

# Benchmarking Deep Learning-based Image Retrieval of Oral Tumor Histology

Ranny Rahaningrum Herdiantoputri<sup>1</sup>, Daisuke Komura<sup>2</sup>, Mieko Ochi<sup>2</sup>, Yuki Fukawa<sup>1</sup>, Kou Kayamori<sup>1</sup>, Maiko Tsuchiya<sup>3</sup>, Yoshinao Kikuchi<sup>3</sup>, Tetsuo Ushiku<sup>4</sup>, Tohru Ikeda<sup>1</sup>, Shumpei Ishikawa<sup>2,5</sup>

<sup>1</sup>Department of Oral Pathology, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, 1-5-45, Yushima, Bunkyo-ku, 1138510, Tokyo, Japan

<sup>2</sup>Department of Preventive Medicine, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, 1130033, Tokyo, Japan

<sup>3</sup>Department of Pathology, Teikyo University School of Medicine, 2-11-1 Kaga, Itabashi-ku, 1738605, Tokyo, Japan

<sup>4</sup>Department of Pathology, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 1130033, Japan

<sup>5</sup>Division of Pathology, National Cancer Center Exploratory Oncology Research & Clinical Trial Center, 6-5-1 Kashiwanoha, Kashiwa, Chiba 277-8577, Japan

Corresponding authors: D. Komura ([kdais-prm@m.u-tokyo.ac.jp](mailto:kdais-prm@m.u-tokyo.ac.jp)) and S. Ishikawa ([ishum-prm@m.u-tokyo.ac.jp](mailto:ishum-prm@m.u-tokyo.ac.jp))

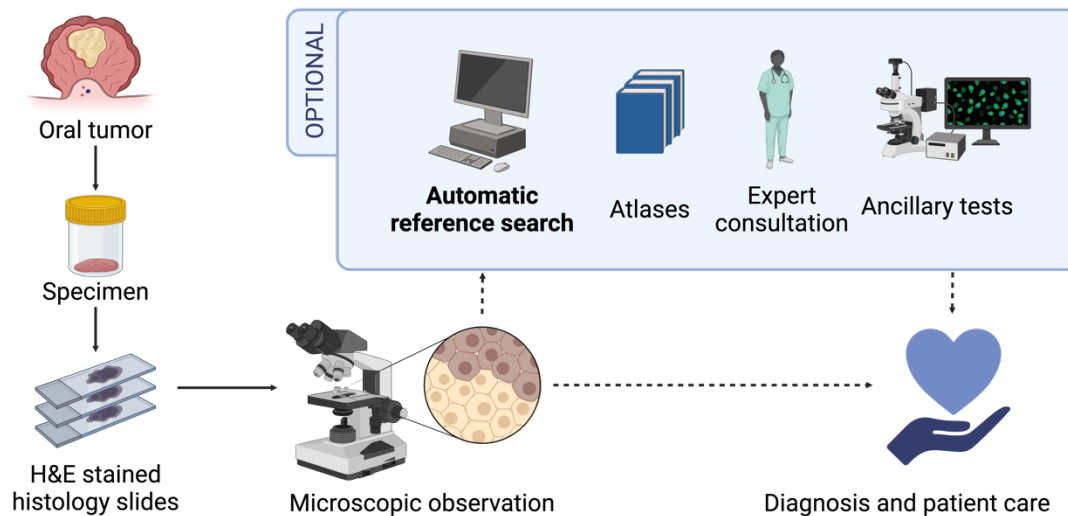
## Abstract

Oral tumors necessitate a dependable computer-assisted pathological diagnosis system considering their rarity and diversity. A content-based image retrieval (CBIR) system using deep neural networks has been successfully devised for digital pathology. No CBIR system for oral pathology has been investigated because of the lack of an extensive image database and feature extractors tailored to oral pathology. This study uses a large CBIR database constructed from 30 categories of oral tumors to compare deep learning methods as feature extractors. The highest average area under the receiver operating curve (AUC) was achieved by models trained on database images using self-supervised learning (SSL) methods (0.900 with SimCLR; 0.897 with TiCo). The generalizability of the models was validated using query images from the same cases taken with smartphones. When smartphone images were tested as queries, both models yielded the highest mean AUC (0.871 with SimCLR and 0.857 with TiCo). We ensured the retrieved image result would be easily observed by evaluating the top-10 mean accuracy and checking for an exact diagnostic category and its differential diagnostic categories. Therefore, training deep learning models with SSL methods using image data specific to the target site is beneficial for CBIR tasks in oral tumor histology to obtain histologically meaningful results and high performance. This result provides insight into the effective development of a CBIR system to help improve the accuracy and speed of histopathology diagnosis and advance oral tumor research in the future.

## Introduction

Oral tumors are generally composed of diverse and rare tumor types, except for major categories like squamous cell carcinoma. Distinguishing oral tumor types is difficult except for well-experienced oral pathologists. The rarity of oral tumors and the diverse tissue types in the oral region make obtaining reference images for diagnosis and research a challenge, potentially leading to delayed diagnosis and a significant burden on pathologists [1]. Consequently, a diagnostic system is needed to improve the speed and accuracy of histopathological diagnosis of these tumors [2]. Artificial intelligence (AI) is a promising solution for efficient histopathological diagnosis of oral tumors.

AI development for oral tumor diagnosis is limited and focused only on a few tumor types. Classification methods have been developed to predict the diagnosis, such as ameloblastoma or odontogenic keratocysts, to which a histopathological image may belong [3,4]. These approaches are helpful in common cases. However, a computer-aided diagnostic system that covers a broader spectrum of tumor types would be more practical and would help narrow the differential diagnoses. Therefore, content-based image retrieval (CBIR) is suitable. CBIR regards histopathological images as query images to find similar images from a database based on their similar morphology [2,5]. This system is useful as a diagnostic aid for finding case references, especially where diagnostic expertise is challenging to find, such as in low- to middle-income countries [1]. The involvement of human intervention is crucial in diagnosis. Conventionally, pathologists diagnose directly after H&E-stained slide analysis or optionally use different methods as diagnostic aids: referring atlases, consulting subspecialist experts, or conducting ancillary tests. An automatic image search can complement these options to expedite image reference search (Figure 1). With scarce pathological expertise, a tool that could provide urgently needed information for rapid diagnosis before conducting tests to raise a definitive one would be significant [6]. CBIR provides interpretability because it presents multiple candidate images, which is beneficial when distinguishing between categories based on histopathological images alone is challenging, such as when information on dental infections or radiographic findings is needed. With CBIR, the retrieved results are to be evaluated by pathologists, reducing the risk of misdiagnosis owing to inaccurate results, especially for categories with very similar histology.



**Figure 1: CBIR's place in the pathological diagnosis workflow**

CBIR optionally providing an interpretable automatic reference search that is fast and easily followed up with more thorough study with atlases, discussions, or ancillary tests. CBIR could help point out similar features from the previously diagnosed image in the database that may lead to testable differential diagnoses more swiftly than directly consulting atlases or senior experts, which may cause the patient delayed treatment. (Image created with BioRender)

The CBIR system consists of two aspects: image feature extraction and nearest-neighbor search. Feature extraction is crucial because it must adequately capture complex histological features such as staining patterns, tissue structures, and cellular morphology to create histologically relevant image representation [2,7,8]. The extracted features must be robust to irrelevant color variations, such as different hematoxylin and eosin (HE) stain brands, glass slide color degradation, and image-capturing devices ranging from whole-slide image (WSI) scanners to smartphone cameras [5,8,9]. At the early stage of CBIR development, traditional image features such as shape, color, texture, or a combination were used. Recent developments showed that deep learning models outperformed traditional features [6,7]. Several deep learning methods, such as supervised learning where models are pre-trained on general images or fine-tuned on histopathological images, have been used to train feature extractors [9–13], and self-supervised learning (SSL), which allows learning from unlabeled images [14–16]. However, no studies have reviewed which method is most suitable for CBIR in oral tumors.

This study aimed to investigate the performance of different deep learning models for oral tumor CBIRs by developing a large dataset of whole-slide images from 541 cases with 51 tumor types

and evaluating the retrieval accuracy by comparing different representational learning techniques.

## Materials & Methods

### Dataset

We collected diagnostic slides of the oral tumor categories described in Chapters 7 and 8 of the WHO Classification of Head and Neck Tumors, 4<sup>th</sup> Edition [17]. Patients were diagnosed in 2001–2022 and underwent surgery at Tokyo Medical and Dental University (TMDU) Hospital. Patients or their surrogates had the option to withdraw from this study through public notices according to the approved protocols. This study was approved by the Institutional Review Board (IRB) of TMDU (No. D2019-087). Some slides that were lost, broken, or required diagnostic confirmation with immunohistochemistry (IHC) staining were remade from the paraffin-embedded tissue blocks. Additional IHC staining was done for the secretory carcinoma and the atypical acinic cell carcinoma cases older than 2017. Categories with fewer than five cases were excluded. All slides that fulfill the inclusion criteria were scanned using a NanoZoomer S210 slide scanner (C13239-01; Hamamatsu Photonics, Japan) at 40× magnification. The tumor areas were annotated by a pathology resident and verified by board-certified pathologists. We included the tumor areas that are typical to the tumor category while excluding the normal tissue and the severe artifacts such as torn or folded tissue. Image patches were then randomly extracted from the annotated tumor areas with three different sizes: 905 μm, 453 μm, and 226 μm. Twenty image patches were extracted with each magnification. The use of three different sizes was to accommodate different magnification levels and preserve the histologic information at the tissue and cellular level as much as possible. The dataset comprises 49,243 image patches from 51 categories, covering approximately 50% of the oral tumor categories (Table 1).

**Table 1. The tumor categories included in the dataset, their corresponding ICD-O codes, and the total number of cases**

Some categories do not correspond to the ICD-O but are described in the WHO Classification of Head and Neck Tumors, 4th Ed. Categories consisting of ten to twenty cases, marked in **bold**, were included in the CBIR database.

Diagnosis	ICD-O Code	Total case
Pleomorphic adenoma	8940/0	27
Radicular cyst	-	21
Dentigerous cyst	-	21
<b>Odontogenic keratocyst</b>	-	<b>17</b>
<b>Salivary duct carcinoma</b>	8500/3	<b>15</b>
<b>Acinic cell carcinoma</b>	8550/3	<b>15</b>
<b>Ossifying fibroma</b>	9262/0	<b>15</b>
Lipoma	8850/0	14
<b>Ameloblastoma</b>	9310/0	<b>14</b>
<b>Adenoid cystic carcinoma</b>	9310/0	<b>13</b>

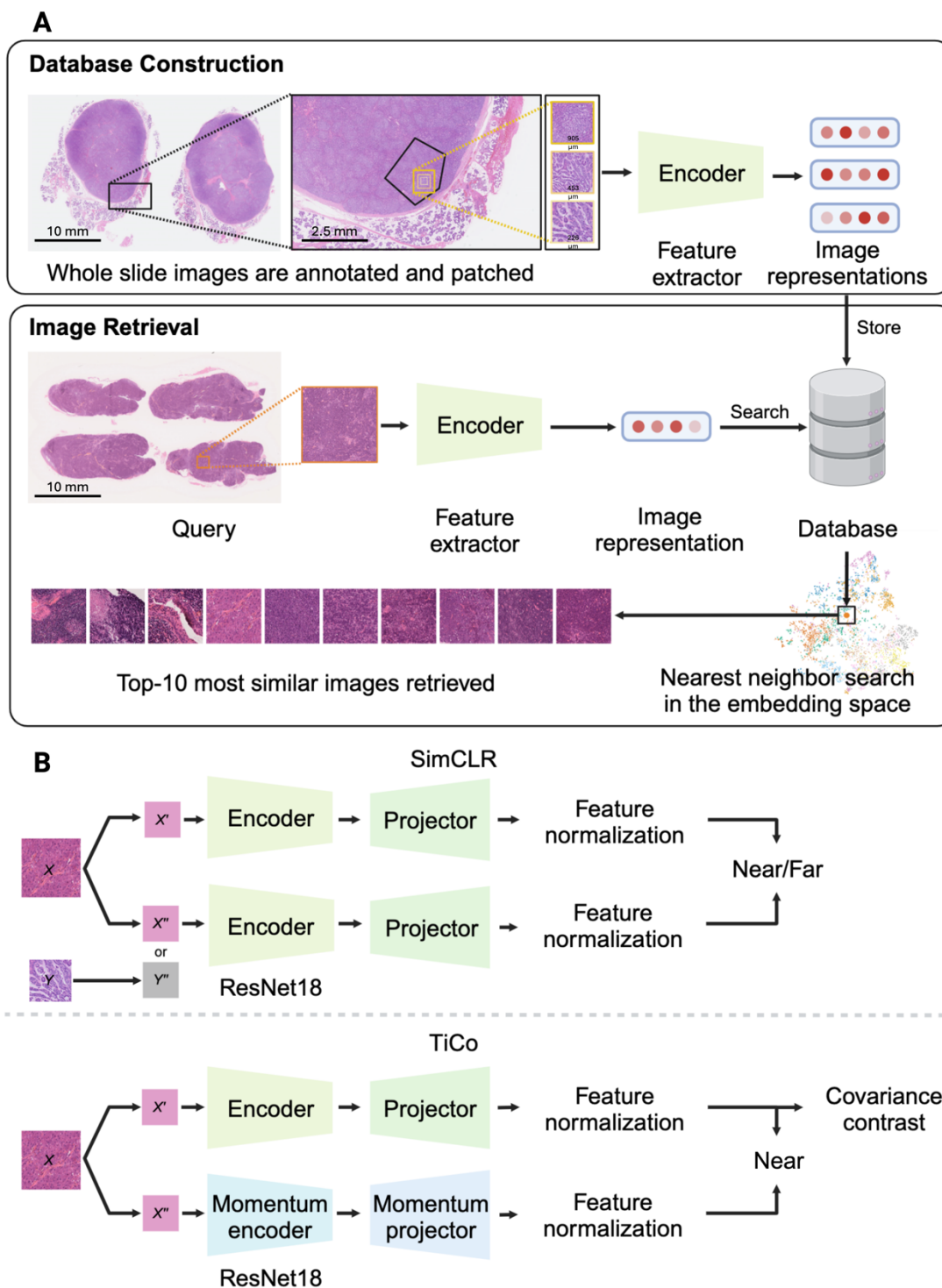
<b>Mucoepidermoid carcinoma</b>	8430/3	<b>13</b>
<b>Fibrous dysplasia</b>	-	<b>13</b>
<b>Adenocarcinoma, NOS</b>	8140/3	<b>12</b>
<b>Myoepithelioma</b>	8982/0	<b>12</b>
<b>Odontogenic fibroma</b>	9321/0	<b>12</b>
<b>Cemento-osseous dysplasia</b>	-	<b>12</b>
<b>Nasopalatine duct cyst</b>	-	<b>12</b>
<b>Hemangioma</b>	9120/0	<b>11</b>
<b>Calcifying odontogenic cyst</b>	9301/0	<b>11</b>
<b>Odontogenic myxoma/myxofibroma</b>	9320/0	<b>11</b>
<b>Ameloblastic fibroma</b>	9330/0	<b>11</b>
<b>Glandular odontogenic cyst</b>	-	<b>11</b>
<b>Inflammatory collateral cyst</b>	-	<b>11</b>
<b>Orthokeratinized odontogenic cyst</b>	-	<b>11</b>
<b>Basal cell adenoma</b>	8147/0	<b>10</b>
<b>Basal cell adenocarcinoma</b>	8147/3	<b>10</b>
<b>Warthin's tumor</b>	8561/0	<b>10</b>
<b>Carcinoma ex pleomorphic adenoma</b>	8941/3	<b>10</b>
<b>Osteoma</b>	9180/0	<b>10</b>
<b>Osteosarcoma, NOS</b>	9180/3	<b>10</b>
<b>Aneurysmal bone cyst</b>	9260/0	<b>10</b>
<b>Cemento-ossifying fibroma</b>	9274/0	<b>10</b>
<b>Adenomatoid odontogenic tumor</b>	9300/0	<b>10</b>
MALT lymphoma	9699/3	9
Ameloblastoma, unicystic type	9310/0	9
Lateral periodontal cyst	-	9
Simple bone cyst	-	8
Cystadenoma	8440/0	8
Epithelial/myoepithelial carcinoma	8562/3	7
Complex odontoma	9282/0	7

Polymorphous adenocarcinoma	8525/3	7
Secretory carcinoma	8502/3	6
Ameloblastoma, extraosseous/peripheral type	-	6
Compound odontoma	9281/0	5
Odontoma	9280/0	5
Oncocytoma	8290/0	5
Cementoblastoma	9273/0	5
Clear cell carcinoma	8310/3	5
Primary intraosseous carcinoma, NOS	9270/3	5
Ductal papilloma	8503/0	5
Myoepithelial carcinoma	8982/3	5

---

### Database construction

A database from a subset of the dataset containing at least ten cases was compiled. Image representations from each model's encoder were stored in the database (Figure 2A). It contains 33,356 image patches from 30 oral tumor categories (Table 1).



**Figure 2: CBIR workflow and SSL models training process**

(A) Schematic representation of CBIR using deep neural networks to retrieve similar oral tumor histopathological images. The similarity is determined by a nearest-neighbor search, which calculates the cosine similarity of the query image's image representation to all database image representations in the embedding space. (B) The training process of the SSL models used ResNet18 as an encoder. The image representations were passed to a projector and subjected to feature normalization. In the SimCLR method, the training loss function yields a low value when the

representation of the original image ( $X'$ ) and its augmentation ( $X''$ ) are close together, while it yields a larger value when  $X'$  and a different image augmentation ( $Y''$ ) are far apart. In the TiCo method, the process still pulls  $X'$  and  $X''$  close, and then the redundancy in the representation is removed using covariance contrast without using a different image ( $Y''$ ). (Image created with BioRender)

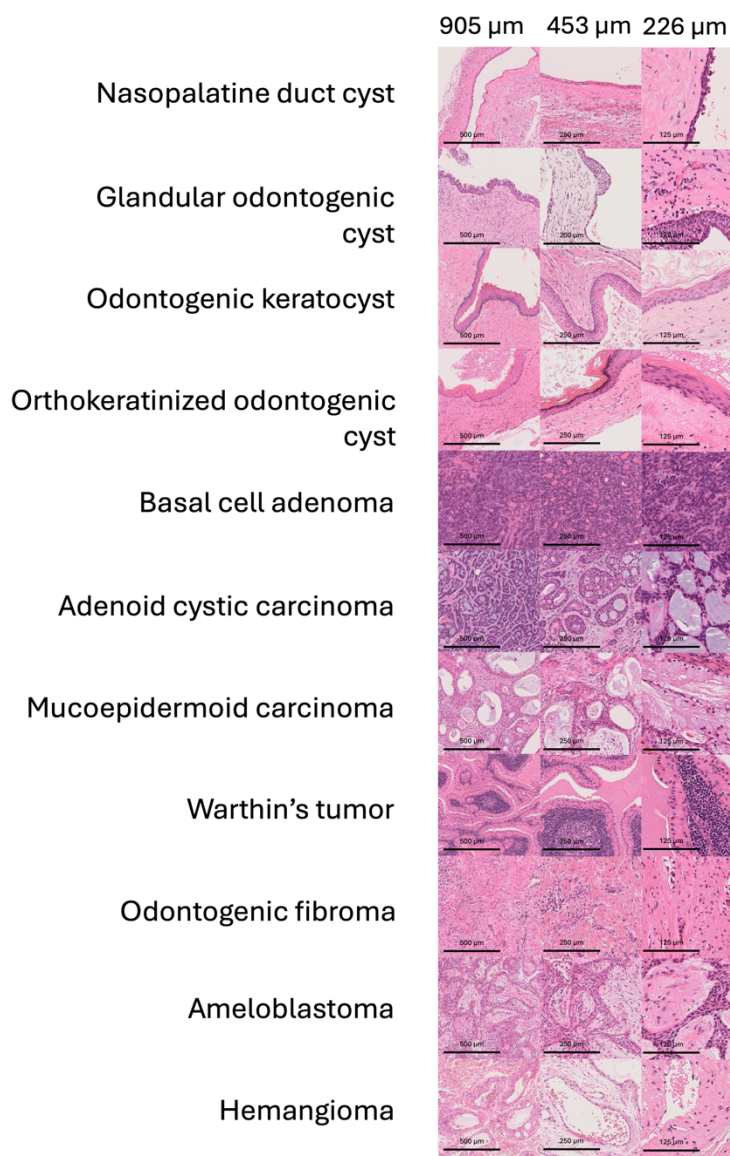
## Test queries

The test queries were the cases available in the hospital repository after the collection of the database case was finished and were representatives of the tumor major categories in the database: odontogenic cysts, odontogenic tumors, benign and malignant tumors of the salivary gland, maxillofacial bone tumor, and soft tissue tumor. Slides that include severe artifacts that are impossible to avoid when extracting image patches were excluded. We prepared three query sets from different hospitals to test the performance: Query case set-A was collected between 2022 and 2023 from the repository of the same institution (TMDU) as the database. Only the tumor categories that have two representative cases were included as the test queries. Finally, eleven tumor categories with two cases in each were used as the test queries. Histopathologic slides were scanned to create WSIs for in-domain queries with the same device as the database image. Three selected tumor areas that are typical of the tumor type from the same slides were photographed with smartphone cameras (Samsung S21FE, iPhone 6, and Motorola g8) using an Olympus BX53 microscope with 10 $\times$ , 20 $\times$ , and 40 $\times$  objective lens magnification to create out-of-domain-phonocam queries. The location (indoor laboratory environment) and the amount of light from the microscope were unchanged when taking the smartphone images. The query case set-B was compiled from the University of Tokyo case (approved by the IRB of The University of Tokyo No. 2019158NI). Eleven cases from eight salivary gland tumor categories were included in this study. Only benign and malignant tumors of the salivary gland could be collected for out-of-domain-B since no other major categories exist in the repository. Histopathologic slides were scanned using a NanoZoomer 2.0HT slide scanner (C9600-12, Hamamatsu Photonics, Japan) to create WSIs for out-of-domain-B queries. The query case set-C was collected from Teikyo University Hospital from 2018-2023 (approved by the IRB of Teikyo University No. 23-054). Only the tumor categories that have two representative cases were included as the test queries. Twenty cases from eight oral tumor categories were included in the study. The histopathologic slides were scanned using a NanoZoomer XR slide scanner (C12000-02; Hamamatsu Photonics, Japan) to create WSIs for out-of-domain-C queries. Patients for query cases set-B and set-C or their surrogates had the option to withdraw from this study through public disclosure according to the approved protocols.

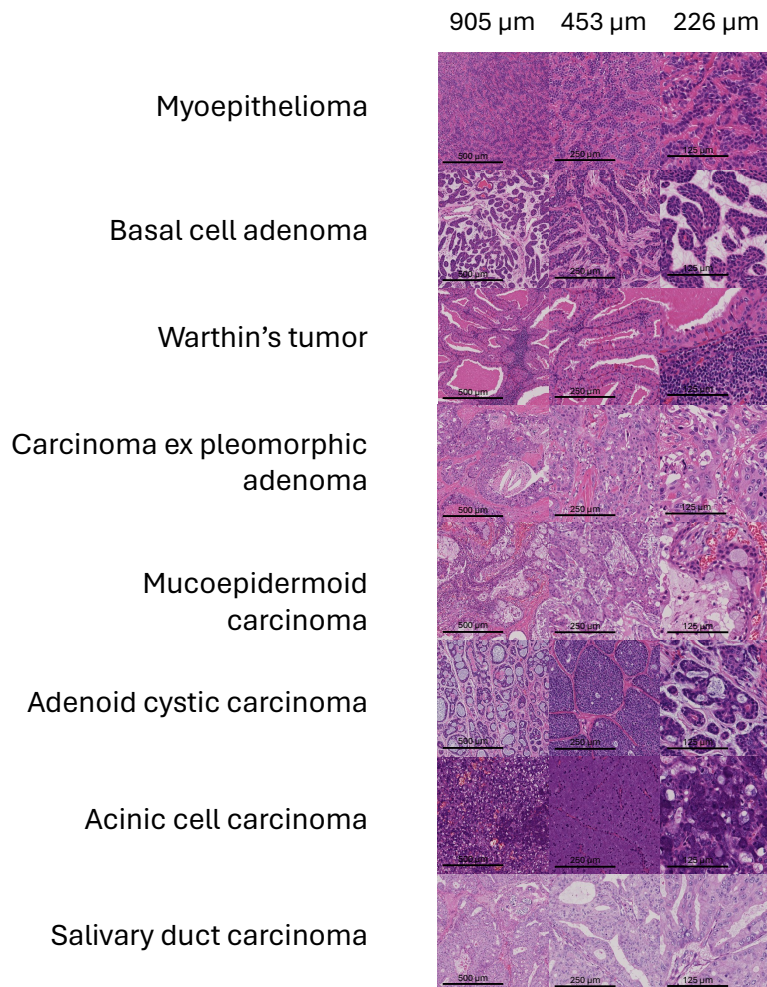
For the WSI queries, the same method was used to create image patches from the WSI as from the database image. All scanned WSIs were annotated to 3–4 representative tumor areas per WSI. From these areas, 20 image patches per magnification level (905  $\mu\text{m}$ , 453  $\mu\text{m}$ , and 226  $\mu\text{m}$ ) per slide were extracted to create image patches. The total number of image patches used for the evaluation was 2,520 images from the WSI and 594 smartphone images.

The representation of each query image was calculated with each tested model. The nearest-neighbor search was performed based on cosine similarity with the database images. Examples of query images for each category in each set can be found in Figures 3–5. The detailed methods for database construction, including the tumor areas selection, patch extraction, feature extraction code, and image retrieval were adapted from our previous study [18].

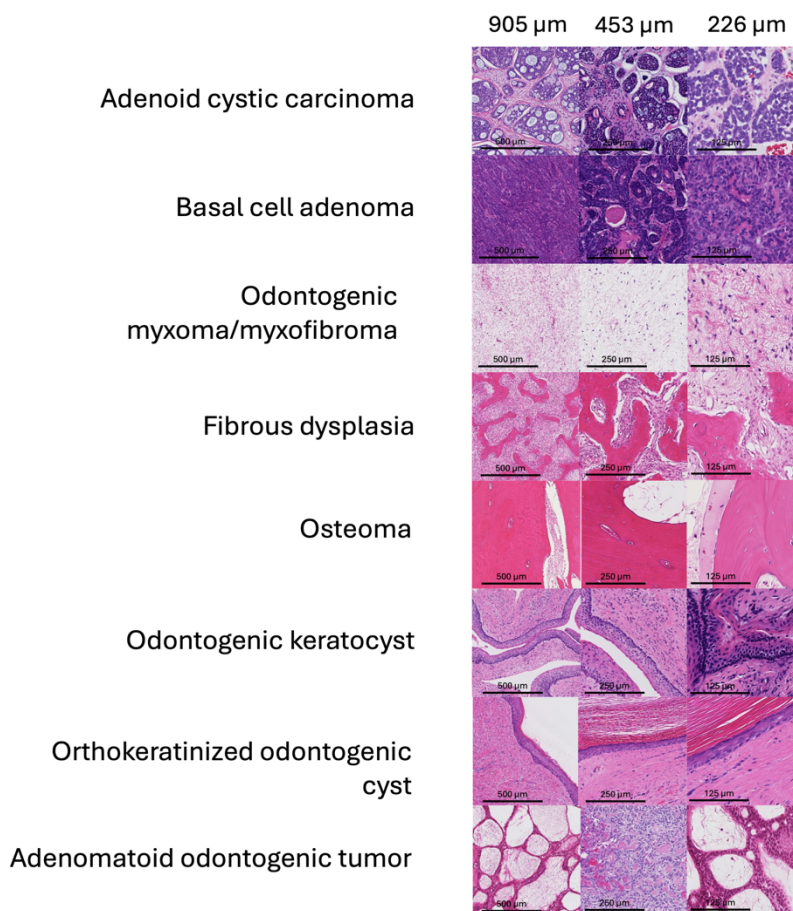




**Figure 3: Examples of set-A query images from each category for each magnification level**  
The total number of image patches extracted from WSIs is 1,320 images.



**Figure 4: Examples of set-B query images from each category for each magnification level**  
The total number of image patches extracted from the WSIs is 660 images.



**Figure 5: Examples of set-C query images from each category for each magnification level**

The total number of image patches extracted from the WSIs is 1,200 images.

### Evaluation metrics and statistical analysis

The area under the receiver operating characteristic curve (AUC) for all query images with top- $k$  retrieved images ( $k$  ranges from 1 to the total number of cases in the database) being the cut-points were averaged into MeanAUC. Based on the top-10 images most similar to the query, three additional metrics were evaluated. MeanAcc denotes the mean of the top-10 diagnostic accuracies (Acc) for each query. %query denotes the percentage of results that contained at least one accurate diagnosis category. The histological similarity in the retrieved results beyond diagnostic accuracy was evaluated by noting the retrieved images that did not belong to the accurate diagnosis category or any of its differential diagnoses [17,19,20] (Table 2). The values

are expressed as histologic inaccuracy (HI) and were averaged to determine the MeanHI. Image retrieval and all statistical analyses were conducted using Python 3.7.12 and R 4.2.2.

**Table 2. List of test query categories and all their respective differential diagnoses that are represented in the database for MeanHI evaluation**

Query Category	Differential Diagnoses		
Acinic cell carcinoma	Salivary duct carcinoma	Secretory carcinoma	Mucoepidermoid carcinoma
Adenoid cystic carcinoma	Basal cell adenocarcinoma	Basal cell adenoma	
Adenomatoid odontogenic tumor	Ameloblastoma	Mucoepidermoid carcinoma	
Ameloblastoma	Adenomatoid odontogenic tumor	Ameloblastic fibroma	
Basal cell adenoma	Adenoid cystic carcinoma	Basal cell adenocarcinoma	Myoepithelioma
Carcinoma ex pleomorphic adenoma	Salivary duct carcinoma	Adenocarcinoma, NOS	Mucoepidermoid carcinoma
Fibrous dysplasia	Ossifying fibroma	Osteosarcoma, NOS	
Glandular odontogenic cyst	Nasopalatine duct cyst	Inflammatory collateral cyst	Mucoepidermoid carcinoma
Hemangioma	Ossifying fibroma	Aneurysmal bone cyst	
Mucoepidermoid carcinoma	Adenocarcinoma, NOS	Warthin's tumor	Carcinoma ex pleomorphic adenoma
Myoepithelioma	Mucoepidermoid carcinoma	Basal cell adenoma	
Nasopalatine duct cyst	Glandular odontogenic cyst	Inflammatory collateral cyst	
Odontogenic fibroma	Odontogenic myxoma/myxofibroma	Ossifying fibroma	
Odontogenic keratocyst	Orthokeratinized odontogenic cyst	Ameloblastoma	
Odontogenic myxoma/myxofibroma	Odontogenic fibroma		
Orthokeratinized odontogenic cyst	Odontogenic keratocyst	Ameloblastoma	
Osteoma	Ossifying fibroma	Fibrous dysplasia	
Salivary duct carcinoma	Acinic cell carcinoma	Secretory carcinoma	Adenocarcinoma, NOS
Warthin's tumor	Mucoepidermoid carcinoma		

## Model preparation and training

### *ImageNet-1k Pre-trained CNN*

We tested VGG16 pre-trained on 1.2 million images from ImageNet after it was shown to perform well as an image classifier in several studies [22,23]. In this study, the block4\_conv3 layer was used as the feature extractor because the middle layer of a convolutional neural network (CNN) architecture has been shown to capture features that are more suitable for histopathology images [12].

### *ImageNet-22k Pre-trained Vision Transformer*

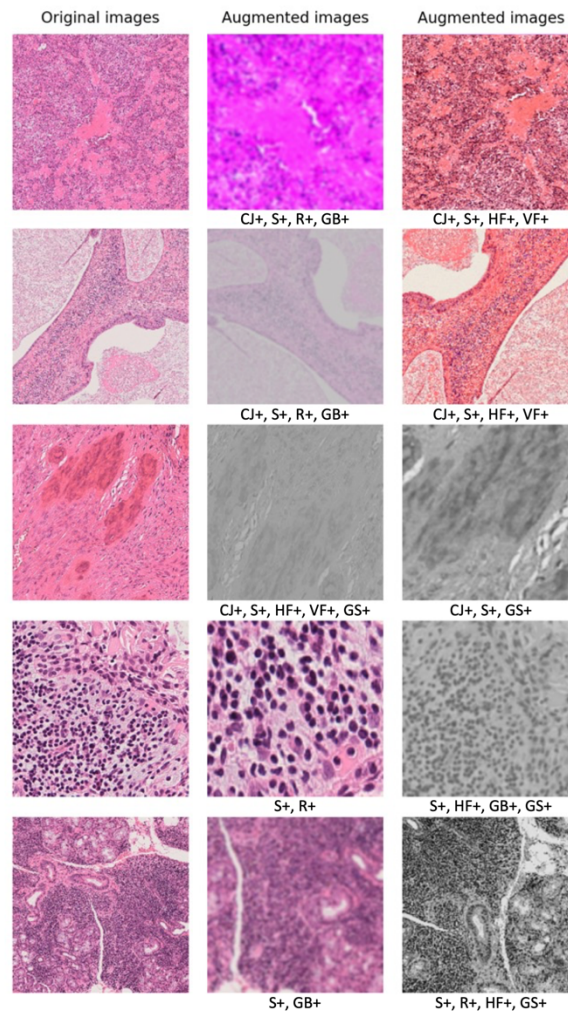
We used the DINOv2 ViT-L/14 model pre-trained with the SSL method on general images from ImageNet-22k [26]. The last layer was used for feature extraction. The images were cropped to 252 pixels owing to the input size restrictions.

### *Fine-tuned CNN*

We fine-tuned all the ImageNet-1k pre-trained ResNet18 models on our dataset using the supervised learning method to classify 51 categories. All the layers were trained with a learning rate of 0.001, a batch size of 32, and 100 epochs in PyTorch 1.11.0. During training, a random 90-degree rotation, random horizontal and vertical flips, color jitter, Gaussian blur, and color normalization transformation were performed. The training-to-test ratio was 8:2.

### *CNN Trained with SSL Methods*

Contrastive (SimCLR) [21] and noncontrastive methods (TiCo) [25] were investigated. ResNet18 is used as the backbone. During training, random color jittering, grayscale, image scaling, horizontal and vertical flipping, 90-degree rotation, Gaussian blurring, and color augmentation were implemented (Figure 6). Both models were trained with Lightly version 1.3.3, with a learning rate of 1.2, a batch size of  $32 \times 32$  (with accumulated gradients), and 1,000 epochs. Examples of original and augmented images are shown in Figure 2B. The ResNet18 backbone model trained on 57 histopathology image datasets (38,594 image patches and 24,923 WSIs) developed by Ciga et al. 2022 was also used for comparison [14].



### Figure 6: Data augmentation examples

Random color jittering (CJ+), grayscale (GS+), image scaling (S+), horizontal and vertical flips (HF+ and VF+), 90-degree rotation (R+), and Gaussian blur (GB+), used during SSL training. Color normalization was used during model training but is not included in this visualization.

The code for SSL model training is available at [https://github.com/rannyrh/oralpath\\_CBIR](https://github.com/rannyrh/oralpath_CBIR).

### *Histopathology Image-trained Vision Transformer*

A vision transformer-based model, Phikon, was trained with 40 million pan-cancer tiles extracted from The Cancer Genomic Atlas (TCGA) using the masked image modeling (MIM) method as an SSL framework. MIM learns meaningful representation by randomly masks portions of an image and trying to reconstruct those masked portions. This model was developed by Owkin. Inc [24]

## Results

### Model performance evaluation

The highest MeanAUC for in-domain queries was achieved by ResNet18+SimCLR (0.900), followed by ResNet18+TiCo (0.897). They achieved this at the query category level for 8 out of the 11 categories. The generalizability of these models was validated using out-of-domain-phonocam queries. The highest performance for out-of-domain-phonocam queries was also achieved by ResNet18+SimCLR (0.871), followed by ResNet18+TiCo (0.857). The highest performance on the query case category levels was achieved by both SSL models for 7 out of the 11 categories. We tested the performance on queries from other institutions to further demonstrate the generalizability. A similar result was yielded by both SSL models, where ResNet18+SimCLR leads with the highest MeanAUC (0.886 for out-of-domain-B; and 0.913 for out-of-domain-D queries), followed by ResNet18+TiCo (0.881 for out-of-domain-B; and 0.905 for out-of-domain-D queries). Phikon, pre-trained on histopathological images, and DINOv2, pre-trained on large-scale general images, performed comparably well with DINOv2 leading the overall MeanAUC in out-of-domain-phonocam query, with Phikon leading in the other three query sets (Table 3).

**Table 3: Mean-AUC (SD) of each test query category in in- and out-of-domain image queries.**

The performances of the SSL models are superior for most test query categories and the overall averages. The highest AUC for each category is marked in **bold**, and the second highest is marked in *italics*.

Query	Category	Pre-trained VGG16	Pre-trained DINOv2	Fine-tuned Resnet18	Resnet18 + SimCLR	Resnet18 + TiCo	Ciga model	Phikon
In-domain (n = 120 images from 2 cases)	Nasopalatine duct cyst	0.758 (0.09)	0.815 (0.07)	0.684 (0.07)	<b>0.857</b> <b>(0.06)</b>	<i>0.855</i> <i>(0.05)</i>	0.811 (0.07)	0.812 (0.06)
	Glandular odontogenic cyst	0.842 (0.09)	0.890 (0.08)	0.770 (0.08)	<b>0.899</b> <b>(0.04)</b>	<b>0.899</b> <b>(0.03)</b>	0.832 (0.07)	0.817 (0.03)
	Odontogenic keratocyst	0.823 (0.07)	0.921 (0.04)	0.922 (0.04)	<i>0.963</i> <i>(0.03)</i>	<b>0.968</b> <b>(0.03)</b>	0.855 (0.07)	0.962 (0.02)
	Orthokeratinized odontogenic cyst	0.838 (0.08)	0.922 (0.10)	0.823 (0.10)	<i>0.976</i> <i>(0.03)</i>	<b>0.978</b> <b>(0.03)</b>	0.861 (0.07)	0.927 (0.04)
	Basal cell adenoma	0.937 (0.05)	0.905 (0.01)	<b>0.973</b> <b>(0.01)</b>	<i>0.961</i> <i>(0.03)</i>	<i>0.961</i> <i>(0.03)</i>	0.924 (0.03)	0.954 (0.03)
	Adenoid cystic carcinoma	0.739 (0.12)	0.831 (0.05)	0.862 (0.05)	<b>0.918</b> <b>(0.05)</b>	<i>0.912</i> <i>(0.07)</i>	0.743 (0.11)	0.807 (0.12)
	Mucoepidermoid carcinoma	0.745 (0.11)	0.776 (0.07)	0.669 (0.07)	<i>0.815</i> <i>(0.09)</i>	0.799 (0.08)	0.728 (0.13)	<b>0.837</b> <b>(0.05)</b>
	Warthin's tumor	0.966 (0.03)	0.939 (0.02)	0.968 (0.02)	<b>0.997</b> <b>(0.01)</b>	<i>0.996</i> <i>(0.08)</i>	0.945 (0.05)	0.990 (0.03)
	Odontogenic fibroma	0.730 (0.14)	<b>0.805</b> <b>(0.10)</b>	0.642 (0.10)	<i>0.769</i> <i>(0.12)</i>	0.763 (0.13)	0.695 (0.09)	0.755 (0.07)
	Ameloblastoma	0.730 (0.11)	0.730 (0.04)	0.664 (0.04)	<b>0.809</b> <b>(0.10)</b>	<i>0.808</i> <i>(0.09)</i>	0.701 (0.09)	0.744 (0.10)
	Hemangioma	0.886 (0.07)	0.842 (0.06)	0.881 (0.06)	<b>0.935</b> <b>(0.03)</b>	<i>0.929</i> <i>(0.03)</i>	0.842 (0.09)	0.886 (0.06)
<b>Average</b>		0.818 (0.08)	0.852 (0.06)	0.805 (0.12)	<b>0.900</b> <b>(0.07)</b>	<i>0.897</i> <i>(0.08)</i>	0.812 (0.08)	0.863 (0.08)

Query	Category	Pre-trained VGG16	Pre-trained DINOv2	Fine-tuned Resnet18	Resnet18 + SimCLR	Resnet18 + TiCo	Ciga model	Phikon
Out-of-domain phonecam (n = 54 images from 2 cases)	Nasopalatine duct cyst	0.466 (0.08)	0.826 (0.05)	0.721 (0.07)	<b>0.871</b> <b>(0.04)</b>	0.859 (0.02)	0.807 (0.06)	0.764 (0.05)
	Glandular odontogenic cyst	0.671 (0.10)	0.876 (0.04)	0.763 (0.06)	<b>0.915</b> <b>(0.02)</b>	0.908 (0.02)	0.876 (0.03)	0.795 (0.04)
	Odontogenic keratocyst	0.622 (0.13)	0.891 (0.07)	0.614 (0.13)	0.961 (0.02)	<b>0.968</b> <b>(0.02)</b>	0.892 (0.04)	0.907 (0.04)
	Orthokeratinized odontogenic cyst	0.647 (0.13)	0.891 (0.05)	0.751 (0.10)	0.920 (0.08)	0.917 (0.08)	0.875 (0.06)	<b>0.939</b> <b>(0.03)</b>
	Basal cell adenoma	<b>0.896</b> <b>(0.07)</b>	0.815 (0.07)	0.576 (0.14)	0.884 (0.08)	0.846 (0.09)	0.769 (0.08)	0.796 (0.08)
	Adenoid cystic carcinoma	0.779 (0.09)	0.793 (0.09)	0.634 (0.10)	<b>0.850</b> <b>(0.13)</b>	0.799 (0.17)	0.651 (0.16)	0.652 (0.11)
	Mucoepidermoid carcinoma	0.536 (0.06)	0.700 (0.10)	0.562 (0.05)	<b>0.828</b> <b>(0.09)</b>	0.793 (0.07)	0.724 (0.07)	0.707 (0.10)
	Warthin's tumor	0.739 (0.09)	0.917 (0.06)	0.819 (0.08)	<b>0.977</b> <b>(0.04)</b>	0.969 (0.06)	0.825 (0.14)	0.894 (0.08)
	Odontogenic fibroma	0.570 (0.12)	<b>0.725</b> <b>(0.11)</b>	0.290 (0.05)	0.663 (0.13)	0.671 (0.12)	0.648 (0.10)	0.649 (0.10)
	Ameloblastoma	<b>0.862</b> <b>(0.04)</b>	0.761 (0.10)	0.645 (0.04)	0.814 (0.12)	0.793 (0.11)	0.670 (0.10)	0.711 (0.06)
	Hemangioma	0.737 (0.17)	0.812 (0.05)	0.585 (0.10)	0.899 (0.06)	<b>0.904</b> <b>(0.05)</b>	0.795 (0.11)	0.889 (0.05)
<b>Average</b>	0.684 (0.13)	0.819 (0.07)	0.633 (0.14)	<b>0.871</b> <b>(0.08)</b>	0.857 (0.9)	0.776 (0.09)	0.791 (0.10)	
Query	Category	Pre-trained VGG16	Pre-trained DINOv2	Fine-tuned Resnet18	Resnet18 + SimCLR	Resnet18 + TiCo	Ciga model	Phikon
Out-of-domain-B	Myoepithelioma (n = 60 images from 1 case)	0.744 (0.08)	<b>0.846</b> <b>(0.05)</b>	0.714 (0.05)	0.814 (0.05)	0.821 (0.04)	0.812 (0.05)	0.835 (0.05)
	Basal cell adenoma (n = 60 images from 1 case)	0.867 (0.09)	0.799 (0.07)	0.840 (0.10)	<b>0.909</b> <b>(0.04)</b>	0.885 (0.08)	0.789 (0.10)	0.882 (0.06)
	Warthin's tumor (n = 60 images from 1 case)	0.943 (0.04)	0.981 (0.02)	0.838 (0.07)	<b>0.999</b> <b>(0.01)</b>	0.997 (0.02)	0.861 (0.05)	0.981 (0.01)
	Carcinoma ex pleomorphic adenoma (n = 60 images from 1 case)	0.766 (0.06)	0.808 (0.10)	0.684 (0.07)	<b>0.913</b> <b>(0.06)</b>	0.903 (0.05)	0.774 (0.11)	0.865 (0.06)
	Mucoepidermoid carcinoma (n = 120 images from 2 cases)	0.797 (0.05)	<b>0.822</b> <b>(0.06)</b>	0.619 (0.09)	0.814 (0.04)	0.802 (0.04)	0.745 (0.07)	0.770 (0.06)
	Adenoid cystic carcinoma (n = 120 images from 2 cases)	0.795 (0.08)	0.825 (0.05)	0.835 (0.05)	0.880 (0.08)	<b>0.896</b> <b>(0.09)</b>	0.741 (0.07)	0.815 (0.06)



	Acinic cell carcinoma (n = 120 images from 2 cases)	0.779 (0.13)	0.803 (0.07)	0.759 (0.07)	0.802 (0.17)	0.789 (0.21)	0.660 (0.23)	<b>0.816</b> <b>(0.12)</b>
	Salivary duct carcinoma (n = 60 images from 1 case)	0.904 (0.02)	0.889 (0.05)	0.798 (0.03)	<b>0.961</b> <b>(0.01)</b>	0.953 (0.02)	0.864 (0.07)	0.952 (0.02)
	<b>Average</b>	0.825 (0.07)	0.847 (0.06)	0.761 (0.08)	<b>0.886</b> <b>(0.07)</b>	0.881 (0.07)	0.781 (0.06)	0.864 (0.07)
<b>Query</b>	<b>Category</b>	<b>Pre-trained VGG16</b>	<b>Pre-trained DINOv2</b>	<b>Fine-tuned Resnet18</b>	<b>Resnet18 + SimCLR</b>	<b>Resnet18 + TiCo</b>	<b>Ciga model</b>	<b>Phikon</b>
	Adenoid cystic carcinoma (n = 180 images from 3 cases)	0.846 (0.07)	0.807 (0.06)	0.812 (0.04)	<b>0.887</b> <b>(0.05)</b>	0.871 (0.08)	0.774 (0.09)	0.842 (0.06)
	Basal cell adenoma (n = 180 images from 3 cases)	<b>0.936</b> <b>(0.04)</b>	0.868 (0.08)	0.811 (0.06)	0.919 (0.05)	0.935 (0.04)	0.855 (0.06)	0.925 (0.03)
	Odontogenic myxoma/myxofibroma (n = 180 images from 3 cases)	0.899 (0.10)	0.935 (0.08)	0.784 (0.09)	<b>0.970</b> <b>(0.06)</b>	0.962 (0.07)	0.955 (0.06)	0.914 (0.08)
	Fibrous dysplasia (n = 120 images from 2 cases)	0.811 (0.09)	0.850 (0.07)	0.759 (0.04)	<b>0.907</b> <b>(0.04)</b>	0.848 (0.06)	0.813 (0.10)	0.824 (0.06)
Out-of-domain-C	Osteoma (n = 180 images from 3 cases)	0.918 (0.09)	0.913 (0.09)	0.824 (0.04)	0.963 (0.05)	<b>0.975</b> <b>(0.03)</b>	0.929 (0.07)	0.944 (0.08)
	Odontogenic keratocyst (n = 120 images from 2 cases)	0.806 (0.10)	<b>0.893</b> <b>(0.06)</b>	0.540 (0.18)	0.871 (0.11)	0.875 (0.10)	0.837 (0.11)	0.878 (0.09)
	Orthokeratinized odontogenic cyst (n = 120 images from 2 cases)	0.758 (0.13)	0.907 (0.04)	0.634 (0.09)	<b>0.949</b> <b>(0.04)</b>	0.932 (0.07)	0.836 (0.10)	0.895 (0.09)
	Adenomatoid odontogenic tumor (n = 120 images from 2 cases)	0.819 (0.07)	0.820 (0.08)	0.756 (0.04)	0.836 (0.05)	<b>0.839</b> <b>(0.05)</b>	0.723 (0.13)	0.799 (0.09)
	<b>Average</b>	0.849 (0.06)	0.874 (0.09)	0.740 (0.09)	<b>0.913</b> <b>(0.04)</b>	0.905 (0.05)	0.840 (0.07)	0.878 (0.05)

Overall, the MeanAcc of test query set-A was highest with the SSL models: ResNet18+TiCo outperformed other models for in-domain queries (4.64), followed by ResNet18+SimCLR (4.53) with no significant difference (Wilcoxon signed-rank test with Bonferroni adjustment). The reverse was observed for the out-domain-phonecam queries (ResNet18+SimCLR (3.33) and ResNet18+TiCo (3.31)) with no significant difference (Wilcoxon signed-rank test with Bonferroni adjustment). Phikon yielded the highest MeanAcc for out-of-domain-B queries (3.79), followed by ResNet18+SimCLR (3.68). Pre-trained DINOv2 outperformed other models for out-of-domain-C queries (3.64), followed closely by ResNet18+TiCo (3.63) (Table 4). The highest overall MeanAcc was consistently achieved by the SSL models at different magnification

levels, except for the high-magnification in-domain queries (Figures 7A, 7B). The highest %query was obtained with SSL models for most query categories (Table 5).

**Table 4. MeanAcc (SD) of each test query category in in- and out-of-domain image queries**

The performances of the SSL models are superior for most categories, further validating the robustness of the models under a wide range of histopathological image conditions. The highest MeanAcc for each category is marked in **bold** and the second highest is in *italics*.

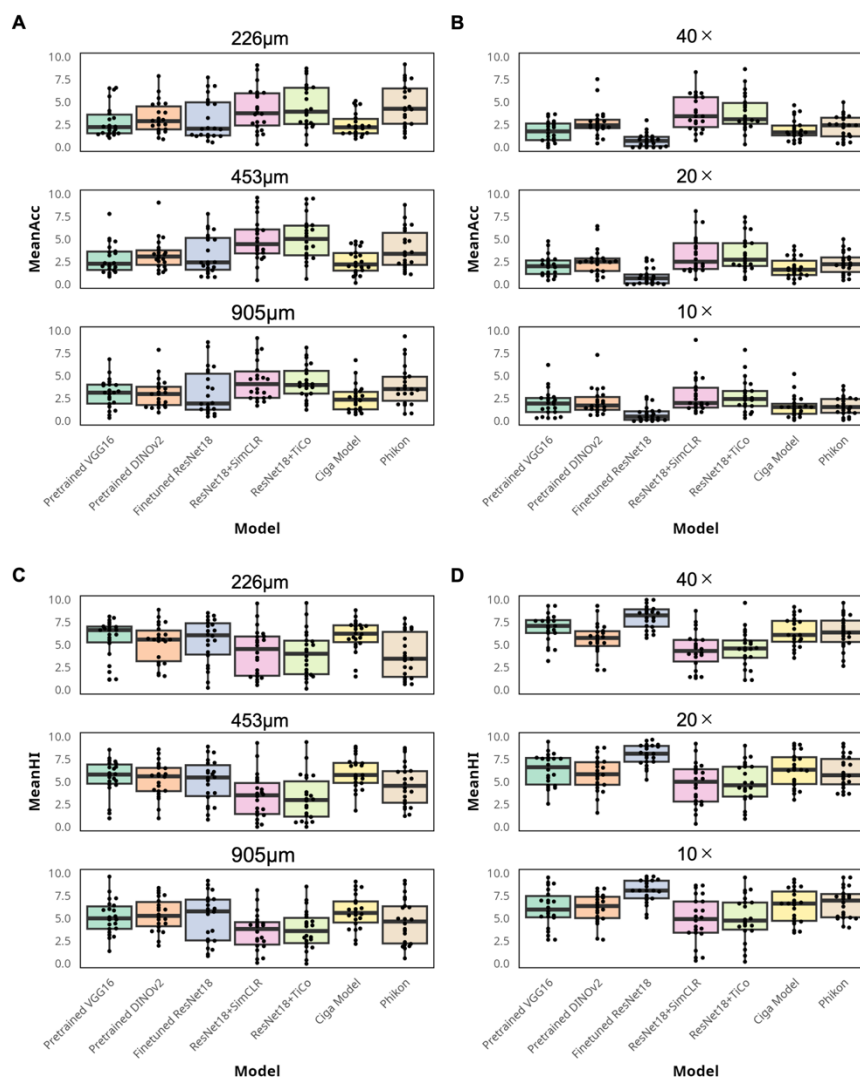
Query	Category	Pretrained VGG16	Pretrained DINOv2	Finetuned Resnet18	ResNet18 + SimCLR	ResNet18 + TiCo	Ciga model	Phikon
In-domain (n = 120 images from 2 cases)	Nasopalatine duct cyst	1.38 (0.28)	1.48 (0.44)	1.53 (1.02)	<i>2.59</i> (0.94)	<b>3.48</b> (0.82)	1.93 (0.43)	2.38 (0.66)
	Glandular odontogenic cyst	2.55 (1.00)	2.70 (0.65)	1.72 (0.52)	<i>3.13</i> (0.77)	<b>3.22</b> (0.71)	2.07 (0.73)	2.58 (0.65)
	Odontogenic keratocyst	2.41 (0.53)	4.96 (0.33)	5.39 (0.50)	<i>6.79</i> (0.92)	<i>6.73</i> (0.76)	2.79 (0.82)	<b>7.13</b> (0.78)
	Orthokeratinized odontogenic cyst	1.68 (0.96)	3.94 (1.49)	3.04 (1.83)	<i>6.19</i> (1.84)	<b>6.83</b> (1.93)	2.38 (1.01)	4.92 (1.37)
	Basal cell adenoma	5.25 (0.85)	3.43 (1.51)	<b>6.11</b> (0.44)	5.13 (1.34)	5.61 (1.72)	4.34 (0.58)	5.73 (1.14)
	Adenoid cystic carcinoma	2.86 (1.10)	3.28 (0.66)	3.15 (1.37)	<i>4.09</i> (0.99)	<b>4.36</b> (0.98)	2.14 (1.13)	3.13 (1.02)
	Mucoepidermoid carcinoma	<i>3.10</i> (0.69)	2.78 (0.89)	1.33 (0.26)	2.87 (0.83)	2.89 (1.10)	1.85 (0.94)	<b>3.71</b> (0.75)
	Warthin's tumor	6.13 (1.32)	5.93 (2.65)	7.63 (0.88)	<b>8.94</b> (0.56)	<i>8.38</i> (1.05)	4.97 (1.16)	8.23 (1.10)
	Odontogenic fibroma	<i>2.05</i> (1.36)	<b>2.45</b> (0.89)	0.99 (0.32)	2.03 (1.33)	2.04 (1.40)	1.29 (0.75)	1.74 (1.01)
	Ameloblastoma	1.67 (0.79)	1.83 (0.80)	1.41 (0.65)	<b>2.73</b> (1.19)	<i>2.42</i> (0.88)	1.09 (0.32)	1.71 (0.69)
	Hemangioma	3.20 (1.04)	2.51 (1.07)	3.03 (1.23)	<b>5.39</b> (1.06)	<i>5.11</i> (1.15)	2.93 (1.07)	4.56 (1.60)
<b>Average</b>		3.20 (1.50)	3.21 (1.32)	3.21 (2.22)	<i>4.53</i> (2.16)	<b>4.64</b> (2.06)	2.53 (1.19)	4.16 (2.29)
<b>Query</b>	<b>Category</b>	<b>Pre-trained VGG16</b>	<b>Pre-trained DINOv2</b>	<b>Fine-tuned Resnet18</b>	<b>Resnet18 + SimCLR</b>	<b>Resnet18 + TiCo</b>	<b>Ciga model</b>	<b>Phikon</b>
Out-of-domain- phonecam	Nasopalatine duct cyst	0.98 (0.59)	1.85 (0.44)	0.87 (0.39)	<i>2.15</i> (0.40)	<b>2.52</b> (0.21)	1.06 (0.40)	1.57 (0.59)
	Glandular odontogenic cyst	1.91 (0.86)	2.48 (0.77)	2.00 (0.54)	<b>3.28</b> (0.99)	<i>2.98</i> (1.10)	2.00 (0.33)	2.91 (0.64)

(n = 54 images from 2 cases)	Odontogenic keratocyst	2.83 (0.63)	4.24 (1.34)	0.31 (0.56)	5.19 (0.76)	<b>5.52</b> <b>(0.81)</b>	3.41 (1.16)	3.93 (0.98)
	Orthokeratinized odontogenic cyst	1.87 (1.04)	2.67 (0.40)	1.67 (1.25)	<b>3.78</b> <b>(1.68)</b>	3.59 (1.86)	2.20 (0.82)	2.72 (0.42)
	Basal cell adenoma	2.00 (0.39)	2.15 (0.43)	0.19 (0.25)	<b>2.48</b> <b>(0.97)</b>	2.43 (1.00)	1.19 (0.62)	2.24 (1.10)
	Adenoid cystic carcinoma	2.00 (1.20)	2.67 (0.96)	0.74 (0.62)	<b>2.91</b> <b>(0.98)</b>	2.87 (1.42)	1.69 (1.33)	1.50 (1.27)
	Mucoepidermoid carcinoma	<b>2.43</b> <b>(0.67)</b>	1.93 (0.76)	0.31 (0.33)	1.93 (0.52)	1.78 (1.00)	0.87 (0.70)	0.89 (0.70)
	Warthin's tumor	3.17 (1.80)	5.06 (2.33)	1.26 (0.89)	<b>7.28</b> <b>(2.31)</b>	7.26 (1.46)	2.83 (1.69)	2.31 (1.45)
	Odontogenic fibroma	0.56 (0.22)	0.91 (0.22)	0.04 (0.09)	1.09 (0.59)	1.33 (0.79)	<b>1.44</b> <b>(0.37)</b>	0.70 (0.30)
	Ameloblastoma	1.15 (0.85)	2.02 (0.37)	1.22 (0.34)	<b>3.26</b> <b>(1.25)</b>	2.76 (0.88)	1.28 (0.44)	2.02 (0.59)
	Hemangioma	2.39 (0.30)	1.37 (0.81)	0.26 (0.24)	3.35 (0.65)	<b>3.43</b> <b>(0.62)</b>	1.98 (0.84)	2.00 (1.08)
	<b>Average</b>	1.93 (1.11)	2.48 (1.47)	0.81 (0.83)	<b>3.33</b> <b>(1.93)</b>	3.31 (1.92)	1.81 (1.11)	2.07 (1.21)

Query	Category	Pre-trained VGG16	Pre-trained DINOv2	Fine-tuned Resnet18	ResNet18 + SimCLR	ResNet18 + TiCo	Ciga model	Phikon
Out-of-domain-B	Myoepithelioma (n = 60 images from 1 case)	0.88 (0.29)	<b>1.97</b> <b>(0.34)</b>	1.27 (0.89)	1.00 (0.26)	1.17 (0.26)	1.38 (0.48)	1.73 (0.33)
	Basal cell adenoma (n = 60 images from 1 case)	2.62 (0.35)	1.37 (0.40)	<b>2.92</b> <b>(0.75)</b>	2.42 (0.18)	1.98 (0.20)	1.62 (0.40)	<b>3.03</b> <b>(0.56)</b>
	Warthin's tumor (n = 60 images from 1 case)	5.70 (0.96)	7.63 (0.71)	2.30 (2.46)	<b>8.88</b> <b>(0.26)</b>	8.23 (0.33)	2.52 (0.13)	7.28 (0.87)
	Carcinoma ex pleomorphic adenoma (n = 60 images from 1 case)	1.45 (0.00)	1.48 (0.78)	1.15 (0.64)	<b>3.33</b> <b>(0.53)</b>	3.07 (0.18)	1.52 (0.21)	1.72 (0.29)
	Mucoepidermoid carcinoma (n = 120 images from 2 cases)	2.09 (1.02)	<b>2.14</b> <b>(0.53)</b>	0.62 (0.51)	1.61 (0.84)	1.73 (0.87)	1.90 (0.90)	1.61 (1.12)
	Adenoid cystic carcinoma (n = 120 images from 2 cases)	3.35 (1.32)	<b>4.15</b> <b>(1.24)</b>	1.82 (0.91)	3.31 (2.39)	3.23 (2.02)	2.55 (1.79)	3.88 (1.76)

	Acinic cell carcinoma (n = 120 images from 2 cases)	3.99 (2.37)	5.00 (1.76)	2.43 (0.91)	4.95 (1.90)	5.63 (2.38)	3.53 (2.51)	<b>5.78</b> <b>(1.41)</b>
	Salivary duct carcinoma (n = 60 images from 1 case)	3.95 (0.36)	4.12 (1.68)	0.90 (0.74)	5.05 (0.26)	4.68 (1.07)	4.23 (3.38)	<b>5.40</b> <b>(2.04)</b>
	<b>Average</b>	3.04 (1.78)	3.56 (2.13)	1.66 (1.19)	3.68 (2.50)	3.67 (2.47)	2.47 (1.78)	<b>3.79</b> <b>(2.33)</b>
<b>Query</b>	<b>Category</b>	<b>Pre-trained VGG16</b>	<b>Pre-trained DINOv2</b>	<b>Fine-tuned Resnet18</b>	<b>ResNet18 + SimCLR</b>	<b>ResNet18 + TiCo</b>	<b>Ciga model</b>	<b>Phikon</b>
Out-of-domain-C	Adenoid cystic carcinoma (n = 180 images from 3 cases)	3.39 (1.06)	3.04 (1.52)	1.44 (0.58)	2.74 (1.34)	3.04 (0.95)	1.85 (0.64)	<b>3.62</b> <b>(0.56)</b>
	Basal cell adenoma (n = 180 images from 3 cases)	<b>3.94</b> <b>(1.45)</b>	2.35 (0.85)	0.95 (0.55)	2.23 (1.90)	2.75 (2.11)	1.19 (0.56)	3.01 (1.63)
	Odontogenic myxoma/myxofibroma (n = 180 images from 3 cases)	4.37 (2.25)	6.00 (1.66)	0.96 (0.43)	5.62 (1.48)	<b>6.05</b> <b>(1.64)</b>	5.20 (1.60)	4.42 (1.70)
	Fibrous dysplasia (n = 120 images from 2 cases)	2.37 (1.22)	<b>2.90</b> <b>(0.59)</b>	0.48 (0.19)	2.07 (1.03)	1.48 (0.79)	1.52 (0.99)	1.45 (1.10)
	Osteoma (n = 180 images from 3 cases)	4.22 (1.52)	4.63 (1.39)	1.97 (0.19)	4.27 (2.18)	4.66 (0.86)	3.93 (0.74)	<b>4.93</b> <b>(1.65)</b>
	Odontogenic keratocyst (n = 120 images from 2 cases)	2.65 (1.62)	<b>3.99</b> <b>(1.26)</b>	0.28 (0.56)	3.09 (3.27)	3.13 (3.16)	3.56 (1.87)	3.38 (2.05)
	Orthokeratinized odontogenic cyst (n = 120 images from 2 cases)	1.53 (0.98)	3.60 (1.39)	0.02 (0.04)	3.95 (2.32)	<b>4.06</b> <b>(2.07)</b>	2.15 (1.30)	3.59 (2.31)
	Adenomatoid odontogenic tumor (n = 120 images from 2 cases)	1.79 (0.42)	1.90 (0.62)	2.19 (0.89)	<b>3.07</b> <b>(0.61)</b>	2.85 (0.77)	0.90 (0.64)	2.37 (1.35)
	<b>Average</b>	3.22 (1.71)	<b>3.64</b> <b>(1.76)</b>	1.52 (1.36)	3.45 (2.13)	3.63 (2.08)	2.64 (1.81)	3.47 (1.81)



**Figure 7: MeanAcc and MeanHI comparisons**

(A) MeanAcc comparison for the in-domain query of all models by magnification showing the highest performance of Phikon at the highest magnification and that of ResNet18+SimCLR and ResNet18+TiCo at the moderate and lowest magnification. (B) MeanAcc comparison for out-of-domain-phonocam queries of all models by magnification shows the highest performance of ResNet18+SimCLR and ResNet18+TiCo at the highest and lowest magnification. Both model performances were comparable to that of pre-trained DINOv2 at moderate magnification. (C) MeanHI comparison for in-domain queries by magnification shows that ResNet18+SimCLR and ResNet18+TiCo outperformed other models except for the highest magnification where Phikon leads with a wider interquartile range. (D) MeanHI comparison for out-of-domain-phonocam query by magnification showing ResNet18+SimCLR and ResNet18+TiCo outperformed other models. (C-D) Please note that a lower MeanHI value denotes a higher model performance.

**Table 5. Percentage of queries that retrieved at least one correct diagnosis among the top-10 results for each test query category (%query) in in- and out-of-domain image queries**

The performance of the SSL models was superior in most categories. The highest %query for each category is marked in **bold**, and the second highest is in *italics*. (%)

Query	Category	Pre-trained VGG16	Pre-trained DINOv2	Fine-tuned Resnet18	ResNet18 + SimCLR	ResNet18 + TiCo	Ciga model	Phikon
In-domain (n = 120 images from 2 cases)	Nasopalatine duct cyst	80.00	80.00	71.67	93.33	<b>96.67</b>	87.50	95.83
	Glandular odontogenic cyst	92.50	95.00	93.33	<b>99.17</b>	<b>99.17</b>	91.67	95.00
	Odontogenic keratocyst	88.33	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	96.67	88.33	<b>100.00</b>
	Orthokeratinized odontogenic cyst	84.17	96.67	91.67	98.33	<b>100.00</b>	92.50	<b>100.00</b>
	Basal cell adenoma	<b>100.00</b>	96.67	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>
	Adenoid cystic carcinoma	94.17	96.67	98.33	<b>100.00</b>	<b>100.00</b>	81.67	<b>100.00</b>
	Mucoepidermoid carcinoma	94.17	88.33	94.17	<b>98.33</b>	97.50	80.00	<b>98.33</b>
	Warthin's tumor	99.17	97.50	<b>100.00</b>	<b>100.00</b>	99.17	95.83	98.33
	Odontogenic fibroma	85.83	<b>94.17</b>	70.83	73.33	71.67	68.33	85.00
	Ameloblastoma	84.17	81.67	80.00	92.50	<b>93.33</b>	70.00	89.17
	Hemangioma	91.67	92.50	98.33	<b>100.00</b>	<b>100.00</b>	93.33	<b>100.00</b>
<b>Average</b>		90.38	92.65	90.76	95.91	95.83	86.29	<b>96.52</b>
Query	Category	Pre-trained VGG16	Pre-trained DINOv2	Fine-tuned Resnet18	ResNet18 + SimCLR	ResNet18 + TiCo	Ciga model	Phikon
Out-domain-phonecam (n = 54 images from 2 cases)	Nasopalatine duct cyst	53.70	92.59	68.52	98.15	<b>100.00</b>	74.07	87.04
	Glandular odontogenic cyst	88.89	98.15	81.48	<b>100.00</b>	<b>100.00</b>	96.30	<b>100.00</b>
	Odontogenic keratocyst	90.74	88.89	16.67	98.15	<b>100.00</b>	96.30	92.59

Orthokeratinized odontogenic cyst	75.93	92.59	62.96	88.89	88.89	88.89	<b>96.30</b>
Basal cell adenoma	<b>96.30</b>	90.74	12.96	85.19	88.89	75.93	83.33
Adenoid cystic carcinoma	85.19	85.19	42.59	<b>88.89</b>	74.07	64.81	64.81
Mucoepidermoid carcinoma	<b>94.44</b>	81.48	25.93	<b>94.44</b>	70.37	61.11	53.70
Warthin's tumor	72.22	<b>98.15</b>	51.85	96.30	<b>98.15</b>	72.22	75.93
Odontogenic fibroma	51.85	55.56	3.70	79.63	79.63	<b>90.74</b>	66.67
Ameloblastoma	61.11	83.33	88.89	94.44	<b>98.15</b>	79.63	88.89
Hemangioma	<b>94.44</b>	66.67	24.07	81.48	90.74	81.48	85.19
<b>Average</b>	78.62	84.85	43.60	<b>91.41</b>	89.90	80.13	81.31

Query	Category	Pre-trained VGG16	Pre-trained DINOv2	Fine-tuned Resnet18	ResNet18 + SimCLR	ResNet18 + TiCo	Ciga model	Phikon
Out-of-domain-B	Myoepithelioma (n = 60 images from 1 case)	43.33	<b>93.33</b>	63.33	55.00	56.67	75.00	78.33
	Basal cell adenoma (n = 60 images from 1 case)	<b>100.00</b>	78.33	96.67	<b>100.00</b>	95.00	93.33	<b>100.00</b>
	Warthin's tumor (n = 60 images from 1 case)	<b>100.00</b>	<b>100.00</b>	71.67	<b>100.00</b>	<b>100.00</b>	90.00	<b>100.00</b>
	Carcinoma ex pleomorphic adenoma (n = 60 images from 1 case)	75.00	66.67	68.33	96.67	<b>98.33</b>	80.00	71.67
	Mucoepidermoid carcinoma (n = 120 images from 2 cases)	81.24	<b>90.17</b>	52.71	78.63	89.09	82.98	80.08
	Adenoid cystic carcinoma (n = 120 images from 2 cases)	97.50	<b>100.00</b>	90.00	80.00	85.83	86.67	97.50
	Acinic cell carcinoma (n = 120 images from 2 cases)	81.67	<b>97.50</b>	89.17	90.83	85.83	69.17	<b>97.50</b>

	Salivary duct carcinoma (n = 60 images from 1 case)	<b>100.00</b>	<b>100.00</b>	58.33	<b>100.00</b>	<b>100.00</b>	76.67	<b>100.00</b>
	<b>Average</b>	84.84	<i>90.75</i>	73.78	87.64	88.84	81.73	<b>90.92</b>
<b>Query</b>	<b>Category</b>	<b>Pre-trained VGG16</b>	<b>Pre-trained DINOv2</b>	<b>Fine-tuned Resnet18</b>	<b>ResNet18 + SimCLR</b>	<b>ResNet18 + TiCo</b>	<b>Ciga model</b>	<b>Phikon</b>
Out-of-domain-C	Adenoid cystic carcinoma (n = 180 images from 3 cases)	<i>94.44</i>	92.22	80.00	86.11	91.11	81.11	<b>95.56</b>
	Basal cell adenoma (n = 180 images from 3 cases)	<b>97.18</b>	<i>89.94</i>	72.90	69.58	73.27	75.99	88.44
	Odontogenic myxoma/myxofibroma (n = 180 images from 3 cases)	95.00	<b>100.00</b>	76.67	92.78	<i>98.33</i>	<i>98.33</i>	93.33
	Fibrous dysplasia (n = 120 images from 2 cases)	85.83	<b>100.00</b>	<i>99.17</i>	77.50	70.00	70.00	72.50
	Osteoma (n = 180 images from 3 cases)	97.78	98.89	<b>100.00</b>	82.22	<b>100.00</b>	99.44	99.44
	Odontogenic keratocyst (n = 120 images from 2 cases)	80.00	<b>97.50</b>	10.83	56.67	55.83	80.83	<i>91.67</i>
	Orthokeratinized odontogenic cyst (n = 120 images from 2 cases)	69.17	<b>95.00</b>	0.83	75.83	78.33	68.33	<i>86.67</i>
	Adenomatoid odontogenic tumor (n = 120 images from 2 cases)	89.17	93.33	95.83	<b>98.33</b>	94.17	59.17	80.83
	<b>Average</b>	88.57	<b>95.86</b>	67.03	79.88	82.63	79.15	<i>89.68</i>

The accuracy calculation excluded the histologic similarity between the query and retrieved images, which provides additional information about the histologic features during diagnosis. To verify whether the SSL models retrieved histologically similar images despite the low MeanAcc, MeanHI was introduced. MeanHI excludes accurate diagnosis and differential diagnosis categories, which are similar to the query and include other inaccurate categories. The lowest



overall inaccuracy was consistently achieved by the SSL models, except for the high-magnification in-domain queries, indicating that these models are best at retrieving the most histologically similar images beyond accurate diagnosis (Figures 7C, 7D, Table 6). The top-10 retrieved images of representative cases found by all the tested models are shown in Figures 8 and 9.

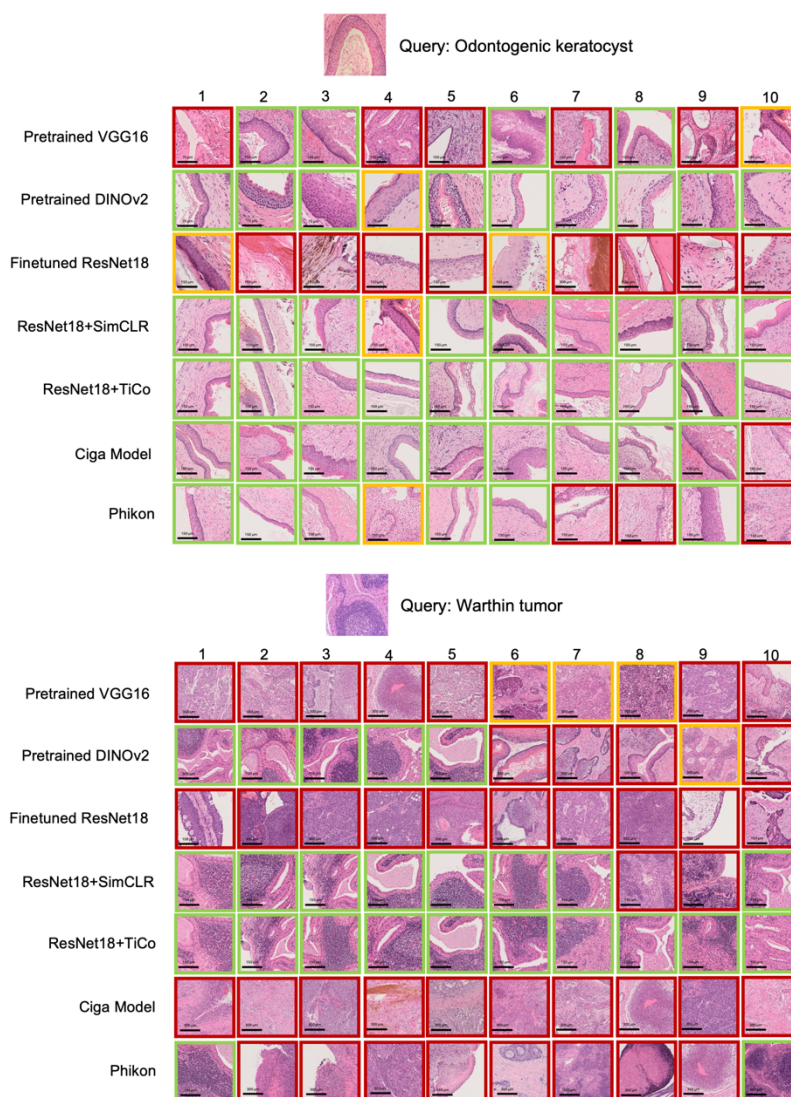
**Table 6. MeanHI (SD) for each test query category for in- and out-of-domain image queries**

The best histologic similarity was consistently retrieved by the SSL models. The histologic similarity of the results was consistently best retrieved by the SSL models. It is noteworthy that a lower MeanHI is preferable. The lowest MeanHI for each category is marked in **bold** and the second lowest is in *italics*.

Query	Category (n = 2 cases)	Pre-trained VGG16	Pre-trained DINOv2	Fine-tuned Resnet18	ResNet18 + SimCLR	ResNet18 + TiCo	Ciga model	Phikon
In-domain (n = 120 images from 2 cases)	Nasopalatine duct cyst	6.93 (0.37)	6.11 (0.91)	6.03 (0.46)	<i>4.91 (1.33)</i>	<b>3.90 (1.02)</b>	6.03 (0.46)	6.13 (1.00)
	Glandular odontogenic cyst	6.00 (0.80)	6.06 (0.77)	5.81 (0.95)	<b>4.94 (0.71)</b>	<i>4.79 (0.81)</i>	5.81 (0.95)	5.11 (1.15)
	Odontogenic keratocyst	6.03 (0.98)	3.10 (0.44)	5.24 (0.53)	<b>1.63 (0.77)</b>	<i>1.73 (0.95)</i>	5.24 (0.53)	1.91 (0.54)
	Orthokeratinized odontogenic cyst	5.88 (1.38)	2.95 (0.96)	5.53 (1.40)	<i>1.33 (1.55)</i>	<b>1.06 (1.18)</b>	5.53 (1.40)	2.78 (0.82)
	Basal cell adenoma	1.53 (0.74)	3.98 (1.51)	2.51 (0.96)	1.66 (0.94)	<b>1.24 (0.71)</b>	2.51 (0.96)	<i>1.51 (0.64)</i>
	Adenoid cystic carcinoma	5.08 (1.18)	4.79 (0.72)	6.30 (1.91)	<i>2.60 (0.77)</i>	<b>2.17 (0.79)</b>	6.30 (1.91)	4.06 (2.27)
	Mucoepidermoid carcinoma	<i>4.84 (0.88)</i>	5.90 (1.00)	7.24 (0.98)	5.35 (1.58)	5.28 (1.84)	7.24 (0.98)	<b>4.73 (1.16)</b>
	Warthin's tumor	3.01 (1.11)	3.66 (2.38)	4.38 (1.12)	<b>1.01 (0.58)</b>	<i>1.58 (1.00)</i>	4.38 (1.12)	1.62 (1.04)
	Odontogenic fibroma	<i>6.67 (1.59)</i>	<b>6.66 (1.25)</b>	7.54 (1.07)	7.03 (2.07)	7.05 (2.10)	7.54 (1.07)	7.21 (1.59)
	Ameloblastoma	7.92 (1.05)	7.83 (0.96)	8.16 (0.75)	<b>5.98 (1.97)</b>	<i>6.63 (1.50)</i>	8.16 (0.75)	7.93 (0.88)
	Hemangioma	6.32 (1.11)	7.13 (1.22)	6.40 (1.05)	<b>4.53 (1.12)</b>	<i>4.78 (1.24)</i>	6.40 (1.05)	5.33 (1.68)
	<b>Average</b>	5.47 (1.82)	5.29 (1.68)	5.92 (1.56)	3.72 (2.13)	<b>3.65 (2.20)</b>	5.92 (1.56)	4.39 (2.44)

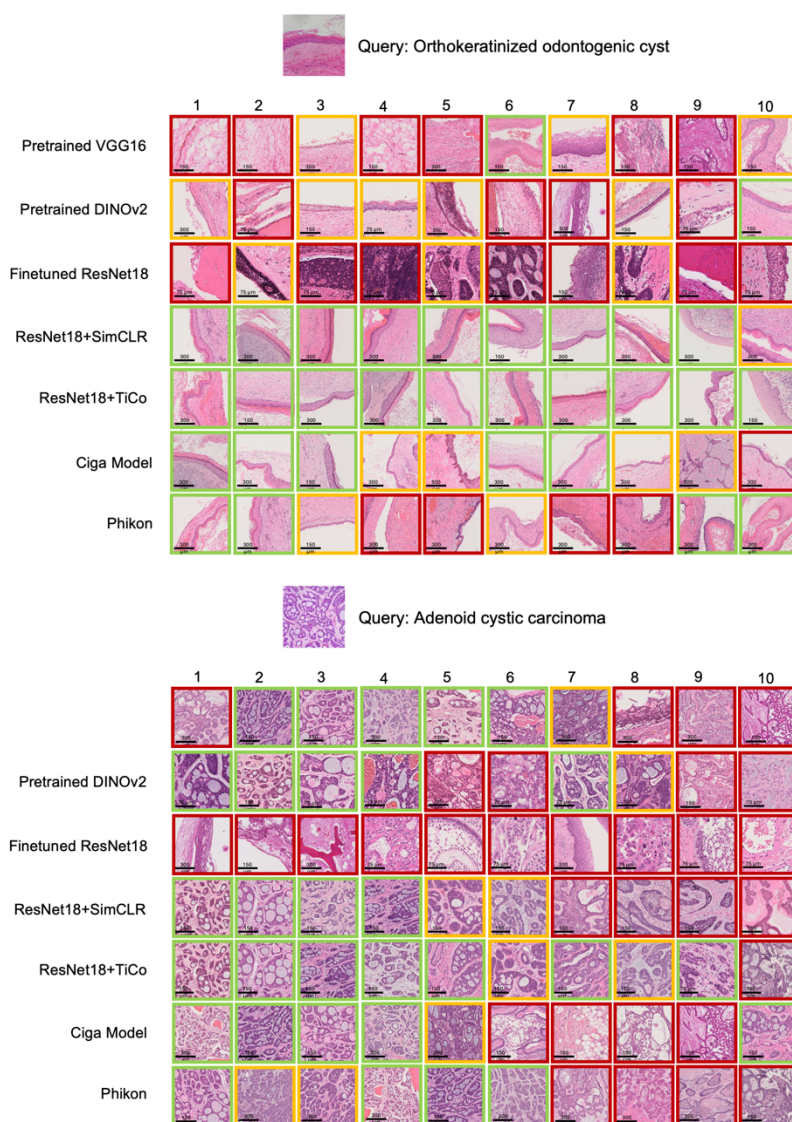
Out-of-domain- phonecam (n = 54 images from 2 cases)	Nasopalatine duct cyst	7.83 (0.52)	5.96 (0.98)	7.30 (0.68)	4.96 (0.60)	<b>4.24</b> <b>(0.32)</b>	6.78 (0.98)	6.67 (1.00)
	Glandular odontogenic cyst	6.48 (1.06)	5.69 (0.66)	6.52 (1.36)	<b>4.91</b> <b>(1.19)</b>	5.04 (0.82)	5.30 (0.63)	5.44 (0.54)
	Odontogenic keratocyst	5.33 (1.11)	3.20 (1.20)	7.87 (0.61)	<b>2.33</b> <b>(1.70)</b>	2.35 (1.58)	3.78 (0.50)	4.15 (0.79)
	Orthokeratinized odontogenic cyst	5.43 (1.50)	4.83 (1.31)	6.65 (0.79)	<b>2.76</b> <b>(2.25)</b>	3.06 (2.41)	5.24 (1.49)	4.41 (1.14)
	Basal cell adenoma	3.28 (0.82)	5.26 (0.66)	8.65 (0.61)	<b>2.54</b> <b>(0.70)</b>	3.67 (0.83)	4.39 (0.67)	4.89 (0.73)
	Adenoid cystic carcinoma	6.91 (1.58)	5.56 (1.46)	9.06 (0.51)	<b>5.35</b> <b>(1.96)</b>	5.59 (2.34)	7.35 (1.81)	5.96 (1.49)
	Mucoepidermoid carcinoma	<b>5.61</b> <b>(0.83)</b>	7.13 (1.21)	8.37 (0.68)	6.61 (1.05)	6.65 (1.15)	8.76 (0.39)	8.50 (0.72)
	Warthin's tumor	5.74 (1.83)	4.69 (2.28)	8.65 (0.75)	2.28 (1.51)	<b>2.26</b> <b>(1.06)</b>	6.81 (1.81)	7.69 (1.45)
	Odontogenic fibroma	8.65 (0.75)	7.63 (0.90)	9.59 (0.41)	7.98 (1.05)	7.78 (1.38)	<b>6.78</b> <b>(0.44)</b>	8.80 (0.54)
	Ameloblastoma	8.13 (1.32)	7.00 (0.83)	7.26 (0.59)	<b>4.96</b> <b>(1.65)</b>	6.22 (1.32)	7.80 (0.80)	6.64 (0.50)
	Hemangioma	7.33 (0.65)	7.69 (1.15)	8.24 (1.60)	6.59 (1.91)	<b>6.46</b> <b>(1.21)</b>	7.46 (1.16)	7.33 (0.80)
	<b>Average</b>	6.43 (1.55)	5.88 (1.39)	8.01 (0.99)	<b>4.66</b> <b>(1.96)</b>	4.85 (1.87)	6.40 (1.53)	6.39 (1.75)
Query	Category	Pre-trained VGG16	Pre-trained DINOv2	Fine-tuned Resnet18	ResNet18 + SimCLR	ResNet18 + TiCo	Ciga model	Phikon
Out-of-domain-B	Myoepithelioma (n = 60 images from 1 case)	5.88 (1.16)	5.58 (0.49)	5.98 (0.18)	5.67 (0.68)	<b>4.55</b> <b>(0.44)</b>	6.17 (0.84)	5.28 (1.40)
	Basal cell adenoma (n = 60 images from 1 case)	3.63 (2.04)	4.57 (0.36)	3.82 (1.96)	<b>1.70</b> <b>(0.09)</b>	1.90 (0.61)	3.73 (0.75)	1.82 (0.08)
	Warthin's tumor (n = 60 images from 1 case)	3.55 (1.26)	2.23 (0.58)	7.70 (2.46)	<b>1.12</b> <b>(0.26)</b>	1.75 (0.35)	7.07 (0.19)	2.72 (0.87)
	Carcinoma ex pleomorphic adenoma (n = 60 images from 1 case)	5.45 (0.64)	5.22 (1.81)	7.17 (1.07)	3.68 (0.98)	4.07 (0.58)	4.93 (1.09)	<b>3.28</b> <b>(1.64)</b>
	Mucoepidermoid carcinoma (n = 120 images from 2 cases)	<b>5.55</b> <b>(1.41)</b>	6.29 (0.56)	6.98 (1.24)	6.22 (1.16)	6.24 (1.34)	6.71 (0.87)	6.93 (1.11)

	Adenoid cystic carcinoma (n = 120 images from 2 cases)	<b>2.93</b> <b>(1.35)</b>	<i>4.03</i> <i>(1.36)</i>	5.96 (0.81)	4.38 (2.78)	4.89 (2.48)	5.88 (2.37)	4.18 (2.55)
	Acinic cell carcinoma (n = 120 images from 2 cases)	5.18 (2.50)	<b>3.94</b> <b>(2.16)</b>	6.96 (0.88)	5.03 (1.87)	4.26 (2.21)	6.21 (2.36)	<i>4.11</i> <i>(1.35)</i>
	Salivary duct carcinoma (n = 60 images from 1 case)	4.57 (0.53)	4.62 (0.58)	8.97 (0.68)	<i>4.10</i> <i>(0.41)</i>	4.40 (1.02)	4.80 (2.98)	<b>3.40</b> <b>(1.78)</b>
	<b>Average</b>	4.59 (1.10)	4.56 (1.23)	6.69 (1.51)	<b>3.99</b> <b>(1.80)</b>	<i>4.01</i> <i>(1.50)</i>	5.69 (1.11)	4.27 (2.23)
<b>Query</b>	<b>Category</b>	<b>Pre-trained VGG16</b>	<b>Pre-trained DINOv2</b>	<b>Fine-tuned Resnet18</b>	<b>ResNet18 + SimCLR</b>	<b>ResNet18 + TiCo</b>	<b>Ciga model</b>	<b>Phikon</b>
Out-of-domain-C	Adenoid cystic carcinoma (n = 180 images from 3 cases)	<i>4.66</i> <i>(1.15)</i>	5.14 (1.29)	8.03 (0.82)	5.23 (1.69)	5.06 (1.28)	6.87 (0.67)	<b>4.19</b> <b>(0.93)</b>
	Basal cell adenoma (n = 180 images from 3 cases)	<b>2.90</b> <b>(1.35)</b>	5.03 (1.03)	8.81 (0.69)	5.51 (2.68)	4.65 (2.95)	5.12 (1.62)	<i>4.62</i> <i>(1.76)</i>
	Odontogenic myxoma/myxofibroma (n = 180 images from 3 cases)	4.75 (2.36)	2.98 <i>(1.31)</i>	8.09 (0.82)	3.38 (1.52)	<b>2.42</b> <b>(1.54)</b>	4.02 (1.29)	4.58 (1.67)
	Fibrous dysplasia (n = 120 images from 2 cases)	5.32 (1.02)	<b>3.80</b> <b>(0.63)</b>	<i>4.18</i> <i>(0.35)</i>	4.19 (0.87)	4.57 (0.70)	5.52 (0.67)	5.22 (1.14)
	Osteoma (n = 180 images from 3 cases)	<i>3.05</i> <i>(0.86)</i>	3.32 (0.86)	3.21 (0.35)	3.08 (0.88)	3.07 (0.40)	3.59 (0.73)	<b>2.59</b> <b>(1.49)</b>
	Odontogenic keratocyst (n = 120 images from 2 cases)	<i>5.61</i> <i>(1.80)</i>	<b>4.45</b> <b>(0.95)</b>	9.58 (0.68)	5.91 (2.69)	6.44 (3.06)	5.67 (2.03)	5.73 (1.92)
	Orthokeratinized odontogenic cyst (n = 120 images from 2 cases)	5.68 (1.21)	2.93 (0.46)	9.69 (0.27)	<b>1.58</b> <b>(0.85)</b>	<i>1.91</i> <i>(1.09)</i>	4.53 (1.45)	3.86 (2.77)
	Adenomatoid odontogenic tumor (n = 120 images from 2 cases)	5.94 (1.23)	6.76 (0.73)	6.38 (1.25)	<b>4.78</b> <b>(1.11)</b>	5.60 (1.26)	7.48 (1.14)	<i>5.24</i> <i>(1.86)</i>
	<b>Average</b>	4.74 (1.17)	4.30 (1.32)	7.25 (2.44)	<b>4.21</b> <b>(1.45)</b>	<b>4.21</b> <b>(1.59)</b>	5.35 (1.34)	4.40 (1.86)



**Figure 8: Comparison of the top-10 results of all models for out-of-domain-phonecam queries from different categories**

ResNet18+SimCLR and ResNet18+TiCo are consistent with the result that provides the highest Acc for different query categories. More examples are presented in Figure 9 (Green outline: accurate diagnosis category; Yellow outline: differential diagnosis categories; Red outline: inaccurate diagnosis category)



**Figure 9: Further examples of comparisons of the top 10 results of all models for out-of-domain-phonecam queries from different categories**

ResNet18+SimCLR and ResNet18+TiCo are consistent with the result providing the highest Acc for different query categories. The categories of retrieved images belonging to the differential diagnoses show which retrieved images have histologic similarity to the query. This comparison demonstrates the SSL model's capability to retrieve histologically similar images when the exact accurate diagnosis is not retrieved. The pre-trained DINOv2 and Phikon also show such potential albeit less consistently across the query category than the SSL models. (Green outline: accurate diagnosis category; Yellow outline: differential diagnosis categories; Red outline: inaccurate diagnosis categories)

## Discussion

The diagnosis of oral pathology has long depended on histopathological image observation, which can be a burden for pathologists, especially when dealing with rare cases. In the last decade, various machine learning methods have been proposed to aid in histopathological

diagnosis and improve speed and accuracy to avoid delays in diagnosis. However, the exploration for oral histopathology diagnosis has been hampered by the difficulty of obtaining an adequately extensive database that includes rare cases and constructing an effective model. To our knowledge, this is the first study to construct a large database of 30 oral tumor categories, with an additional 21 categories used as the model-training dataset.

Many have argued that CBIR has greater advantages in this field. Pathologists can review CBIR results to make a final decision. However, decision bias may occur when the algorithm is unreliable. To find the best way to represent images for CBIR, we compared different methods of training the feature encoder. We then ranked the similarity of all images in the database to the test queries. The gradual concept of similarity and the multiple-ranked results of CBIR pose a challenge in interpretation. Four evaluation measures were used: MeanAUC from the whole database similarity rank; MeanAcc, %query, and MeanHI from the top-10 most similar images. MeanAUC assumes all rank cut-points are relevant to model performance in extracting histologic features, while the top-10 similar results are relevant during image observation by pathologists in future CBIR implementation. Our findings suggest that model training for feature extraction using an in-house dataset with SSL methods outperforms other popular methods in retrieving images with an accurate diagnosis and similar histology: in-domain queries in 73% of categories and out-of-domain-phonocam queries in 64% of categories (Table 3). This was supported by their MeanAcc which was superior in 63% of in-domain query categories and 82% of out-of-domain-phonocam query categories (Table 4). There was no significant difference in MeanAcc between the SimCLR and TiCo models for both query categories. The MeanAcc superiority of the SSL models was consistent at the in-domain low and moderate magnification levels (Figure 7A) and at out-of-domain-phonocam at all magnification levels (Figure 7B). Additionally, both SSL models retrieved fewer images without histologic similarity at all magnification levels for both query categories (Figures 7C, 7D), meaning when low accuracy is achieved in the result, the users get several options that are histologically similar upon CBIR implementation with the SSL models because they belong to the textbook differential diagnosis categories. From this result, users could proceed with the additional tests more easily than having to do the preliminary reference search manually. Our dataset had overlapping image patches with similar histologic features. In this situation, the SSL method was superior because it compensated for the lack of a labeled dataset for learning representations that cluster the data during training based on semantic classes in conjunction with convolutional neural networks as feature extractors, regardless of the category [20].

The most impressive performance was shown for Warthin's tumor query, with MeanAUC values greater than 0.960 in every query set using the SSL models (Table 3). Histologically, Warthin's tumors consist of varying proportions of papillary cystic structures lined by two layers of oncocytic epithelial cells and a lymphoid stroma with germinal centers. It is one of the most common tumors of the salivary gland, especially the parotid gland, and is generally easy to diagnose microscopically owing to its characteristic pattern [19].

The SLL models were successful for most of the test query categories. Out of those categories, the MeanAcc of the SLL model for the ameloblastoma query was lower (Table 3). Although ameloblastoma is one of the most common odontogenic tumors, it has diverse histologic variants: follicular, plexiform, acanthomatous, granular, basaloid, desmoplastic, or a mixture of these [17]. This diversity requires an adequate representation of each subtype in the database for greater accuracy. However, the %query indicated that the models retrieved a similar

ameloblastoma type in the top-10 for more than 93% of the queries tested (Table 5). With several differential diagnoses of ameloblastoma included in the database, the best MeanHI obtained for the in-domain ameloblastoma query was 5.98 by SimCLR (Table 6). These categories can be considered histologically similar only if the characteristics of certain subtypes are captured. For example, islands of odontogenic epithelium with ameloblastic features in the follicular type may resemble ameloblastic fibroma [19]. Updating the database with newly encountered subtypes continuously would improve the accuracy of rare tumor subtypes.

Although CBIR works by retrieving similar images that can be considered a digital second opinion, the result may contain images from different categories, with many having similar or indistinguishable histology. Arguably, the range of MeanAcc values obtained with the SSL models, 1.00 to 8.94 (Tables 3 and 4), is considerably wide. However, 55% to 100% of the total queries retrieved at least one of 10 images from the correct category (Table 5), and 6.46 to 1.12 out of the 10 images had no histologic similarity to the query image (Table 6). This implies that displaying the complete top-10 results, including the correct diagnosis and differential diagnosis, as shown in Figures 8 and 9, could be significant for pathologists to narrow the differential diagnoses and conduct further research efficiently. To further improve usability, it is necessary to include clinical and other findings, such as the location of the tumor, patient history, and diagnostic criteria, which are usually essential to making a diagnosis by pathologists, when developing a CBIR system, especially in the oral region where tissue types are diverse.

This study implements patch-based CBIR. Some CBIR systems can analyze WSIs of which implementation is prospective in developed countries. As expensive WSI scanners are not universally installed in oral laboratories, the image-capturing equipment accessible to pathologists differs considerably across regions. Microscope images captured directly using a smartphone camera could be the easiest mode for education, image sharing, and case consultation [27,28]. By using patch-based CBIR where pathologists only need to select the tumor areas and capture them with smartphone cameras to create input, this technology is more accessible globally. Variations in image color and resolution resulting from these differences hinder obtaining reliable results. We tested the robustness of each model to domain shifts by testing the models on out-of-domain queries using WSIs from multiple institutions captured by different scanners and smartphone cameras. SSL models performed best for most query categories, with SimCLR or TiCo achieving the best MeanAUC for over 68% of out-of-domain query categories, from 0.839 to 0.999 (Table 3), confirming the previous finding that SSL is more robust to domain shifts than supervised learning in some datasets, including pathological images [29]. Interestingly, the performances of the vision transformer models (pre-trained DINOv2 and Phikon) always come second best to the SSL models in MeanAUC and are comparable in MeanAcc to that of the SSL models on out-of-domain query sets (Tables 3 and 4). Although further investigation is needed, this result may be considered when choosing methods for a CBIR system. If the system is designed for an in-house database and query or a scenario applicable in large hospitals, SSL models trained on in-house cases are the optimal choice. However, where the system is designed to handle out-of-domain queries, using a pre-trained vision transformer model becomes a viable alternative to eliminate the need to train the SSL model, which could be computationally expensive.

## Study limitations

Limitations of this study include that the experiments involved a single query to retrieve similar images. An algorithm that supports more information in the query, such as multiple query algorithms and filters for location or other diagnostic criteria, would improve retrieval accuracy and provide better support for diagnosis. Our study is limited to test queries from the same geographical area as the SSL model-training dataset. Collecting query cases from a more diverse area would be beneficial in future CBIR development to further challenge the generalizability of the result. The comparative methodology did not emphasize histopathology characteristics that differentiate between benign or malignant tumors, such as capsule invasion and mitotic activity in basal cell adenoma vs. basal cell adenocarcinoma but focused on how such image retrieval tools would be beneficial in reducing to several differential diagnoses and recalling diagnosis criteria before following up with ancillary tests if necessary. Image retrieval is less likely to mislead decision-makers owing to model overfit than a conventional classification method that predicts the possible tumor diagnosis. Nonetheless, a sequel of observations would still be needed when image retrieval is utilized. This study provides insights as the first step to developing a CBIR algorithm by observing retrieval accuracy with strict tumor category criteria on a relatively small database and did not investigate the impact of the result on decision-making in clinical settings. The implementation of CBIR as a well-rounded system to be incorporated into the comprehensive diagnostic process is beyond the scope of this study and observation of the interaction between pathologists and a CBIR system for common and rare diagnoses is needed before the system is used in clinical settings.

## Conclusion

This study highlighted various methods to develop an effective CBIR model and presented key measures to determine the best approach for future clinical usage. We have shown that using SSL methods for deep neural network training is an effective way to develop a CBIR system for histopathological diagnosis of oral tumors compared to other commonly used methods. Vision transformer models, though slightly less effective than SSL models, still provided strong performance and could be a viable alternative for out-of-domain queries. These approaches have considerable potential to create a clinically useful image retrieval system that accelerates the diagnostic process and improves accuracy.

## Author Contributions

R. R. Herdiantoputri: Contributed to the conception, design, data acquisition, drafting, and critical revision of the manuscript. D. K. contributed to the conception, design, drafting, and critical revision of the manuscript. M. Ochi: Contributed to the design and critical revision of the manuscript. Y. Fukawa: Contributed to data acquisition and critically revised the manuscript. K. Kayamori: Contributed to data acquisition, and critically revised the manuscript., M. Tsuchiya: Contributed to data acquisition, and critically revised the manuscript. Y. Kikuchi: Contributed to data acquisition and critically revised the manuscript. T. Ushiku: Contributed to the data acquisition and critically revised the manuscript. T. Ikeda: Contributed to the data acquisition and critically revised the manuscript. Ishikawa: Contributed to the conception, design, drafting, and critical revision of the manuscript. All authors provided their final approval and agreed to be accountable for all aspects of this work.



## Acknowledgments

We thank Editage ([www.editage.jp](http://www.editage.jp)) for the English language review.

We thank the pathologists who contributed to the image verification.

## Declaration of Conflicting Interests

The authors declare no potential conflicts of interest concerning the research, authorship, or publication of this article.

## Funding

This study was supported by AMED Practical Research for Innovative Cancer Control under grant number JP 23ck0106640 to S.I. and the JSPS KAKENHI Grant-in-Aid for Scientific Research (B) under grant number 21H03836 to D.K.

## References

1. Kiehl T-R: Digital and Computational Pathology: A Specialty Reimagined. In: Ehsani S, Glauner P, Plugmann P, Thieringer FM, eds. *The Future Circle of Healthcare*. Springer International Publishing, Cham; 2022. 10.1007/978-3-030-99838-7\_12
2. Tommasino C, Merolla F, Russo C, Staibano S, Rinaldi AM: Histopathological Image Deep Feature Representation for CBIR in Smart PACS: *J Digit Imaging*. 2023, 36:2194-209. 10.1007/s10278-023-00832-x
3. Rao RS, Shivanna DB, Lakshminarayana S, et al.: Ensemble Deep-Learning-Based Prognostic and Prediction for Recurrence of Sporadic Odontogenic Keratocysts on Hematoxylin and Eosin Stained Pathological Images of Incisional Biopsies. *J Pers Med*. 2022, 12:1220. 10.3390/jpm12081220
4. Giraldo-Roldan D, Ribeiro ECC, Araújo ALD, et al.: Deep learning applied to the histopathological diagnosis of ameloblastomas and ameloblastic carcinomas. *J Oral Pathology Medicine*. 2023, 13481:10.1111/jop.13481
5. Kumar N, Gupta R, Gupta S: Whole Slide Imaging (WSI) in Pathology: Current Perspectives and Future Directions. *J Digit Imaging*. 2020, 33:1034-40. 10.1007/s10278-020-00351-z
6. Van Der Laak J, Litjens G, Ciompi F: Deep learning in histopathology: the path to the clinic. *Nat Med*. 2021, 27:775-84. 10.1038/s41591-021-01343-4
7. Hameed IM, Abdhussain SH, Mahmmod BM: Content-based image retrieval: A review of recent trends. *Cogent Engineering*. 2021, 8:1927469. 10.1080/23311916.2021.1927469
8. Dehkharghanian T, Bidgoli AA, Riasatian A, et al.: Biased data, biased AI: deep networks predict the acquisition site of TCGA images. *Diagn Pathol*. 2023, 18:67. 10.1186/s13000-023-01355-3
9. Springenberg M, Frommholz A, Wenzel M, Weicken E, Ma J, Strodthoff N: From modern CNNs to vision transformers: Assessing the performance, robustness, and classification strategies of deep learning models in histopathology. *Medical Image Analysis*. 2023, 87:102809. 10.1016/j.media.2023.102809
10. Kalra S, Tizhoosh HR, Choi C, Shah S, Diamandis P, Campbell CJV, Pantanowitz L: Yottixel - An Image Search Engine for Large Archives of Histopathology Whole Slide Images. *Medical Image Analysis*. 2020, 65:101757. 10.1016/j.media.2020.101757

11. Albashish D: Ensemble of adapted convolutional neural networks (CNN) methods for classifying colon histopathological images. *PeerJ Computer Science*. 2022, 8:1031. [10.7717/peerj-cs.1031](https://doi.org/10.7717/peerj-cs.1031)
12. Komura D, Kawabe A, Fukuta K, et al.: Universal encoding of pan-cancer histology by deep texture representations. *Cell Reports*. 2022, 38:110424. [10.1016/j.celrep.2022.110424](https://doi.org/10.1016/j.celrep.2022.110424)
13. Wang X, Yang S, Zhang J, et al.: Transformer-based unsupervised contrastive learning for histopathological image classification. *Medical Image Analysis*. 2022, 81:102559. [10.1016/j.media.2022.102559](https://doi.org/10.1016/j.media.2022.102559)
14. Ciga O, Xu T, Martel AL: Self supervised contrastive learning for digital histopathology. *Machine Learning with Applications*. Machine Learning with Applications. 2022, 7:100198. [10.1016/j.mlwa.2021.100198](https://doi.org/10.1016/j.mlwa.2021.100198)
15. Shmatko A, Ghaffari Laleh N, Gerstung M, Kather JN: Artificial intelligence in histopathology: enhancing cancer research and clinical oncology. *Nat Cancer*. 2022, 3:1026-38. [10.1038/s43018-022-00436-4](https://doi.org/10.1038/s43018-022-00436-4)
16. Ben-Shaul I, Shwartz-Ziv R, Galanti T, Dekel S: Reverse Engineering Self-Supervised Learning. *arXiv*. 2023, [10.48550/ARXIV.2305.15614](https://arxiv.org/abs/10.48550/ARXIV.2305.15614)
17. Organisation mondiale de la santé, Centre international de recherche sur le cancer: WHO classification of head and neck tumours, 4th Edition, Volume 9 . IARC, Lyon; 2017.
18. Herdiantoputri RR, Komura D, Fujisaka K, Ikeda T, Ishikawa S: Deep texture representation analysis for histopathological images. *STAR Protocols*. *STAR Protocols*. 2023, 4:102161. [10.1016/j.xpro.2023.102161](https://doi.org/10.1016/j.xpro.2023.102161)
19. Gnepp DR, Bishop JA: Gnepp's Diagnostic Surgical Pathology of the Head and Neck. 3. Edition. Elsevier, Philadelphia; 2020.
20. Bishop JA, Thompson LDR, Wakely PE, Weinreb I: Tumors of the salivary glands. American Registry of Pathology, Arlington, VA; 2021.
21. Chen T, Kornblith S, Norouzi M, Hinton G: A Simple Framework for Contrastive Learning of Visual Representations. *arXiv*. 2020, [10.48550/arXiv.2002.05709](https://arxiv.org/abs/10.48550/arXiv.2002.05709)
22. Liu M, Yi M, Wu M, Wang J, He Y: Breast Pathological Image Classification Based on VGG16 Feature Concatenation. *J Shanghai Jiaotong Univ (Sci)*. 2022, 27:473-84. [10.1007/s12204-021-2398-x](https://doi.org/10.1007/s12204-021-2398-x)
23. Zhou P, Cao Y, Li M, et al.: HCCANet: histopathological image grading of colorectal cancer using CNN based on multichannel fusion attention mechanism. *Sci Rep*. 2022, 12:15103. [10.1038/s41598-022-18879-1](https://doi.org/10.1038/s41598-022-18879-1)
24. Filiot A, Ghermi R, Olivier A, et al.: Scaling Self-Supervised Learning for Histopathology with Masked Image Modeling. *medRxiv*. 2023, [10.1101/2023.07.21.23292757](https://doi.org/10.1101/2023.07.21.23292757)
25. Zhu J, Moraes RM, Karakulak S, Sobol V, Canziani A, LeCun Y: TiCo: Transformation Invariance and Covariance Contrast for Self-Supervised Visual Representation Learning. *arXiv*. 2022, [10.48550/ARXIV.2206.10698](https://arxiv.org/abs/10.48550/ARXIV.2206.10698)
26. Oquab M, Darcet T, Moutakanni T, et al.: DINOv2: Learning Robust Visual Features without Supervision. *arXiv*. 2023, [10.48550/arXiv.2304.07193](https://arxiv.org/abs/10.48550/arXiv.2304.07193)
27. Morrison AO, Gardner JM: Microscopic Image Photography Techniques of the Past, Present, and Future. *Archives of Pathology & Laboratory Medicine*. 2016, 140:618-618. [10.5858/arpa.2016-0903-CO](https://doi.org/10.5858/arpa.2016-0903-CO)
28. Maekawa Y, Tanahashi N, Yoneda M, Morishita Y: Smartphone-based imaging technology for histopathology image education in lieu of a tabletop microscope. *Japanese Journal of Medical technology*. 2023, 72:614-618. [10.14932/jamt.23-31](https://doi.org/10.14932/jamt.23-31)

29. Shi Y, Torr PHS, Daunhauer I, Vogt JE, Sanyal A: How Robust is Unsupervised Representation Learning to Distribution Shift. arXiv. 2022, 10.48550/arXiv.2206.08871