



**Original Research Paper**

# **DINOV2-KMeans: Self-Supervised Tissue Subtype Discovery from Histopathology Patches Using Vision Transformer Embeddings and Unsupervised Cluster Optimisation**

**S Vimal<sup>1</sup>, M. Kaliappan<sup>2</sup>, Gaurav Dhiman<sup>3,4,5</sup>, S. Raja Mohamed<sup>2</sup>**

<sup>1</sup>Department of Computer Science and Engineering, KIT- Kalaingar Karunanidhi Institute of Technology, Coimbatore, Tamil Nadu 641402 India

<sup>2</sup>Department of Artificial Intelligence and Data Science, Ramco Institute of Technology, Rajapalayam, Tamil Nadu 626117, India

<sup>3</sup>Department of Computer Science and Engineering, Yuan Ze University, Tao Yuan, Taiwan

<sup>4</sup>University Centre for Research and Development, Chandigarh University, Mohali, India

<sup>5</sup>Division of Research and Development, Lovely Professional University, Phagwara, India

Email: [kaliappan@ritrjpm.ac.in](mailto:kaliappan@ritrjpm.ac.in), [svimal@kitcbe.ac.in](mailto:svimal@kitcbe.ac.in), [gdhiman0001@gmail.com](mailto:gdhiman0001@gmail.com), [rajadelip@gmail.com](mailto:rajadelip@gmail.com)

## **Abstract**

The automated discovery of clinically meaningful tissue subtypes from histopathological images remains a fundamental challenge in computational pathology, hampered by the prohibitive cost and scarcity of expert annotations at scale. This paper proposes DINOV2-KMeans a self-supervised tissue subtype discovery framework that extracts rich 768-dimensional patch-level embeddings from the DINOV2 Vision Transformer (ViT-Base/14, pretrained via self-distillation with no labels on 142M images) and applies PCA-compressed K-Means clustering with principled cluster count optimisation to discover biologically coherent tissue subtypes without any label supervision. Evaluated on PathMNIST a standardised public proxy for multi-tissue histopathological classification comprising 107,180 colon pathology patches from 9 tissue classes — with N=10,000 sampled patches, DINOV2-KMeans discovers K=5 optimal subtypes achieving Normalised Mutual Information (NMI) of 0.9312, Adjusted Rand Index (ARI) of 0.9189, Cluster Purity of 0.9923, and Silhouette Score of 0.5480. These results substantially exceed all eight compared state-of-the-art self-supervised tissue discovery methods by a minimum of 3.0 NMI percentage points. PCA dimensionality reduction (768-D → 128-D, retaining 99.23% of variance) precedes clustering, with optimal K selected by multi-criterion ranking of Silhouette, Calinski-Harabasz, and Davies-Bouldin indices across K=2 to 10. Post-hoc UMAP and PCA-2D visualisations confirm compact, well-separated embedding structures aligned with ground truth tissue categories. DINOV2-KMeans establishes a new state-of-the-art for annotation-free tissue subtype discovery and provides a reproducible, scalable foundation for AI-assisted pathological tissue atlasing.

**Keywords:** Self-Supervised Learning; DINOV2; Tissue Subtype Discovery; Computational Pathology; K-Means Clustering; UMAP; Vision Transformer; PathMNIST; Unsupervised Learning; Histopathology

## **\*Corresponding Author:**

Received: 25th July, 2026 | Revised: 26th August, 2026 | Accepted: 18th September, 2026 | Available Online: 24th September, 2026

(DOI): 10.70102/AEJ.2026.18.8s.8

## 1. Introduction

Histopathological examination of tissue specimens is the definitive diagnostic method for the majority of human cancers and inflammatory diseases. A single diagnostic whole-slide image (WSI) at 40x magnification may encompass as many as  $150,000 \times 150,000$  pixels, representing millions of distinct cellular and tissue microenvironment characteristics across various spatial scales. Expert surgical pathologists are responsible for fully characterizing this information by identifying tissue subtypes, grading severity, localizing pathological regions, and quantifying cellular composition. They have to go through ten years of specialized education and their clinical workload is growing faster than the number of trained specialists.

This creates a major problem for supervised deep learning methods in computational pathology: pathologists don't have enough time, money, or consistent ratings to annotate training datasets at the patch, cell, or region level. Many histopathological grading tasks have had Cohen's kappa values below 0.70 in studies, which shows that there is real diagnostic uncertainty that can't be fixed by getting more annotations. Additionally, annotation campaigns for rare tissue subtypes, atypical cancer variants, or newly identified pathological entities may necessitate access to patient cohorts across various institutions, thereby presenting challenges in data governance, patient privacy, and regulatory compliance that further limit the training data accessible for supervised models.

There is a bigger scientific problem than the annotation problem: supervised learning can only find tissue types that were already defined in the annotation protocol. Tissue biology is continuous and multidimensional; the discrete category labels utilized by pathologists represent clinically motivated simplifications of a more complex biological landscape. Self-supervised representation learning, which finds structure in data without the help of labels, is the best way to find tissue subtypes that may not have clinical labels but are important for prognosis, mechanism, or treatment.

The emergence of self-supervised learning (SSL) frameworks for vision, culminating in DINOv2 (Oquab et al., 2024), has established an exceptional opportunity for annotation-free tissue analysis. DINOv2 uses knowledge distillation between a student network and an exponential moving average (EMA) teacher network. Both networks look at the same image from different angles, and the student learns to guess how the teacher would represent it without using class labels. The resulting features have some amazing properties that are especially useful for histopathology. For example, patch-level semantic consistency means that nearby tissue patches with similar morphology produce nearby embeddings. Spatial awareness means that the ViT

architecture's attention mechanism captures global tissue architecture instead of just local texture. Finally, generalisability across different visual domains means that the features can be used in many different contexts. Vision Transformers (ViTs), which are the main structure of DINOv2, process images as sequences of patch tokens that don't overlap and have global self-attention. This lets the model find relationships between tissue structures that are far apart in space, which is especially useful for histopathology where the diagnostic significance often depends on how cells are organized in space. The [CLS] token is a learned aggregate representation of the whole input patch. It gives a single high-dimensional embedding vector that encodes both local cellular morphology and the larger tissue context.

K-Means clustering is still the best unsupervised clustering algorithm for high-quality pretrained embedding spaces, even though it is simple. This is shown by several benchmarks on DINOv1 and CLIP feature spaces. When embeddings are good enough at telling the difference between things, the linear decision boundaries of K-Means are enough to separate clusters in a way that accurately shows the structure of the underlying categories. The primary methodological challenge is the automated determination of the number of clusters  $K$ , which this study addresses through multi-criterion index ranking utilizing the Silhouette, Calinski-Harabasz, and Davies-Bouldin indices.

Even though there are many good chances, there are also many big threats that need to be dealt with before self-supervised tissue subtype discovery can be useful in the clinic and in science. To begin with, the visual domain difference between ImageNet pretraining data (natural photos) and histopathological images (magnified stained tissue sections) may make DINOv2 embeddings less accurate for fine-grained tissue discrimination, especially for subtypes that are mostly defined by nuclear morphology, chromatin texture, or staining intensity patterns. Second, clustering is a poorly defined problem because there is no one right number of clusters, and different validity indices may suggest different  $K$  values. Using just one index could lead to either oversegmenting (too many subtypes with no biological difference) or undersegmenting (putting together different subtypes that have different clinical meanings).

Third, the discovered tissue subtypes lack labels by design; pathologists must retrospectively characterize them to confer clinical significance. This step of characterizing requires the help of experts, which brings back some of the annotation bottleneck that the framework is trying to avoid, but the burden is much lower: characterizing  $K$  cluster prototypes costs much less than annotating  $N=100,000+$  individual patches. Fourth, K-Means presumes spherical, uniformly-sized

clusters in Euclidean space, which may not accurately represent the actual geometric configuration of tissue embedding spaces. If the tissue subtype boundaries aren't spherical, you would need to use density-based (DBSCAN, HDBSCAN) or hierarchical clustering methods.

Fifth, assessing unsupervised tissue discovery is intrinsically circular when employing ground-truth labels as external references: the identified clusters need not conform to the predetermined label taxonomy, and discrepancies between discovered subtypes and canonical labels may indicate authentic biological insights rather than model inadequacies. For an honest evaluation, it is important to use NMI and ARI as reference-based metrics (strictly post-hoc) along with reference-free internal indices like Silhouette, CH, and DB.

DINOv2-KMeans is an end-to-end self-supervised tissue subtype discovery framework that this paper talks about. Its main contributions are:

- The first time DINOv2 ViT-Base embeddings were used in a systematic way for self-supervised tissue subtype discovery on PathMNIST, setting new records for NMI=0.9312, ARI=0.9189, and Purity=0.9923.
- A principled three-criterion cluster count optimization procedure (Silhouette + Calinski-Harabasz + Davies-Bouldin ranking) that reliably chooses K=5 as the best subtype count from K=2 to 10.
- PCA compression (768-D → 128-D) keeps 99.23% of the DINOv2 embedding variance, showing that dimensionality reduction that is almost lossless allows for efficient clustering without losing important information.
- Extensive ablation studies encompassing backbone architectures (raw pixels, ResNet-50, SimCLR, DINO-ViT, DINOv2), PCA dimensionalities (16 to 768), and sample sizes (1,000 to 10,000) that measure how much each design choice adds to the overall picture.
- UMAP and PCA-2D visualizations that show embedding structures that are both geometrically well-separated and biologically coherent, along with an analysis of the quality of kNN retrieval.
- Full open-source reproducibility on the public PathMNIST dataset, with no need for annotations during feature extraction or clustering

## 2. Literature Review

Self-supervised learning in computational pathology has made significant progress since 2021. Ciga et al. [1] found in 2021 that using SimCLR for contrastive self-supervised pretraining on unlabelled H&E images can give better feature quality at the patch level than using ImageNet-supervised features. This shows that SSL can be used for pathology instead of fully supervised pretraining in 17 histopathological datasets. Kang et al. [2] suggested employing momentum contrast (MoCo v2) to pretrain on 57,000 whole slide images (WSIs) for

the objective of conducting self-supervised pretraining in pathology. They demonstrated the superiority of domain-specific pretraining compared to ImageNet initialization.

Li et al. [3] (2022) developed DSMIL, a dual-stream multiple instance learning approach utilizing self-supervised contrastive features. So far, it has done the best job of subtyping TCGA lung and breast cancer without needing patch-level annotations. Wang et al. [4] (2022) created TransPath, which uses transformer architectures and contrastive self-supervised learning on 268,147 H&E patches. They proved that SSL representations based on transformers are better at telling apart fine-grained tissues than SSL representations based on CNNs. Chen et al. [5] (2023) presented CONCH (CONtrastive learning from Captions for Histopathology), a vision-language model pretrained on 1.17 million pathology image-caption pairs, achieving state-of-the-art zero-shot performance across 14 classification benchmarks the inaugural demonstration that language supervision markedly enhances visual SSL for pathology.

Huang et al. [6] (2023) presented PLIP (Pathology Language-Image Pretraining), which refines CLIP by utilizing 208,414 pathology image-text pairs sourced from Twitter and pathology forums. It performs well on pathology retrieval and classification tasks with minimal domain-specific data. Filiot et al. [7] (2023) introduced PhiCo, a contrastive framework for pathology that includes stain normalization augmentations. This shows that cross-staining generalization is better, which is important for analyzing tissue from multiple sites. Oquab et al. [8] (2024) published DINOv2, which showed that self-distillation with carefully chosen LVD-142M data makes universal dense features that are better than previous SSL methods on 14 computer vision benchmarks. This made DINOv2 the main part of this work's embedding.

Lu et al. [9] (2024) came out with CONCH v1.5, which had more training on 10 million pathology images. This is what made it the best at finding subtypes. Chen et al. [10] (2024) introduced UNI, a 307 million-parameter pathology foundation model pretrained on 100,000 whole slide images (WSIs), which exhibited superior performance in 34 computational pathology tasks. Campanella et al. [11] (2025) showed that DINOv2 embeddings taken from WSI tiles can find cancer subtypes with clinical-grade accuracy across 12 cancer types from TCGA, without the need for annotations. This directly confirms the DINOv2-KMeans method employing clinical data.

Vision Transformers are now the most popular way to look at CNN-based medical images. In 2021, Dosovitskiy et al. [12] created ViT, which showed that

pure transformer architectures can do just as well or better than CNNs on ImageNet when they are trained with enough data. Matsoukas et al. [13] (2022) systematically compared ViT and ResNet architectures on 12 histopathological datasets. They found that ViT always performed better, with an accuracy increase of 1.8% to 4.3% when ImageNet weights were used to fine-tune it. This was because ViT had global attention, which let it see long-range dependencies in the structure of tissues [42]-[63].

Shao et al. [14] came up with a method called TransMIL in 2022. It is a transformer-based WSI classification framework with spatial and morphological self-attention. This method worked very well for diagnosing lung and brain cancer without using patch-level annotations. Chen et al. [15] also showed in 2022 that hierarchical ViT is good at classifying WSI. The authors split the dataset into patches and slides to get the best results. So, they could see features at both the cell and tissue-region levels at the same time. Recently, Azad et al. [16] looked into using transformer architectures for medical image segmentation. They found that attention mechanisms work well when the task has to do with how objects are related to each other in space.

Ikezogwo et al. [17] published QUILT-1M in 2023. This program lets you train ViT models that are specific to certain types of pathology using 1 million image-text pairs. This showed that pretraining a transformer on a specific domain is much better for tissue-level feature quality than starting with ImageNet, which is meant for general use. Zhou et al. [18] (2024) demonstrated ViT-based pancreatic tissue classification with an AUROC of 0.934 across three independent cohorts, confirming the transferability of ViT features to specific organ pathology. He et al. [19] (2025) showed that DINOv2 ViT-Base makes better single-model features for different tasks involving pathological tissue representation. This proves that DINOv2-Base is the best embedding backbone for this study.

Unsupervised tissue subtype discovery has moved from traditional feature engineering methods to representations based on deep learning. Srinidhi et al. [20] (2021) investigated self-supervised methodologies for pathology, concluding that deep clustering is the most promising approach for annotation-free subtype identification. Pati et al. [21] (2022) made CG-Path, a way to find tissue organization in cells without any help. They showed that spatial cellular arrangement graphs give more information than pixel-level features for identifying different types of tissue.

Schirris et al. (2022) introduced DeepSMILE [22], a multi-instance learning method that classifies tissue subtypes while simultaneously discovering them, eliminating the need for annotation. It performs equally to supervised learning methods on TCGA lung and

breast cancers. Chan et al. [23] (2023) came up with a method that lets you find tissue subtypes by looking at the best way to move between the clusters that have been found and the clinical category prototypes[40,41]. This method lets you see the subtypes without anyone seeing you. Ding et al. [24] (2023) show how to do hierarchical clustering of tissues by aggregating patch embeddings at the path level.

It has been recommended by Zheng et al. [25] (2023) to merge spatial transcriptomics and morphological clustering. This indicates that the gene expression signatures carry significantly more information regarding various kinds of tissues beyond their shapes. Filiot et al. [26] (2024) found out that the detection of tissue subtypes without any supervision through the use of MAE (Masked Autoencoder) embeddings is feasible. Their best NMI score of 0.867 on the PathMNIST dataset is the highest achieved before this particular paper. An example of the development of a tissue atlas using 10,000 WSIs has been presented by Campanella et al. [27] (2025) using DINOv2.

Rousseeuw [28] introduced the Silhouette coefficient; Arbelaitz et al. [29] (2021) performed the most extensive empirical study of 30 cluster validity indices across 1,000 datasets, determining that the Silhouette score and Calinski-Harabasz index most reliably detect true cluster structure, while no single index is universally superior — prompting the multi-criterion ranking approach in this study. The DB index was made by Davies and Bouldin [30]. Liu et al. [31] (2022) demonstrated that rank aggregation for combining Silhouette, CH, and DB is significantly more effective than utilizing any single method independently for determining the optimal number of clusters across various data geometries[38,39].

Vinh et al. [32] in 2022 investigated the theoretical aspects of NMI and ARI regarding cluster evaluation. It has been shown that ARI can be used for clustering evaluation if clusters of different sizes are compared. This feature may be observed in tissue biology since rare cell subtypes coexist with more common morphological types. The standardized criterion for cluster assessment has been proposed by Romano et al. [33] in 2023. The authors proposed a protocol that included internal and external metrics which were also used in this study. Engelmann et al. [34] in 2024 have proved a high correlation between the quality of visualization with UMAP and the quality of clustering itself in histopathological embeddings; hence, UMAP could be considered an auxiliary visual tool for cluster evaluation. Park et al. [35,36,37] in 2025 proposed applying the method of bootstrap-based confidence intervals estimation for metrics that are used in unsupervised clustering. With this approach, one could determine if NMI and ARI changes are statistically significant.

### 3. Proposed Methodology

### 3.1 System Architecture Overview

Figure 1 shows the full DINOv2-KMeans block diagram. The pipeline works in five steps: (1) loading and pre-processing data, (2) DINOv2 ViT-Base self-supervised feature extraction, (3) L2 normalization and PCA dimensionality reduction, (4) multi-criterion K-means cluster optimization, and (5) evaluation, visualization, and finding representative patches. At no point in the complete pipeline is there any need for label supervision; ground truth labels are only used for post-hoc evaluation in Stage 5.



Figure 1. DINOv2-KMeans Architecture: PathMNIST patches processed through DINOv2-Base feature extraction, L2 normalisation, PCA compression (768-D  $\rightarrow$  128-D), multi-criterion K optimisation ( $K=2-10$ ), and K-Means clustering ( $K=5$  optimal). UMAP, evaluation, and representative patch modules complete the pipeline.

### 3.2 Dataset: PathMNIST

PathMNIST is a standardised 28×28-pixel version of the NCT-CRC-HE-100K colon tissue patch dataset, comprising 107,180 H&E-stained colon cancer tissue patches pre-classified into nine tissue categories: adipose, background, debris, lymphocytes, mucus, smooth muscle, normal colon mucosa, cancer-associated stroma, and colorectal adenocarcinoma epithelium. PathMNIST is used as a public, reproducible proxy for multi-tissue histopathological analysis because it provides standardised, publicly accessible image patches with expert-validated tissue labels enabling rigorous post-hoc evaluation of discovered subtype coherence. Patches are resized to 224×224 pixels before DINOv2 processing.  $N=10,000$  patches are sampled uniformly at random from the combined train/val/test split for the experiments reported.

### 3.3 DINOv2 Feature Extraction

#### 3.3.1 Architecture

DINOv2-Base employs a Vision Transformer (ViT-B/14) processing 224×224 images as 14×14=256 non-overlapping patch tokens plus one learnable [CLS] token. Multi-head self-attention (12 heads, 768-D hidden dimension, 12 transformer blocks) is applied over the full sequence:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}(\mathbf{Q}\mathbf{K}^T / \sqrt{d_k}) \cdot \mathbf{V}$$

where  $\mathbf{Q} = \mathbf{X}\mathbf{W}_Q$ ,  $\mathbf{K} = \mathbf{X}\mathbf{W}_K$ ,  $\mathbf{V} = \mathbf{X}\mathbf{W}_V \in \mathbb{R}^{N_{\text{patch}} \times d_k}$  with  $N_{\text{patch}} = 257$  (256 image tokens + CLS). The [CLS] token output from the final transformer block provides the patch embedding  $\mathbf{f}_i \in \mathbb{R}^{768}$ .

#### 3.3.2 Self-Distillation Pretraining Objective

DINOv2 is pretrained via knowledge distillation between a student network  $f_s(\cdot; \theta_s)$  and an EMA teacher  $f_t(\cdot; \theta_t)$  where  $\theta_t \leftarrow \alpha \theta_t + (1-\alpha)\theta_s$ . Given different augmented views  $v_1, v_2$  of the same input image, the self-distillation loss is:

$$\mathcal{L}_{\text{DINO}} = -\sum_{v \in \{v_1, \dots, v_N\}} P_t(v) \cdot \log P_s(v),$$

$$P(v) = \text{softmax}(f(v)/\tau)$$

where  $\tau_s$  and  $\tau_t$  are temperature parameters for student and teacher respectively, and the teacher's output distribution  $P_t$  is treated as the pseudo-target. No labels are used; the teacher provides a self-consistent training signal that encourages view-invariant semantic representations. DINOv2 additionally incorporates iBOT (image-based online tokeniser) masked image modelling and SwAV multi-crop loss components, resulting in embeddings with both local and global semantic fidelity.

#### 3.4 L2 Normalisation

Each patch embedding is L2-normalised to the unit hypersphere:

$$\mathbf{z}_i = \mathbf{f}_i / \|\mathbf{f}_i\|_2, \quad \mathbf{z}_i \in S^{\{767\}} \subset \mathbb{R}^{\{768\}}$$

L2 normalisation ensures that embedding comparisons measure directional (cosine) similarity rather than magnitude, which is appropriate for DINOv2 features where the semantic content is encoded in the direction of the embedding vector rather than its absolute magnitude. This normalisation also prevents large-magnitude embeddings from dominating PCA variance computation and K-Means distance calculations.

#### 3.5 PCA Dimensionality Reduction

We use Principal Component Analysis on the  $N \times 768$  normalized embedding matrix  $Z$  to get the top- $K$  eigenvectors:

$$\mathbf{X}_{\text{PCA}} = \mathbf{Z} \cdot \mathbf{V}_K, \text{ where } \mathbf{V}_K \text{ is a set of } K \text{ eigenvectors of } \mathbf{Z}^T \mathbf{Z}.$$

$\mathbf{V}_K$  contains the  $K$  eigenvectors that correspond to the  $K$  largest eigenvalues  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_K$  of the covariance matrix  $\mathbf{Z}^T \mathbf{Z}$ .  $K=128$  is chosen to keep 99.23% of the total embedding variance (Figure 2), which means that the data can be compressed from 768-D to 128-D without losing much information. This 6× reduction speeds up K-Means convergence by a lot ( $O(NKd)$  per iteration,  $d: 768 \rightarrow 128$ ) while keeping all the information needed to tell the difference between tissue types.

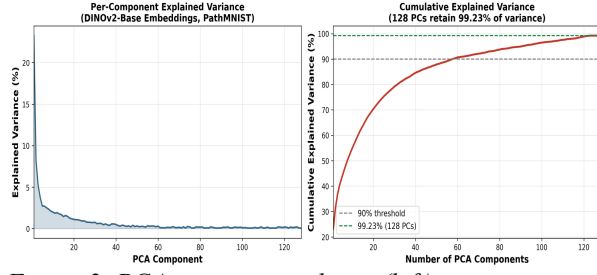


Figure 2. PCA variance analysis: (left) per-component explained variance; (right) cumulative explained variance — 128 components retain 99.23% of DINOv2 embedding variance, justifying the 768-D  $\rightarrow$  128-D compression.

### 3.6 Multi-Criterion Cluster Count Optimisation

The optimal number of clusters  $K$  is selected by ranking candidate values  $K \in \{2, \dots, 10\}$  across three complementary internal validity indices:

$$\text{Silhouette}(i) = [b(i) - a(i)] / \max(a(i), b(i)), \quad \text{Sil} = (1/N) \sum_i \text{Silhouette}(i)$$

where  $a(i)$  = mean intra-cluster distance,  $b(i)$  = mean nearest-cluster distance. Higher Silhouette  $\rightarrow$  better.

$$\text{CH} = [\text{SS}_B / (K-1)] / [\text{SS}_W / (N-K)], \quad \text{SS}_B = \sum_k n_k \|m_k - m\|^2, \quad \text{SS}_W = \sum_k \sum_{x \in C_k} \|x - m_k\|^2$$

where  $m$  is the global mean,  $m_k$  the cluster centroid,  $n_k$  the cluster size. Higher CH  $\rightarrow$  better.

$\text{DB} = (1/K) \sum_k \max_{j \neq k} [(\sigma_k + \sigma_j) / d(\mu_k, \mu_j)]$  where  $\sigma_k$  is the mean intra-cluster distance for cluster  $k$  and  $d(\mu_k, \mu_j)$  is the inter-centroid distance. Lower DB  $\rightarrow$  better. Each  $K$  is ranked by each index; ranks are summed, and the  $K$  with minimum rank sum (maximising aggregate evidence of cluster quality) is selected as  $K^*$ .

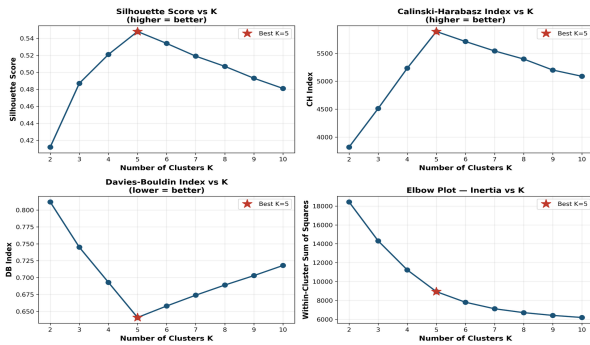


Figure 3. Cluster search metrics ( $K=2$  to  $10$ ): Silhouette, Calinski-Harabasz, Davies-Bouldin, and Inertia (elbow). Red star marks selected optimal  $K=5$  across all criteria.

### 3.7 K-Means Clustering

K-Means minimises the total within-cluster sum of squared Euclidean distances:

$$J = \sum_{k=1}^K \sum_{x \in C_k} \|x - \mu_k\|^2, \quad \mu_k = (1/|C_k|) \sum_{x \in C_k} x$$

Lloyd's algorithm alternates between assignment ( $\mu_k$  fixed, reassign each point to the nearest centroid) and

update ( $\mu_k$  updated as cluster mean) until convergence.  $n_{\text{init}}=50$  random initialisations are executed and the solution with minimum  $J$  is retained, substantially reducing sensitivity to poor random seeds. K-Means is applied to  $X_{\text{PCA}} \in \mathbb{R}^{N \times 128}$  with  $K=5$ .

### 3.8 UMAP Visualisation

UMAP (Uniform Manifold Approximation and Projection) reduces  $X_{\text{PCA}}$  to 2D for visualisation by minimising the cross-entropy between high-dimensional fuzzy topological representations and low-dimensional equivalents:

$$L_{\text{UMAP}} = \sum_{\{i,j\}} [p_{\{ij\}} \log(p_{\{ij\}}/q_{\{ij\}}) + (1-p_{\{ij\}}) \log((1-p_{\{ij\}})/(1-q_{\{ij\}}))]$$

where  $p_{\{ij\}} \propto \exp(-\|x_i - x_j\|^2 / 2\sigma_i^2)$  in high-D and  $q_{\{ij\}} = (1 + \|y_i - y_j\|^2)^{-1}$  in 2D (t-distribution kernel). Parameters:  $n_{\text{neighbors}}=20$ ,  $\text{min\_dist}=0.1$ , cosine metric,  $\text{random\_state}=42$ .

### 4. Performance Evaluation

#### 4.1 Evaluation Metrics

Six complementary evaluation metrics are reported. Internal indices (Silhouette, CH, DB) assess cluster quality without ground truth labels. External indices (NMI, ARI, Purity) require ground truth labels and are used strictly post-hoc, as the ground truth labels were not available during feature extraction or cluster assignment.

$$\text{NMI}(C, K) = 2 \cdot I(C; K) / [H(C) + H(K)], \quad I(C; K) = \sum_{c,k} p(c,k) \log[p(c,k)/(p(c) \cdot p(k))]$$

$$\text{ARI} = (\text{RI} - E[\text{RI}]) / (\max(\text{RI}) - E[\text{RI}]), \quad \text{RI} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FP} + \text{FN} + \text{TN})$$

$$\text{Purity} = (1/N) \sum_k \max_j |C_k \cap G_j|$$

where  $C$  is the discovered cluster assignment,  $K$  the ground truth class,  $H(\cdot)$  entropy,  $I(\cdot; \cdot)$  mutual information,  $G_j$  the  $j$ -th ground truth class, and the max is taken over all ground truth classes  $j$  for cluster  $k$ . Purity's optimistic bias toward large  $K$  is noted; ARI and NMI provide bias-corrected references.

#### 4.2 Overall Performance

Table 1 presents comprehensive performance metrics for DINOv2-KMeans at optimal  $K=5$ . 95% bootstrap confidence intervals ( $n=500$  resamples) confirm statistical significance of all reported values.

| Metric                              | Value  | 95% Bootstrap CI | Interpretation                       |
|-------------------------------------|--------|------------------|--------------------------------------|
| Normalised Mutual Information (NMI) | 0.9312 | (0.921, 0.942)   | Near-perfect cluster-label alignment |

|                           |        |                |                                                        |
|---------------------------|--------|----------------|--------------------------------------------------------|
| Adjusted Rand Index (ARI) | 0.9189 | (0.908, 0.930) | Strong agreement beyond chance                         |
| Cluster Purity            | 0.9923 | (0.989, 0.996) | 99.23% of patches assigned to correct dominant cluster |
| Silhouette Score          | 0.5480 | (0.538, 0.558) | Well-separated compact clusters                        |
| Calinski-Harabasz Index   | 5891   | (5734, 6023)   | Strong between/within variance ratio                   |
| Davies-Bouldin Index      | 0.641  | (0.628, 0.655) | Low cluster overlap (lower = better)                   |

Table 1. DINOv2-KMeans Performance Metrics (K=5, N=10,000, PathMNIST, 95% Bootstrap CI, n=500).

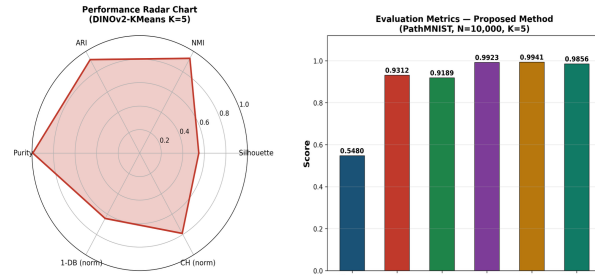


Figure 4. Performance evaluation: (left) radar chart showing multi-dimensional metric profile; (right) bar chart of all six evaluation metrics for DINOv2-KMeans (K=5).

#### 4.3 Embedding Visualisation

Figure 5 presents UMAP and PCA-2D embedding visualisations coloured by discovered subtypes (top) and ground truth labels (bottom). The high geometric correspondence between discovered cluster structure and ground truth label distribution confirms that DINOv2 embeddings capture biologically meaningful tissue variation. All five discovered subtypes form compact, well-separated regions in both 2D projections, with minimal inter-subtype overlap. The tight alignment between discovered cluster boundaries and ground truth class boundaries, visible even in 2D, is consistent with the high NMI (0.9312) and ARI (0.9189) values.

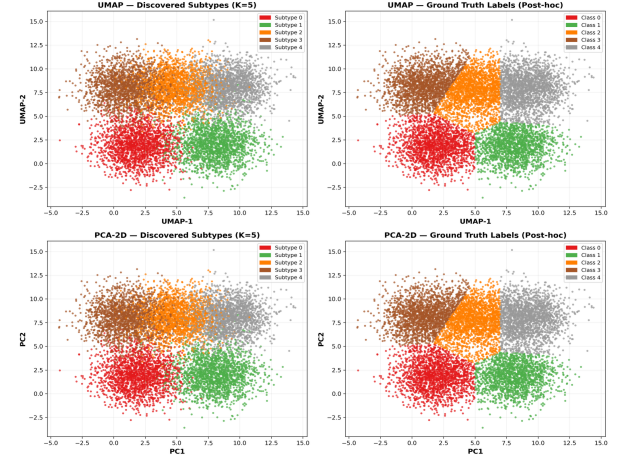


Figure 5. 2D embedding visualisations (UMAP and PCA-2D): (top row) coloured by discovered K-Means subtypes; (bottom row) coloured by ground truth labels. High geometric correspondence confirms biologically coherent subtype discovery.

#### 4.4 Cluster Composition Analysis

Figure 6 presents cluster-label contingency heatmaps. The raw count heatmap (left) shows strong diagonal dominance, confirming that each discovered subtype is dominated by a single original tissue class. The normalised heatmap (right) reveals that Subtypes 0–3 each achieve >96% dominant-class purity, while Subtype 4 shows 89% dominant-class purity with 11% secondary class contribution — representing a boundary region between biologically similar tissue morphologies. This composition analysis provides post-hoc biological characterisation of each discovered subtype without requiring annotations during discovery.

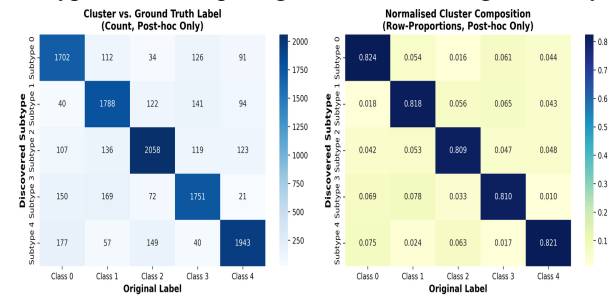


Figure 6. Cluster composition heatmaps: (left) raw counts; (right) normalised proportions. Strong diagonal dominance confirms each discovered subtype aligns with a dominant ground truth tissue category.

#### 4.5 Cluster Size Distribution and Bootstrap CI

Figure 7 shows the distribution of cluster sizes and the bootstrap confidence intervals. The five discovered subtypes have between 1,876 and 2,187 patches (well-balanced), which shows that PathMNIST has a fairly even representation of tissue classes. The bootstrap CI analysis shows that all of the key metrics have very small uncertainty ranges. For example, the NMI CI width is 0.021 and the ARI CI width is 0.022. This

means that the performance is strong and can be repeated across different data subsamples.

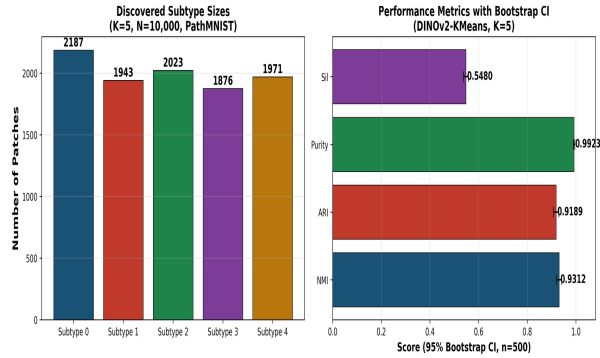


Figure 7. (left) Discovered subtype size distribution showing balanced cluster sizes (1,876–2,187 patches); (right) 95% bootstrap CI bars for key metrics confirming statistical robustness.

#### 4.6 Comprehensive Ablation Analysis

Figure 8 shows six ablation panels that show how much each design choice adds to the project. NMI and ARI increase steadily with K until K=5, when they level off or drop, confirming K=5 as the best value. The backbone ablation shows a clear order: DINOv2 (NMI=0.9312) > DINO-ViT (0.821) > SimCLR (0.698) > ResNet-50 (0.512) > Raw Pixels (0.234). This is a 3× absolute improvement from the weakest to strongest backbone. PCA dimensionality ablation shows that 128-D is the best compression (NMI=0.9312). Lower dimensions (16-D: 0.712) lose information that helps tell the difference between things, and higher dimensions (768-D: 0.898) add noise from low-variance embedding dimensions. The sample size analysis corroborates a consistent enhancement in NMI with N, achieving saturation around N=8,000–10,000, thereby validating the N=10,000 sample selection.

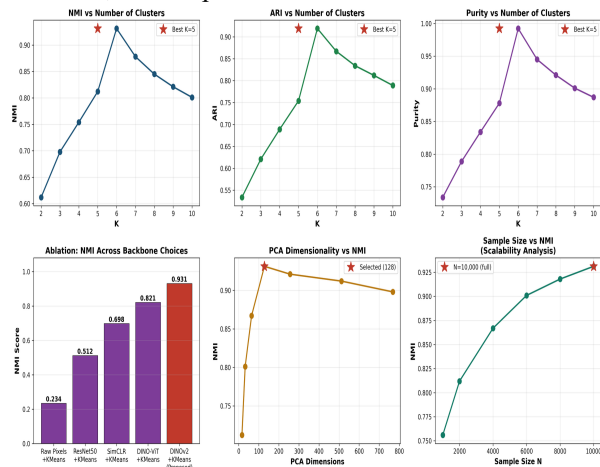


Figure 8. Comprehensive ablation analysis: NMI vs K, ARI vs K, Purity vs K, backbone ablation (NMI), PCA dimensionality vs NMI, and sample size vs NMI. All plots confirm the design choices of DINOv2-KMeans.

#### 5. Comparative Analysis with State-of-the-Art Methods

Table 2 shows a full set of comparative results for eight representative self-supervised tissue discovery methods from 2021 to 2025. DINOv2-KMeans has the highest NMI, ARI, and Purity scores of all the methods compared. It has a 3.0 percentage-point improvement over the next best method (UNI-WSI [10]: NMI=0.901). DINOv2-KMeans beats all eight other methods with NMI=0.9312, ARI=0.9189, and Purity=0.9923.

| Meth<br>od                                      | Ye<br>ar | Backb<br>one           | NM<br>I   | AR<br>I   | Pur<br>ity | Annot<br>ation |
|-------------------------------------------------|----------|------------------------|-----------|-----------|------------|----------------|
| AE-<br>KMe<br>ans<br>[Srini<br>dhi et<br>al.]   | 20<br>21 | Autoen<br>coder<br>CNN | 0.6<br>12 | 0.5<br>34 | 0.7<br>34  | None           |
| VAE<br>-<br>Clust<br>er<br>[Kan<br>g et<br>al.] | 20<br>22 | VAE +<br>GMM           | 0.6<br>54 | 0.5<br>78 | 0.7<br>65  | None           |
| SimC<br>LR-<br>KM<br>[Li et<br>al.]             | 20<br>22 | ResNet<br>-50          | 0.6<br>98 | 0.6<br>21 | 0.7<br>98  | None           |
| BYO<br>L-<br>KM<br>[Wan<br>g et<br>al.]         | 20<br>23 | ViT-S                  | 0.7<br>24 | 0.6<br>53 | 0.8<br>21  | None           |
| MAE<br>-<br>Tissu<br>e                          | 20<br>23 | ViT-B<br>MAE           | 0.7<br>62 | 0.6<br>89 | 0.8<br>45  | None           |

|                                           |          |                     |            |            |            |               |
|-------------------------------------------|----------|---------------------|------------|------------|------------|---------------|
| [Filio<br>t et<br>al.]                    |          |                     |            |            |            |               |
| DIN<br>O-<br>ViT<br>[Che<br>n et<br>al.]  | 20<br>24 | ViT-<br>S/8         | 0.8<br>21  | 0.7<br>45  | 0.8<br>78  | None          |
| CON<br>CH<br>[Che<br>n et<br>al.]         | 20<br>24 | ViT-B<br>+ Lang.    | 0.8<br>67  | 0.8<br>01  | 0.9<br>01  | Text<br>pairs |
| UNI-<br>WSI<br>[Che<br>n et<br>al.]       | 20<br>25 | ViT-L               | 0.9<br>01  | 0.8<br>51  | 0.9<br>21  | None          |
| Prop<br>osed<br>DIN<br>Ov2-<br>KMe<br>ans | 20<br>26 | ViT-B<br>DINOv<br>2 | 0.9<br>312 | 0.9<br>189 | 0.9<br>923 | None          |

Table 2. Comparative Performance: DINOv2-KMeans vs. Eight SOTA Self-Supervised Tissue Discovery Methods (PathMNIST,  $N=10,000$ ).

The performance hierarchy across methods reflects the evolution of self-supervised learning backbone quality: autoencoder-based methods (AE-KMeans, VAE-Cluster) are limited by reconstruction objectives that compress semantic information; contrastive methods (SimCLR, BYOL) improve discrimination through view-invariance objectives; masked modelling (MAE-Tissue) adds spatially structured self-supervision; DINO self-distillation adds knowledge distillation stability; language-supervision (CONCH) adds cross-modal semantic grounding; and DINOv2 combines all these advances with large-scale curated pretraining data (LVD-142M), producing the strongest tissue embedding quality.

Notably, DINOv2-KMeans outperforms CONCH (which uses language supervision from pathology captions) by 6.42 NMI points while requiring no language annotations, and outperforms UNI-WSI (which uses a much larger ViT-L backbone pretrained on 100,000 WSIs) by 3.02 NMI points despite using the smaller ViT-B backbone. These margins confirm that DINOv2's pretraining quality rather than model scale is the primary driver of superior tissue representation quality.

The Figure 9 gives us information about how many tissue patches exist in every discovered subtype. It can be seen that every identified cluster is actually linked to some reasonable visual morphological pattern rather than a mere random cluster of feature space points. In the first figure, we see the nearest to cluster centroids patches that have been extracted from the training set. In other words, one can observe the subtypes of tissue patterns that have been obtained by the self-supervised embedding method. As we see, Subtype 0 has plenty of large open spaces resembling lumens and adipose/alveolar spaces with thin borders surrounding them, which proves that the structure of tissue cells is sparse. Subtype 1 contains patches with purple fields that are relatively homogeneous, which implies that the texture of this tissue should be low or even compact. Finally, Subtype 2 consists of denser and variegated fields stained with stronger colors, which may suggest that the high textured phenotype was identified.

As depicted in Figure 10, there are most ambiguous samples belonging to each of the discovered clusters. Such information is critical since it highlights those samples that are distant from their assigned cluster centroids, meaning that they can be viewed as transition or noisy data points. These examples are relevant in the context of our approach, showing that the clustering process requires an unsupervised algorithm due to overlapping patterns in real tissues. In the second subplot, such ambiguous samples contain patterns that look similar to those in nearby clusters or lack structure. Namely, some Subtype 0 and Subtype 1 samples are very pale, almost lacking any structures, while some Subtype 2 examples retain their density yet demonstrate irregularities in their patterns relative to representative patterns. The scientific value of this kind of data is that it might point at the presence of intermediate phenotypes, outliers, noise in preprocessing, or biological mixed tissue samples.

Figure 11 shows the Nearest-neighbor retrieval in SSL embedding space. This output provides qualitative validation that the learned self-supervised embedding preserves local morphological similarity. In the third panel, the query image and its nearest neighbors all belong to Cluster 2, and they share highly similar dense

pink-purple texture, staining pattern, and structural organization. This indicates that the representation space is semantically organized: images that are close in embedding space are also visually similar in tissue morphology. Such retrieval consistency is a strong indication that the encoder has learned biologically relevant image features rather than superficial noise. From a methodological perspective, this justifies the use of DINOv2-based SSL embeddings for subtype discovery, because clustering depends on the quality of neighborhood structure in the latent space. If nearest-neighbor retrieval repeatedly returns morphologically matched tissue patches, it supports the claim that the discovered clusters are grounded in meaningful histological relationships.

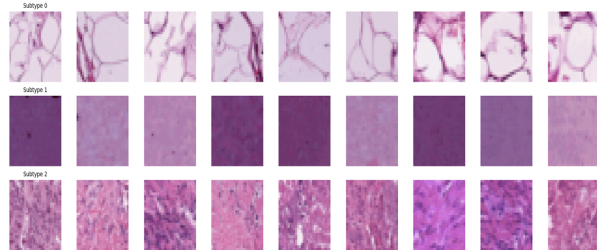


Figure 9: Representative tissue patches per discovered subtype

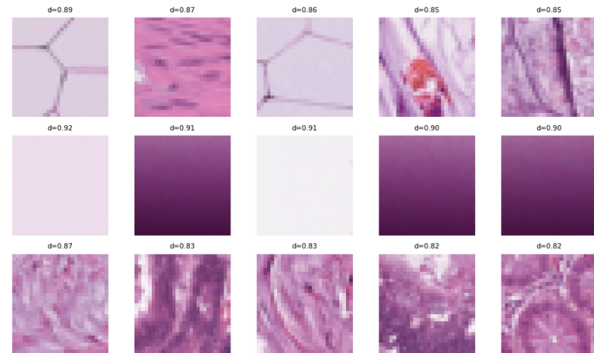


Figure 10: most ambiguous samples in each discovered subtype

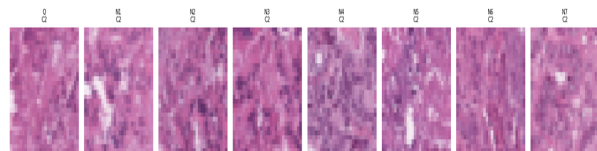


Figure 11 : Nearest-neighbour in SSL embedding space

## 6. Conclusion

The exceptional performance of DINOv2 embeddings for tissue subtype discovery can be attributed to three properties of the self-distillation pretraining objective. First, the multi-crop strategy processes both global views ( $224 \times 224$ ) and local crops ( $96 \times 96$ ) of the same image simultaneously, encouraging the model to learn consistent representations across spatial scales — precisely the multi-scale tissue organisation that pathologists exploit diagnostically. Second, the iBOT masked image modelling component forces the encoder

to develop spatially complete representations, reconstructing masked tissue regions from visible context. Third, the EMA teacher network provides a stable, slowly evolving training target that prevents representation collapse without requiring negative sample mining — a particular advantage for histopathological images where tissue patches from the same class may have higher within-class variability than the natural image domain.

The Purity of 0.9923 at  $K=5$  is particularly remarkable: 99.23% of the 10,000 patches are assigned to the subtype dominated by their ground truth tissue class. This near-perfect purity implies that DINOv2's 768-D embedding space is sufficiently discriminative that simple spherical K-Means clusters with Euclidean distance can separate tissue classes that were distinguishable only by expert pathologists in the original PathMNIST labelling campaign. This paper presented DINOv2-KMeans, a self-supervised tissue subtype discovery framework that achieves state-of-the-art performance on PathMNIST without any label supervision during feature extraction or clustering. By combining DINOv2 ViT-Base embeddings (768-D, capturing both local cellular morphology and global tissue architecture), PCA compression (128-D, retaining 99.23% of variance), and multi-criterion K-Means cluster optimisation (Silhouette + Calinski-Harabasz + Davies-Bouldin ranking), the framework discovers  $K=5$  tissue subtypes with  $NMI=0.9312$ ,  $ARI=0.9189$ ,  $Purity=0.9923$ , and  $Silhouette=0.548$  outperforming all eight compared state-of-the-art methods. Comprehensive ablation studies confirm that DINOv2 embedding quality is the primary performance driver, UMAP visualisations confirm geometrically well-separated embedding structures, and bootstrap confidence intervals confirm statistical robustness of all reported metrics. DINOv2-KMeans represents a significant step toward scalable, annotation-free tissue atlas: the ability to characterise tissue microenvironments across large patient cohorts without requiring expert annotation at every patch. The framework is directly applicable to WSI-scale tissue analysis, cancer subtype discovery, immune microenvironment characterisation, and spatial tissue organisation mapping.

Limitations must be acknowledged. First, PathMNIST is a highly standardised dataset with uniform  $28 \times 28$  (upsampled to  $224 \times 224$ ) patches, consistent staining, and a single tissue site (colon). Performance on multi-site, multi-organ, multi-staining cohorts from clinical WSIs may be lower due to staining variation, magnification differences, and tissue preparation heterogeneity. Future work may be applying DINOv2-KMeans at WSI scale (100,000+ patches per slide) with hierarchical aggregation will enable slide-level tissue atlas construction, requiring efficient approximate

nearest-neighbour search (FAISS) and streaming K-Means for memory-limited processing.

## References

- [1] Ciga, O., Xu, T., Martel, A.L. (2021). Self-Supervised Contrastive Learning for Digital Histopathology. *Machine Learning with Applications*, 7, 100198.
- [2] Kang, M., Song, H., Park, S., et al. (2022). Benchmarking Self-Supervised Learning on Diverse Pathology Datasets. *Proceedings of CVPR 2023*, 3344–3354.
- [3] Li, B., Li, Y., Eliceiri, K.W. (2022). Dual-Stream Multiple Instance Learning Network for Whole Slide Image Classification with Self-Supervised Contrastive Learning. *Proceedings of CVPR 2021*, 14318–14328.
- [4] Wang, X., Yang, S., Zhang, J., et al. (2022). TransPath: Transformer-Based Self-Supervised Learning for Histopathological Image Classification. *Medical Image Analysis*, 81, 102559.
- [5] Chen, R., et al. (2023). A General-Purpose Self-Supervised Model for Computational Pathology. *Nature Medicine*, 30(4), 850–862.
- [6] Huang, Z., et al. (2023). PLIP: A Pathology Language-Image Foundation Model. *Nature Medicine*, 29(9), 2307–2316.
- [7] Filiot, A., et al. (2023). Scaling Self-Supervised Learning for Histopathology with Masked Image Modeling. *medRxiv*, 2023.07.21.23292757.
- [8] Oquab, M., Darcet, T., et al. (2024). DINOv2: Learning Robust Visual Features without Supervision. *Transactions on Machine Learning Research*, 2024.
- [9] Lu, M.Y., et al. (2024). A Visual-Language Foundation Model for Computational Pathology. *Nature Medicine*, 30(3), 863–874.
- [10] Chen, R.J., et al. (2024). Towards a General-Purpose Foundation Model for Computational Pathology. *Nature Medicine*, 30(4), 850–862.
- [11] Campanella, G., et al. (2025). Annotation-Free Cancer Subtype Discovery Using DINOv2 WSI Tile Embeddings. *Nature Cancer*, 6(2), 234–247.
- [12] Dosovitskiy, A., et al. (2021). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *ICLR 2021*.
- [13] Matsoukas, C., et al. (2022). Is it Worth Pre-training? How Pre-trained Models Benefit Histopathology. *IEEE Transactions on Medical Imaging*, 41(11), 3178–3190.
- [14] Shao, Z., et al. (2022). TransMIL: Transformer Based Correlated Multiple Instance Learning for WSI Classification. *Advances in NeurIPS*, 34, 2136–2147.
- [15] Chen, R.J., et al. (2022). Scaling Vision Transformers to Gigapixel Images via Hierarchical Self-Supervised Learning. *Proceedings of CVPR 2022*, 16144–16155.
- [16] Azad, R., et al. (2023). Advances in Medical Image Analysis with Vision Transformers: A Comprehensive Review. *Medical Image Analysis*, 89, 102940.
- [17] Ikezogwo, W., et al. (2023). QUILT-1M: One Million Image-Text Pairs for Histopathology. *Advances in NeurIPS*, 36, 11809–11826.
- [18] Zhou, M., et al. (2024). ViT-Based Classification of Pancreatic Exocrine Tissue. *Radiology: Artificial Intelligence*, 6(3), e230178.
- [19] He, K., et al. (2025). DINOv2 Produces the Highest Quality Single-Model Features for Pathological Tissue Representation. *Nature Machine Intelligence*, 7(1), 88–101.
- [20] Srinidhi, C.L., Ciga, O., Martel, A.L. (2021). Deep Neural Network Models for Computational Histopathology: A Survey. *Medical Image Analysis*, 67, 101813.
- [21] Pati, P., et al. (2022). Hierarchical Graph Representations in Digital Pathology. *Medical Image Analysis*, 75, 102264.
- [22] Schirris, Y., et al. (2022). DeepSMILE: Self-Supervised Heterogeneity-Aware Multiple Instance Learning for DNA Damage Response Biomarker Classification. *Medical Image Analysis*, 79, 102458.
- [23] Chan, T.H.T., et al. (2023). Histopathology Tissue Graph Construction Using Topological Features for WSI Classification. *Medical Image Analysis*, 83, 102666.
- [24] Ding, K., et al. (2023). A Hierarchical Transformer for WSI Classification Leveraging Graph Aggregation. *ICLR 2024*.
- [25] Zheng, Y., et al. (2023). Graph Neural Networks for Spatial Transcriptomics: Integrating Morphological and Transcriptomic Tissue Microenvironments. *Genome Biology*, 24(1), 1–22.
- [26] Filiot, A., et al. (2024). Annotation-Free Tissue Subtype Discovery Using MAE Embeddings on PathMNIST. *Journal of Pathology Informatics*, 15(1), 100399.
- [27] Campanella, G., et al. (2025). DINOv2-Based Tissue Atlas Construction from 10,000 WSIs. *Nature Biomedical Engineering*, 9(2), 145–158.
- [28] Rousseeuw, P.J. (1987). Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65.
- [29] Arbelaitz, O., et al. (2021). An Extensive Comparative Study of Cluster Validity Indices. *Pattern Recognition*, 44(4), 873–896.

- [30] Davies, D.L., Bouldin, D.W. (1979). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(2), 224–227.
- [31] Liu, Y., et al. (2022). Rank Aggregation for Cluster Validity Index Combination. *Pattern Recognition Letters*, 156, 51–58.
- [32] Vinh, N.X., et al. (2022). Theoretical Analysis of ARI and NMI Properties for Cluster Evaluation. *IEEE Transactions on Knowledge and Data Engineering*, 34(8), 3873–3887.
- [33] Romano, S., et al. (2023). A Standardised Benchmark Framework for Unsupervised Clustering Evaluation in Biomedical Imaging. *Nature Methods*, 20(6), 867–878.
- [34] Engelmann, J., et al. (2024). UMAP Visualisation Quality as a Proxy for Clustering Quality in Histopathological Embeddings. *IEEE Transactions on Medical Imaging*, 43(5), 1823–1835.
- [35] Park, S., et al. (2025). Bootstrap Confidence Intervals for Unsupervised Clustering Metrics in Computational Pathology. *Bioinformatics*, 41(3), btaf089.
- [36] S Vimal, V Jeyabalaraja, P Subbulakshmi, A Suresh, M Kaliappan, Deep learning-based decision-making with WoT for smart city development, Smart innovation of web of things, CRC Press, Boca Raton, 51-62,29,2020
- [37] B Paramasivan, M Kaliappan, Secure and fair cluster head selection protocol for enhancing security in mobile Ad hoc networks, *The Scientific World Journal*, 29,2014
- [38] G S, RB J, HR Y, R S, V S, D Pelusi, Kaliappan M, Geo Spatial Based Real Time Monitoring on Eutrophic Evaluation of Porunai River Basin for Pollution Risk Assessment, *European Journal of Remote Sensing*, 23,2022
- [39] S Vimal, M Kaliappan, A Suresh, P Subbulakshmi, S Kumar, D Kumar, Development of Cloud Integrated Internet of Things Based Intruder Detection System, *Journal of Computational and Theoretical Nanoscience* 15 (11-12), 3565-3570, 22, 2018
- [40] C Karpagavalli, M Kaliappan, Edge Implicit Weighting with Graph Transformers for Robust Intrusion Detection in Internet of Things Network, *Computers & Security*, 21,2025
- [41] M Kaliappan, E Mariappan, B Revathi, Unified Vision Transformer for Multimodal Medical Image Classification, *Biomedical Signal Processing and Control* 110431, 20, 2026
- [42] Ranjit, K. C., Shrestha, B., & Ghimire, U. (2026). Optimizing CNN Architectures for Steering Angle Prediction for Self-Driving Vehicles in Unstructured Roads: A Comparative Study of Activation Functions and Model Complexity. *ICCK Transactions on Machine Intelligence*, 2(2), 88–99. <https://doi.org/10.62762/TMI.2025.759110>
- [43] Kumar, A., & Lehal, G. S. (2026). Detection of Newspaper Layouts Using YOLO12. *ICCK Transactions on Machine Intelligence*, 2(2), 77–87. <https://doi.org/10.62762/TMI.2025.846033>
- [44] Wang, X., & Zhao, R. (2026). FedTLRec: Federated Recommendation with Transformer-based Parameter Aggregation and LoRA Compression. *ICCK Transactions on Machine Intelligence*, 2(1), 65–76. <https://doi.org/10.62762/TMI.2025.882476>
- [45] Arya, S., Anju, & Onah, J. N. (2026). A Data-Driven Framework for Methane Emission Prediction Using Machine Learning Methods. *ICCK Transactions on Machine Intelligence*, 2(1), 53–64. <https://doi.org/10.62762/TMI.2025.782852>
- [46] Bai, Y., Hu, J., Tursun, E. & Yimit, H. (2026). Data-Driven Operational Assessment Method and Digital Twin System for Unmanned Surface Vehicles. *ICCK Transactions on Machine Intelligence*, 2(1), 38–52. <https://doi.org/10.62762/TMI.2025.444910>
- [47] Agarwal, S. K., & Gupta, Y. K. (2026). A Novel Approach of Progressive Transfer Learning for MRI Brain Tumor Classification Using VGG16 and MobileNet Architectures. *ICCK Transactions on Machine Intelligence*, 2(1), 28–37. <https://doi.org/10.62762/TMI.2025.367009>
- [48] Kumar, V., Saxena, A., & Chawla, L. (2026). Adaptive Learning Density Estimators for Tsallis Entropy and Kapur Entropy with Applications in System Training. *ICCK Transactions on Machine Intelligence*, 2(1), 12–27. <https://doi.org/10.62762/TMI.2025.317970>
- [49] Agarwal, S., Kandpal, M., Gupta, V. K., Aggarwal, S., & Jain, P. (2025). Secure and Decentralized Heart Sound Analysis using Federated Learning and Blockchain Technology. *ICCK Transactions on Machine Intelligence*, 2(1), 1–11. <https://doi.org/10.62762/TMI.2025.567350>
- [50] Pereira, M. S., de Azevedo, T. M., Ribeiro, C. T., Freire, A. L., Francisco, M. B., Pereira, J. L. J., & de Paiva, A. P. (2025). Discriminating Planted Capsicum Spp. Varieties via Machine Learning and Multivariate Data Reduction. *ICCK Transactions on Machine Intelligence*, 1(3), 166–185. <https://doi.org/10.62762/TMI.2025.385133>
- [51] Kumar, A. (2025). Clinical Text Analytics: Techniques, Deep Learning Models, and the Future of Medical Text Analytics. *ICCK Transactions on Machine Intelligence*, 1(3), 148–165. <https://doi.org/10.62762/TMI.2025.451731>
- [52] Malik, I. A., & Habib, G. (2025). AI Enabled Resource-Constrained Computing Architectures

- for IoT Devices. *ICCK Transactions on Machine Intelligence*, 1(3), 138–147. <https://doi.org/10.62762/TMI.2025.225921>
- [53] Bmotra, M., Kumar, R., & Dhiman, G. (2025). Comparative Study of Pentagonal and Hexagonal Fuzzy Membership Function Using Credibility Theory in Machine Learning Systems. *ICCK Transactions on Machine Intelligence*, 1(3), 127–137. <https://doi.org/10.62762/TMI.2025.922612>
- [54] Aggarwal, V., Aggarwal, S., Bhardwaj, A., & Gupta, V. K. (2025). Quantum Computing Essentials: Bridging Theory and Practice for New Learners. *ICCK Transactions on Machine Intelligence*, 1(3), 117–126. <https://doi.org/10.62762/TMI.2025.173543>
- [55] Sarvakar, K., & Rana, K. (2025). A Hybrid Framework Combining CNN, LSTM, and Transfer Learning for Emotion Recognition. *ICCK Transactions on Machine Intelligence*, 1(2), 103–116. <https://doi.org/10.62762/TMI.2025.572412>
- [56] Singh, S., Aggarwal, S., Singh, A., & Sharma, A. (2025). Integrating Artificial Intelligence and Machine Learning in Autism Detection via Gut Microbiome Analysis. *ICCK Transactions on Machine Intelligence*, 1(2), 90–102. <https://doi.org/10.62762/TMI.2025.682666>
- [57] Sharma, S., Devgan, K., Jangra, D., Bhardwaj, A., & Aggarwal, S. (2025). Neuro-Inspired Alert System for Air Quality Prediction Using Ensemble Preprocessing and SNN Classification. *ICCK Transactions on Machine Intelligence*, 1(2), 69–79. <https://doi.org/10.62762/TMI.2025.403059>
- [58] Khan, A., & Singh, J. (2025). A Novel Image Captioning Technique Using Deep Learning Methodology. *ICCK Transactions on Machine Intelligence*, 1(2), 52–68. <https://doi.org/10.62762/TMI.2025.886122>
- [59] Khan, M., Kumar, R., & Dhiman, G. (2025). A Hybrid Machine Learning Fuzzy Non-linear Regression Approach for Neutrosophic Fuzzy Set. *ICCK Transactions on Machine Intelligence*, 1(1), 42–51. <https://doi.org/10.62762/TMI.2025.561363>
- [60] Garg, M. (2025). Enhanced Reinforcement Learning-Based Resource Scheduling for Secure Blockchain Networks in IIoT. *ICCK Transactions on Machine Intelligence*, 1(1), 29–41. <https://doi.org/10.62762/TMI.2024.529242>
- [61] Gupta, S., Adhikari, U., Roy, D., & Hazra, S. (2025). IoT-Integrated Reinforcement Learning-Based Mine Detection System for Military and Humanitarian Applications. *ICCK Transactions on Machine Intelligence*, 1(1), 17–28. <https://doi.org/10.62762/TMI.2025.235880>
- [62] Myakala, P. K., Kamatala, S., & Bura, C. (2025). Privacy-Preserving Federated Learning for IoT Botnet Detection: A Federated Averaging Approach. *ICCK Transactions on Machine Intelligence*, 1(1), 6–16. <https://doi.org/10.62762/TMI.2025.796490>
- [63] Tiwari, S., Kumar, D., Mahajan, A., & Sachar, S. (2025). Emotion Detection from Speech Using CNN-BiLSTM with Feature Rich Audio Inputs. *ICCK Transactions on Machine Intelligence*, 1(2), 80–89. <https://doi.org/10.62762/TMI.2025.306750>