025.
- [28] Guoli Wang, Qikui Zhu, Chaoda Song, Benzhen Wei, and Shuo Li. Medkaformer: When kolmogorov-arnold theorem meets vision transformer for medical image representation. *IEEE Journal of Biomedical and Health Informatics*, 2025.
- [29] Yu Bai, Liang Bai, Xian Yang, and Jiye Liang. Label-semantic-based prompt tuning for vision transformer adaptation in medical image analysis. *IEEE Transactions on Circuits and Systems for Video Technology*, 2025.
- [30] Aryan Singh, Ciarán Eising, Patrick Denny, and Pepijn van de Ven. Improving gnns for image classification: Addressing homophily challenges. *IEEE Open Journal of the Computer Society*, 2025.
- [31] Zheng Zechen, He Xuelei, Zhao Fengjun, and He Xiaowei. Psnas-net: Hybrid gradient-physical optimization for efficient neural architecture search in customized medical imaging analysis. *Expert Systems with Applications*, 288:128155, 2025.
- [32] Yuexi Du, Jiazhen Zhang, Tal Zeevi, Nicha C Dvornek, and John A Onofrey. Sre-conv: Symmetric rotation equivariant convolution for biomedical image classification. In *2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI)*, pages 1–5. IEEE, 2025.
- [33] Ruhaan Singh and Sreelekha Guggilam. Iterative misclassification error training (imet): An optimized neural network training technique for image classification. *IEEE Access*, 2025.
- [34] Maqsudur Rahman and Jun Zhuang. Nqnn: Noise-aware quantum neural networks for medical image classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 433–442. Springer, 2025.
- [35] Zhuoran Zheng and Xiuyi Jia. Complex mixer for medmnist classification decathlon. *arXiv preprint arXiv:2304.10054*, 2023.
- [36] Yaoyao Zhu, Xiuding Cai, Xueyao Wang, Xiaoqing Chen, Zhongliang Fu, and Yu Yao. Bsda: Bayesian random semantic data augmentation for medical image classification. *Sensors*, 24(23):7511, 2024.
- [37] Dmitrii Zhemchuzhnikov and Sergei Grudinin. Ilpo-net: Network for the invariant recognition of arbitrary volumetric patterns in 3d. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 352–368. Springer, 2024.