



OPEN

Creating an atlas of normal tissue for pruning WSI patching through anomaly detection

Peyman Nejat¹, Areej Alsaafin¹, Ghazal Alabtah¹, Nneka I. Comfere², Aaron R. Mangold³, Dennis H. Murphree^{2,4}, Patricija Zot⁵, Saba Yasir⁵, Joaquin J. Garcia⁵ & H. R. Tizhoosh^{1✉}

Patching whole slide images (WSIs) is an important task in computational pathology. While most of them are designed to classify or detect the presence of pathological lesions in a WSI, the confounding role and redundant nature of normal histology are generally overlooked. In this paper, we propose and validate the concept of an “atlas of normal tissue” solely using samples of WSIs obtained from normal biopsies. Such atlases can be employed to eliminate normal fragments of tissue samples and hence increase the representativeness of the remaining patches. We tested our proposed method by establishing a normal atlas using 107 normal skin WSIs and demonstrated how established search engines like Yottixel can be improved. We used 553 WSIs of cutaneous squamous cell carcinoma to demonstrate the advantage. We also validated our method applied to an external dataset of 451 breast WSIs. The number of selected WSI patches was reduced by 30% to 50% after utilizing the proposed normal atlas while maintaining the same indexing and search performance in leave-one-patient-out validation for both datasets. We show that the proposed concept of establishing and using a normal atlas shows promise for unsupervised selection of the most representative patches of the abnormal WSI patches.

Pathology is the field of studying the cause, development, and effects of diseases. This field involves examining sampled tissues, cells, or bodily fluids to diagnose pathologic disorders or help with determining the prognosis of a disorder's progression. In examining tissue samples of solid tumors, microscopes have been the most commonly used tool in this field before the introduction of digital scanners¹ to convert glass slides with mounted processed and fixed tissue samples into digital images. This digitization introduced a new concept to the field commonly referred to as *digital pathology*². While digital pathology is the general term used to describe integrating innovative digital tools into pathology, computational pathology describes a field that uses computational techniques and algorithms to analyze and interpret pathology data with the goal of improving the diagnosis, prognosis, and treatment of diseases³.

As tissue glass slides constitute a major source of information in pathology, whole slide imaging and whole slide image (WSI) analysis comprise a major and new subset of computational pathology. The current market of digital scanners is very diverse with most of the commercially available digital scanners being able to capture whole slide images at magnifications as high as 40X, generating extremely large image files⁴. While storing and leveraging the information in such a huge amount of data can be challenging, the stored retrospective data can be effectively used to introduce assistive tools into digital pathology increasing the speed and accuracy of pathologists hence reducing the workload. Artificial intelligence (AI) has shown promise in utilizing mass medical archives.

Deep learning uses artificial neural networks to learn complex patterns in any type of data⁵. In contrast to conventional machine learning which mainly uses structured data, e.g. tabular data, the so-called deep models (deep artificial neural networks) can take in unstructured data including text, sound, and image, without pre-processing or pre-extraction of features. This makes deep networks extremely useful in real-life applications, including medicine. Deep models are mainly composed of multiple layers of connected artificial neurons. The general workflow of training and using a deep model in a supervised manner includes presenting the model with enough samples so the model can be properly trained to perform the desired task on new cases⁵.

¹KIMIA Lab, Department of Artificial Intelligence and Informatics, Mayo Clinic, Rochester, MN, USA. ²Department of Dermatology, Mayo Clinic, Rochester, MN, USA. ³Department of Dermatology, Mayo Clinic, Scottsdale, AZ, USA. ⁴Department of Artificial Intelligence and Informatics, Mayo Clinic, Rochester, MN, USA. ⁵Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA. ✉email: tizhoosh.hamid@mayo.edu

It has been shown that searching in an indexed dataset with image-level labels may provide a base for computational consensus^{6–9}. An index dataset can be called an “atlas” of medical information where previously established knowledge is stored and can be retrieved for patient matching. In such an atlas, newly encountered cases, i.e., new patients, are indexed using the same methods and are matched against all the atlas cases to find the most similar patients. Diagnosis, staging, prognosis, and any other decision or prediction can be then inferred from the top similar atlas patients for which all outcomes are known.

Indexing visual content as generating a vector is the major task for atlas creation. While deep networks are mainly used as classifiers to provide definite labels on the input data, it has been shown that the numerical representations of some pre-decision layers of the network, also called embeddings or deep features, contain important information about the content of the input data^{10,11}. The concept of image retrieval in an atlas using deep features has been used in both radiology and pathology^{7,8,12}. Any deep network can be used to obtain deep features, but the architecture of the network and the data it has been trained on can greatly affect the representativeness of the features for the task at hand. For instance, KimiaNet is a dense convolutional neural network fine-tuned using all diagnostic histopathology images from the TCGA repository¹³. Vision transformers have also shown a promising ability to capture the details in a given image^{14,15}.

While deep networks are widely used for generating deep features from images, they also impose limitations regarding image size. State-of-the-art graphical processing units (GPUs) which are commonly used in deep learning can process up to a certain size of the image in each iteration. Therefore, any whole slide image (WSI) processing requires the image to be broken down into many smaller images, also known as *patches*¹⁶. While patching gigapixel WSIs is a crucial step in computational pathology, one WSI can generate several hundred and even several thousands of patches of ordinary size, say 224 by 224 or 512 by 512 pixels, depending on the size of the WSI, magnification, and the biopsy type. One approach would be to feed all the patches into the deep network and use the output for the downstream tasks. Such a *brute-force* approach would require massive computational power, a factor that would limit the adoption of the technology.

Processing WSIs is a difficult problem that requires a *divide & conquer* approach. Patching is generally the divide and extracting deep features is the conquer part. Various methods have been proposed to select a smaller subset of WSI patches as representative ones for downstream tasks. One approach as described in Yottixel is clustering patches based on color distribution or RGB histogram and proximity of patches in WSI to select a diverse but smaller set of patches from all morphological structures within a WSI⁷. This is called a *mosaic* of WSI. The Yottixel mosaic concept is simple and reliable and has been adopted by other works^{17,18}. Still, patch selection remains one of the open fields for research because it can greatly affect the performance of the downstream analyses. Yottixel’s mosaic ignores normal histology and selects normal and malignant patches based on their frequency in the clustering space.

As pathologists have been trained to recognize normal histologies and disregard them when focusing on the pathology, it can be helpful for a better patching strategy to remove normal histology specific to the site. One way to formulate this task is by integrating anomaly detection into the patching scheme. Provided WSIs depicting entirely normal tissue are available, one-class classifiers¹⁹ could learn the typical features of normal histology. The trained classifiers can then accompany the mosaic generation, i.e., the patching, to remove normal patches.

In this paper, we propose and validate an “atlas of normal tissue” solely using samples of WSIs obtained from normal tissue biopsies. The proposed method leverages a weakly supervised multiple-instance learning method using one-class classifiers to exclude the normal patches and focus on clinically important parts of any given WSI. We argue that this method will reduce the number of patches selected, hence removing redundancy, from each WSI while maintaining the representativeness of the patches for downstream tasks.

Experiments setup

Datasets

Two datasets from two different organs were used to test the proposed atlas of normal tissue and the one-class classifier for normal patch detection.

- One dataset included a total number of 660 skin tissue WSIs of patients diagnosed with **cutaneous squamous cell carcinoma** (cSCC) which were pulled from the internal Mayo Clinic database (REDCap). The cases were first identified through enterprise-wide internal search engines which identified patients with a histopathologic diagnosis of cSCC or metastatic cSCC from pathology reports of archived tissue specimens. These cases subsequently underwent chart review to confirm the primary tumor as part of the inclusionary criteria. Samples were either taken at Mayo Clinic locations (Minnesota, Arizona, Florida) or outside facilities and shared with Mayo Clinic for consultation. The tissue may be from either biopsy (punch, shave, etc.) or subsequent excision. All tissue included in the pilot underwent both initial review and re-review by a dermatopathologist at one of the Mayo Clinic sites. The purpose of the re-review was to confirm cSCC, tumor characteristics (depth, differentiation, PNI, etc.), and tumor stage, and to select cases most appropriate for scanning and sequencing.
- Selected cases included 386 well-differentiated, 100 moderately differentiated, and 67 poorly differentiated cSCC. There were also 107 normal WSIs selected to represent the normal skin tissue. The only associated label used is the degree of differentiation or being normal. A total number of 10 slides were randomly selected from well-differentiated and poorly differentiated cases for internal validation of the normal atlas at the patch level. A pathologist selected the regions with the most prominent abnormal morphology in these slides. The annotations were later used to generate normal and abnormal patches for validation purposes only.
- The second collected dataset included 21 breast tissue WSIs obtained from 8 patients with no prior diagnosis of breast cancer at the Mayo Clinic. For validation purposes, 78 WSIs from cases with lobular carcinoma

of the breast and 354 WSIs from cases with ductal carcinoma of the breast were obtained from the TCGA Research Network. These were all the diagnostic WSIs available on the repository for the two selected medical conditions at the time of collecting the data. Tissue samples from cases with no prior malignancy and no treatment at the time of sampling were selected to imitate the diagnostic pipeline of a deployed model. After data curation and data exploration, 2 WSIs were excluded from the ductal carcinoma class due to not having an object power attribute recorded in their file metadata.

Normal atlas creation

The idea of the normal atlas relies upon leveraging anomaly detection algorithms used in machine learning to *detect outliers* in a dataset. One-class classifiers are best suited for this task. Due to the complex nature of the data, we selected Isolation Forest²⁰ and one-class SVM²¹ as the abnormality detection mechanisms in this study. Isolation Forest isolates anomalies by constructing a random forest of decision trees. It identifies outliers by measuring how easily instances can be isolated in the tree structures, where anomalies are expected to require fewer splits to be separated from the majority of the data. One-class SVM, on the other hand, builds a model based on the characteristics of the majority class and then classifies instances as either belonging to that class or being outliers, making it particularly useful when dealing with imbalanced datasets. Both methods were presented with a series of deep features obtained from normal patches. The fitted classifiers were then used to classify normal and abnormal in a given set of deep features they had not seen before. Data preprocessing included tissue segmentation, patching, and color normalization. The tissue segmentation was carried out using an in-house trained U-NET segmentation model. This model is trained on low magnification thumbnails of WSIs only to segment the tissue vs background. This model is not intended to differentiate normal and abnormal parts of the tissue. After removing parts of the WSI with background only, patching was done at $20 \times$ magnification with a patch size of 1024 by 1024 and an overlap of 30% between adjacent patches, both in height and width. A minimum of 75 percent patch area coverage was selected as the criteria for excluding patching with insufficient tissue. All patches were separately fed to three different deep networks to obtain the deep features. These networks included one convolutional neural network with the architecture of DenseNet121 specifically trained on pathology patches named KimiaNet¹³, one pre-trained vision transformer using a method named DINO, and another convolutional neural network with the architecture of ResNet50 trained using the same DINO approach¹⁵.

When training the one-class classifiers with the normal data, the sample was assumed to be contaminated with some abnormal data. In the context of normal WSIs, this may include artifacts, tissue folds, stain sedimentation, pen markings, and out-of-focus slide areas. Both Isolation Forest and one-class SVM use methods to incorporate this assumption. One-class SVM utilizes a factor called ν to set an upper bound on the fraction of margin errors and a lower bound on the fraction of support vectors relative to the total number of training examples²². We set $\nu = 5\%$ of the total cases according to the expert's opinion and let the Isolation Forest classifier deduct this parameter automatically according to the method described in the original paper²⁰. The pseudocode of creating and using the normal atlas is outlined in algorithms 1 and 2. The pipeline of creating the normal atlas is also visually shown in Fig. 1. We used the Python library SciKit Learn for the implementation of Isolation Forest and one-class SVM²³. All the experiments were executed on a Linux-based server with two AMD EPYC 7413 CPUs and one NVIDIA A100 GPU with 80GB memory. The GPU was used for inferring deep features from the patches. Training One-class classifiers and using them as predicted were handled on the CPUs along with index and search.

```

Set  $I_N$  (set of normal WSIs)
Set  $W$  (deep network)
Set  $M$  (unfitted instance of one-class classifier)
Set  $s$  (patch size)
procedure CREATENORMALATLAS( $I_N$ )
   $F \leftarrow \emptyset$  ▷ Initialization
  for  $I \in I_N$  do
     $T \leftarrow \text{TissueSegmentation}(I)$  ▷ Extract the tissue
     $P_I \leftarrow \text{Patching}(I, s)$  ▷ Perform patching
     $P_T \leftarrow T \cap P_I$  ▷ Isolate patches containing tissue
     $F_I \leftarrow \text{DeepNet}(P_T, W)$  ▷ Extract features
     $F \leftarrow F_I \cup F$  ▷ Append features
  end for
   $M_F \leftarrow \text{FitOneClassClassifier}(F, M)$  ▷ Training
  return  $M_F$ 
end procedure

```

Algorithm 1. Pseudocode for creating the normal atlas given a set of normal WSIs I_N .

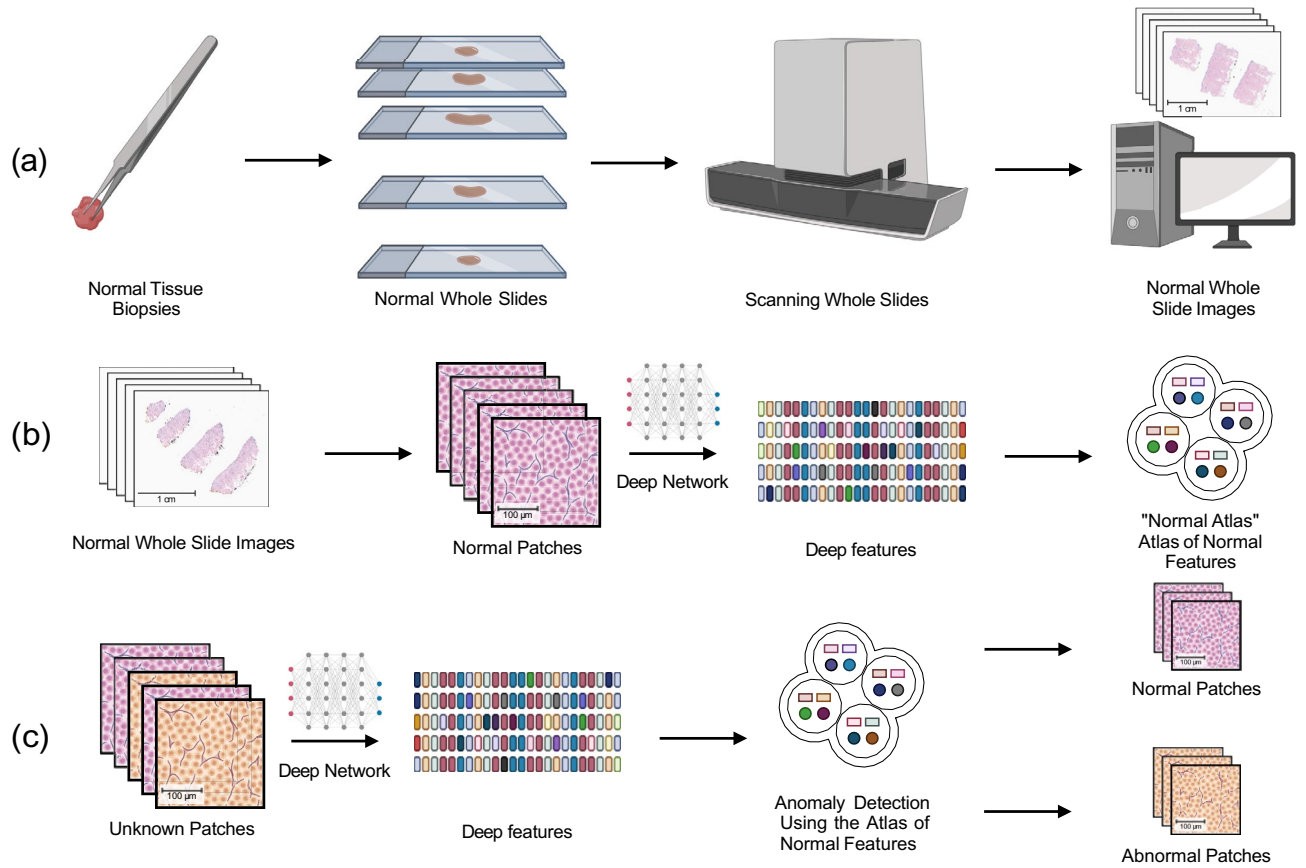


Figure 1. The general workflow of creating the normal atlas: (a) Normal WSIs are generated by scanning slides of known normal tissue biopsies, (b) Deep features of normal WSIs are obtained by passing the patches from each WSI through a deep network; the resulting deep features are stored in the normal atlas, (c) The constructed normal atlas is used to differentiate normal vs. abnormal deep features from unknown patches obtained using the same preprocessing and feature extraction method.

```

Set  $M_F$  (fitted one-class classifier)
Set  $P$  (query patch)
Set  $W$  (deep network)
procedure USENORMALATLAS( $P, M_F$ )
   $F \leftarrow \text{DeepNet}(P, W)$   $\triangleright$  Extract features
   $N \leftarrow \text{DetectAnomaly}(F, M_F)$   $\triangleright$  Is  $F$  normal
  if  $N$  is normal then
    return True  $\triangleright$  patch  $P$  is normal
  else
    return False  $\triangleright$  patch  $P$  is not normal
  end if
end procedure

```

Algorithm 2. Pseudo-code for using the normal atlas given a patch P and fitted one-class classifier M_F .

Index and search workflow

To demonstrate the role of removing the normal patches from the WSI, we have used the general workflow for indexing and searching WSIs for retrieval of similar cases as previously explained in Yottixel with some minor adjustments⁷. The tissue segmentation and patching follow the same method used to obtain the normal patches. To reduce the computational and storage cost, a subset of patches is then selected from each WSI using the patch selection method of Yottixel⁷. This method involves clustering patches into 9 clusters using a histogram of their RGB values and then selecting 15% of the patches from each cluster in a spatially homogenous manner. This results in selecting almost 15 percent of the total patches from each WSI, collectively called mosaic. The deep features for the patches of each WSI are used as an indexed version of that WSI. The median of minimum

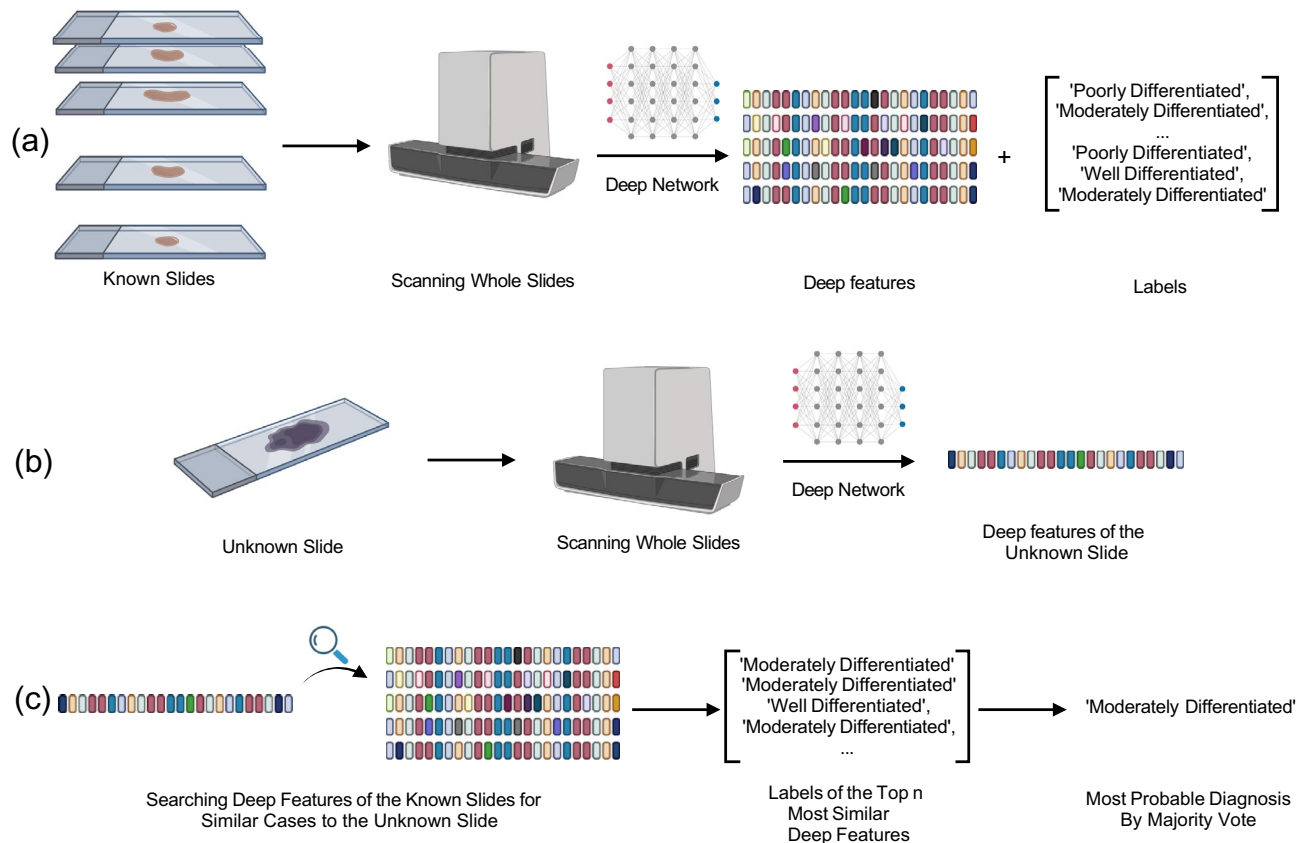


Figure 2. General workflow of index and search: (a) Deep features of known slides are obtained via passing the preprocessed patches through a deep network and are stored with their corresponding labels, (b) The deep features of a new (diagnostically unknown) slide are obtained through the same process, and (c) Stored deep features of the known slides are searched to find the closest matches to the newly generated deep features to extract the most probable label through a majority vote among the top-*n* closest labels.

one-to-one Euclidean distances between all the patches of the two WSI is used to find the most similar indexed WSIs to any given WSI. The same approach for obtaining similarity is used in Yottixel with the difference being the Hamming distance for barcoded features⁷ instead of the Euclidean between deep features which was used in our study. The schematic workflow of indexing and searching WSIs is shown in Fig. 2.

Internal and external validation

The performance of the normal atlas at the patch level was tested by obtaining normal and abnormal patches from 10 randomly selected cSCC test slides which were annotated by a pathologist. Having a minimum of 50 percent surface area of the patch annotated as the abnormal region was set as the threshold to label the patch as *abnormal*. The rest of the patches were considered normal. The patches were obtained in the same size and magnification as the original datasets used, i.e., 1024 by 1024 at 20X. This process yielded 6479 normal and 9352 abnormal patches. Precision, recall, and F1 score were selected as performance metrics for this classification task.

Three different setups of indexing and search pipeline were examined to test the effect of the normal atlas at the WSI level. The base setup included the original indexing and search pipeline for the WSIs as described in the Yottixel paper⁷. Two other setups were almost the same with one added step of excluding normal patches using an atlas of normal tissue after applying the Yottixel's mosaic to acquire patches. One setup used the Isolation Forest as the method of anomaly detection and another one used the one-class SVM. The resulting patches were also used to perform indexing and search in the same manner as the base setup. WSIs with all patches excluded as normal patches were labeled as normal. Performance of the indexing and search pipelines were measured using a leave-one-patient-out validation for all included patches by taking the output attached to top-1 and a majority vote among top-3 and top-5 most similar WSIs. Performance metrics such as precision, recall, and F1 score were calculated for each label inference. The overall performance of each configuration was also reported using the weighted F1 score of all the labels in the dataset according to their number of occurrences. To show the effect of using the normal atlas on computational time, we also measured the average time needed for a query search in the indexed WSI of the three selected setups. The added time for selection using the normal atlas per WSI was also recorded to determine the added overhead of using the one-class classifier.

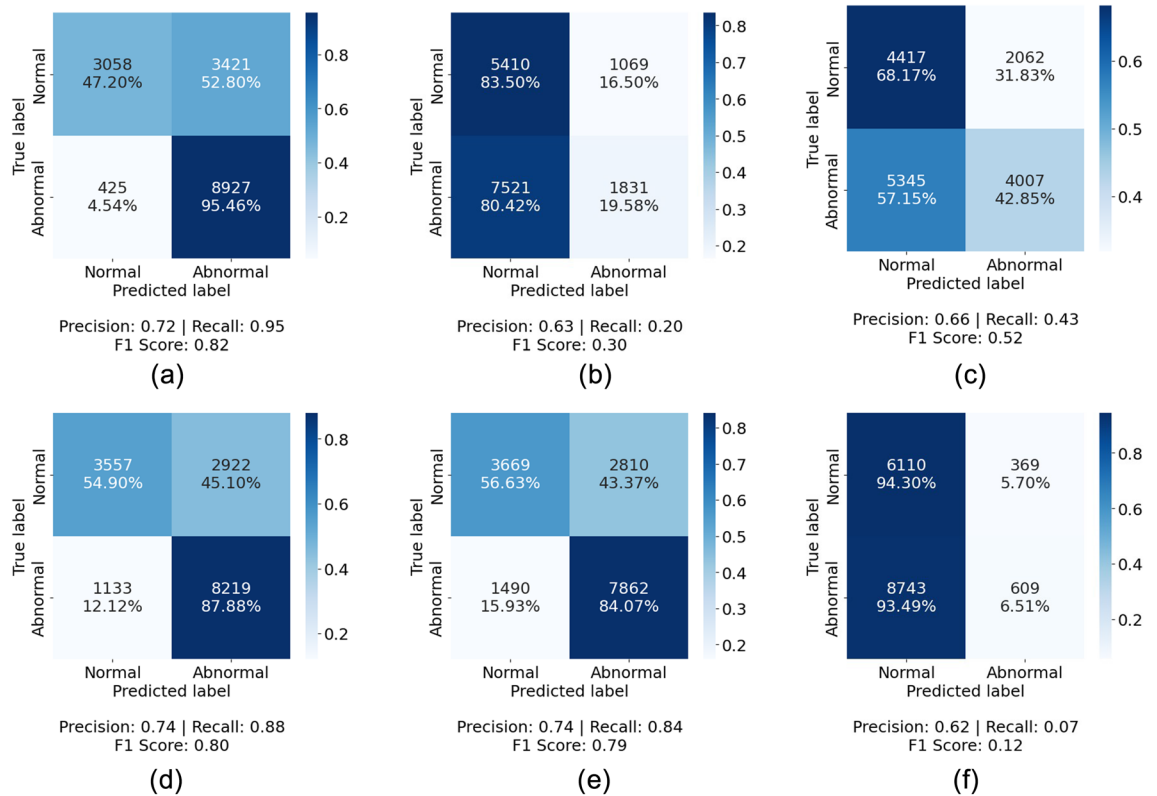


Figure 3. Normal atlas performance at patch-level demonstrated using validation results for 10 manually annotated WSIs by a pathologist shown as confusion matrices using normal atlases established using Isolation Forest (a–c) and one-class SVM (d–f) on deep features extracted by KimiaNet (a,d), ViT-DINO (b,e), and ResNet50-DINO (c,f).

Results

Normal Atlas patch-level validation

The patch-level internal validation for the normal atlases was carried out using the 10 previously randomly selected skin WSIs which were annotated for abnormal and normal regions by a pathologist. The performance varied based on the deep network used to obtain the deep features and the anomaly detection method used. KimiaNet's features showed the best performance both using Isolation Forest and one-class SVM, with F1 scores of 0.82 and 0.80, respectively. Among the networks trained with natural images, the DINO-trained vision transformer combined with one-class SVM showed the best results with an F1 score of 0.79. The detailed patch-level validation results for normal skin atlases are shown in Fig. 3. One visual example of how patch-level classification of different deep networks and anomaly detection algorithms comply with pathologist annotations is shown in Fig. 4 for a cSCC WSI and in Fig. 5 for a breast cancer WSI.

Indexing and search results

Indexing and search were carried out on both datasets of cSCC and breast cancer WSIs using the three previously described experiment setups. The detailed performance results of the indexing and search experiments are reported in Tables S1 and S2 in the supplementary section. As the confusion matrices for the leave-one-patient out classification depicted in Fig. 6 for the skin dataset show the two mostly mislabeled classes are 'well-differentiated' and 'moderately differentiated'. While this is true for the indexing and search using the deep features from all three deep networks, KimiaNet (trained specifically on pathology images¹³) shows the best performance. The classification, regardless of the network used to obtain the deep features, shows almost the same level of performance in all three setup configurations of using no normal atlas, using a normal atlas with Isolation Forest, and using a normal atlas with one-class SVM. To show this in the skin dataset, the F1 score for different experimental setups of the top-5 consensus is illustrated in Fig. 9. Confusion matrices of the indexing and search leave-one-patient-out validation for the breast dataset are shown in Fig. 7. Indexing and searching for breast carcinoma cases using deep features from all three models also show suboptimal performance for lobular carcinoma and infiltrating ductal carcinoma. However, the performance stays almost the same in three setups of not using the normal atlas, using one with Isolation Forest, and using one with a one-class SVM. This is shown in the F1 scores plot of the top-5 consensus of the corresponding experiments in Fig. 10. While the performance of the classification using the indexing and search stays almost the same in both skin and breast datasets, the number of average patches selected per WSI in the whole dataset to represent the WSIs is reduced by at least 12% and at most 86% of the original number of patches as shown in Table 1. The same table also depicts the average total time of search per WSI, which follows the same pattern seen in the number of patches. Supplementary Table S3

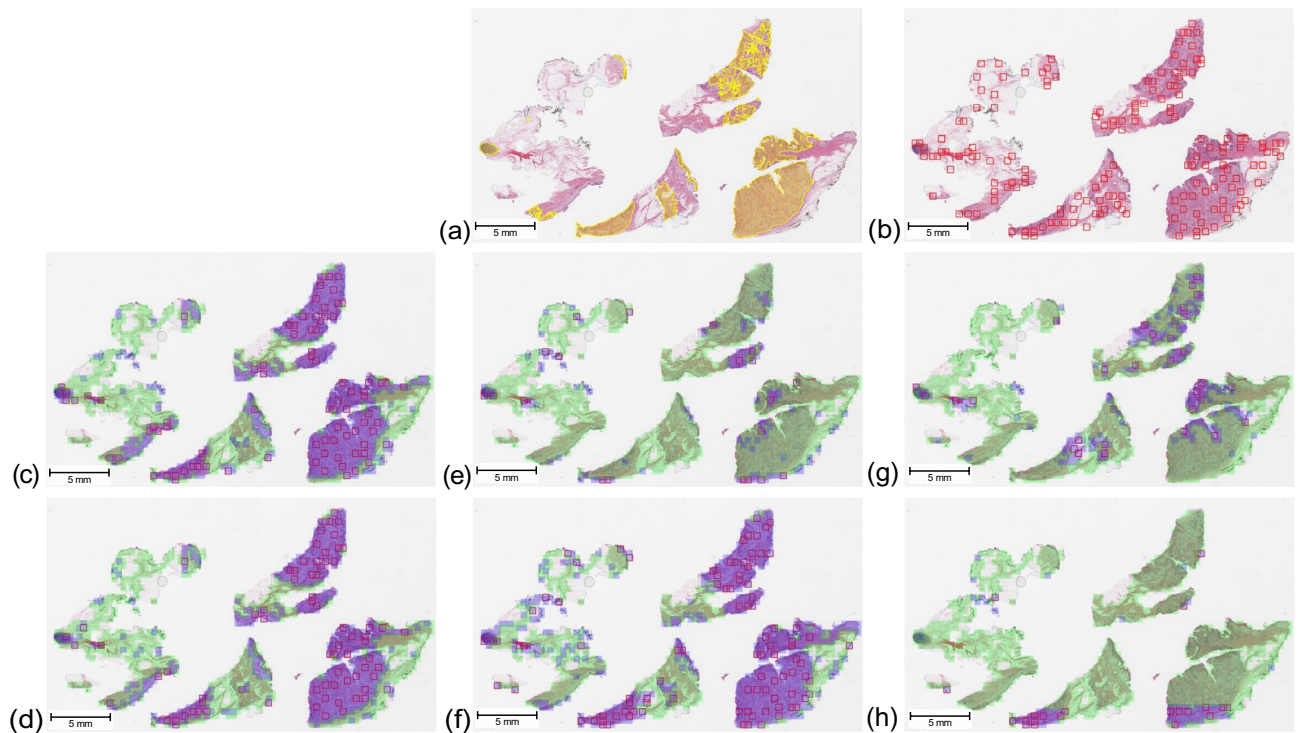


Figure 4. Visual results in a cSCC case: (a) Pathologist ground-truth, (b) Yottixel's mosaic, (c) KimiaNet and Isolation Forest, (d) ViT-DINO and Isolation Forest, (e) ResNet50-DINO and Isolation Forest, (f) KimiaNet and one-class SVM, (g) ViT-DINO and one-class SVM, (h) ResNet50-DINO and one-class SVM. Yellow regions depict the pathologist ground truth annotation for the tumor region. The green and purple areas show patches labeled normal and abnormal by the normal atlas, respectively. Red boxes show the patches selected for indexing and search within each WSI.

provides a breakdown of the selection, search, and total time for each experiment setup. The exact level of patch and search time reduction depends on the deep network used to obtain the deep features and on the anomaly detection algorithm used in the normal atlas. Figure 8 provides a visual representation of the reduction in the average number of patches and total search time per WSI.

Discussion

The urgent need for novel WSI patching is an important aspect of current works in the field of computational pathology. Patching involves dividing an image into smaller regions or patches and using these patches to represent the tissue in the image. While most of the computational pathology tasks are designed to classify or detect the presence of pathological lesions in gigapixel WSIs, the presence of computationally redundant normal tissue in WSIs is often ignored. While a large body of scientific work has been done in the field of WSI retrieval systems to address the numerical representation of the WSI^{13,24–26} and WSI search engines^{7,8,27,28}, no study has addressed the effect of normal tissue in the indexing and search pipeline. Therefore, removing normal tissue is not an established practice in this field. While some works have used normal tissue exclusion as part of their workflow²⁹, the potential effect of redundant normal tissue in WSI retrieval is an unknown area. In this work, we tested the use of two anomaly detection algorithms as a means of excluding abnormal patches from Yottixel's indexing and search⁷. Our methods included no WSI-level annotation and used multiple instance learning for known normal WSIs.

Our results show that removing the areas with normal histology during the patching process helps with selecting a smaller set of representative patches for each WSI, therefore decreasing the computational and storage cost of the search engine, while maintaining the overall retrieval performance of the search engine.

As shown in Fig. 9, by using a one-class SVM to detect and exclude the normal patches in the cSCC dataset using features from KimiaNet, the classification performance for all three classes was either maintained at the same level or even improved. At the same time, as Table 1 depicts, the number of patches for selected cases was reduced by 67% after excluding the normal tissue, followed by a reduction in the total search time. Contrary

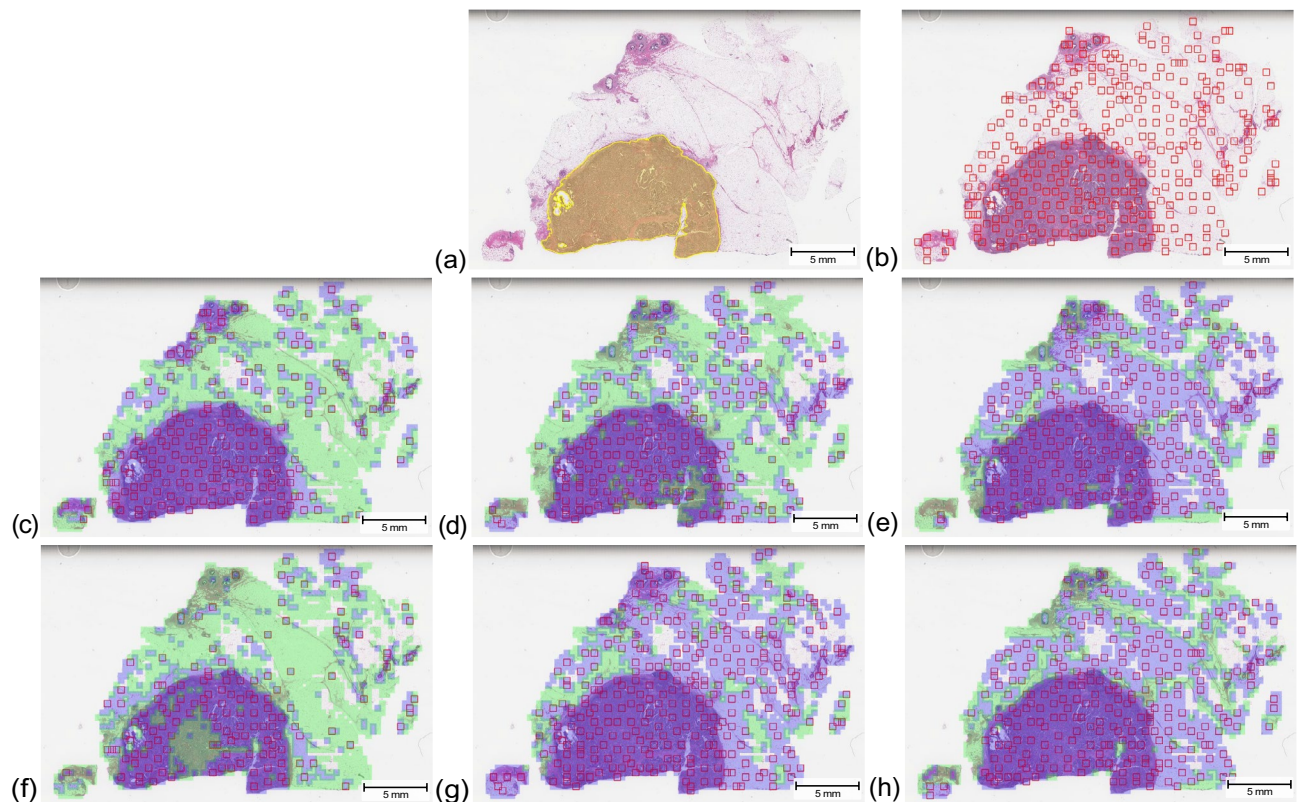


Figure 5. Visual results of a breast carcinoma case: (a) Pathologist ground truth, (b) Yottixel's mosaic, (c) KimiaNet and Isolation Forest, (d) ViT-DINO and Isolation Forest, (e) ResNet50-DINO and Isolation Forest, (f) KimiaNet and one-class SVM, (g) ViT-DINO and one-class SVM, and (h) ResNet50-DINO and one-class SVM. Yellow regions depict the pathologist ground truth annotation for the tumor region. The green and purple areas show patches labeled normal and abnormal by the normal atlas, respectively. Red boxes show the patches selected for indexing and search within each WSI.

to this improvement, the performance for correctly classifying the normal cases decreases, with a decrease in the F1 score from 0.97 to 0.79. This drop can be attributed to the fact that excluding patches that look normal from a normal WSI will leave it with patches mainly containing artifacts and distortions. This shift in the representation patches increases the chances of mismatch when searching. In contrast to the improvement seen for the experiments with deep features obtained by KimiaNet, those done on deep features obtained using DINO-trained ViT and ResNet50 show rather mixed results.

The results for the breast dataset show almost the same pattern as the results for the cSCC dataset. As Fig. 10 shows, the classification performance for ductal carcinoma has shown minor changes regardless of the anomaly detection method selected and the network used to obtain the deep features. However, excluding normal tissue inflicts varying changes to the lobular carcinoma classification performance based on the network and anomaly detection method. Excluding normal tissue from breast WSIs also decreased the number of patches selected by 12% to 47%, followed by a reduction in the total search time as depicted in Table 1. The same effect of decreased performance in normal cases for the cSCC dataset can also be observed in the breast dataset.

The changes in indexing and search performance seen in the skin dataset are also in concordance with the patch-level classification performance seen for annotated validation cases which is shown in Fig. 3. The validation results show that using Isolation Forest and one-class SVM classifiers on deep features obtained from KimiaNet are the best combination with an F1 score of 0.82 and 0.80, respectively. The same configurations also showed the best results in indexing and search as shown in Fig. 9. Applying one-class SVM on DINO-trained ViT and Isolation Forest on DINO-trained ResNet50 also resulted in acceptable performance with an F1 score of 0.79 and 0.52, respectively. These two configurations also showed mediocre performance in indexing and search. Even though these two pairs of combinations showed good results, combining them the other way around does not show promising performance. These figures are in concordance with the visual demonstration of the normal versus abnormal region selection and the pathologist-approved ground-truths shown for two sample cases in

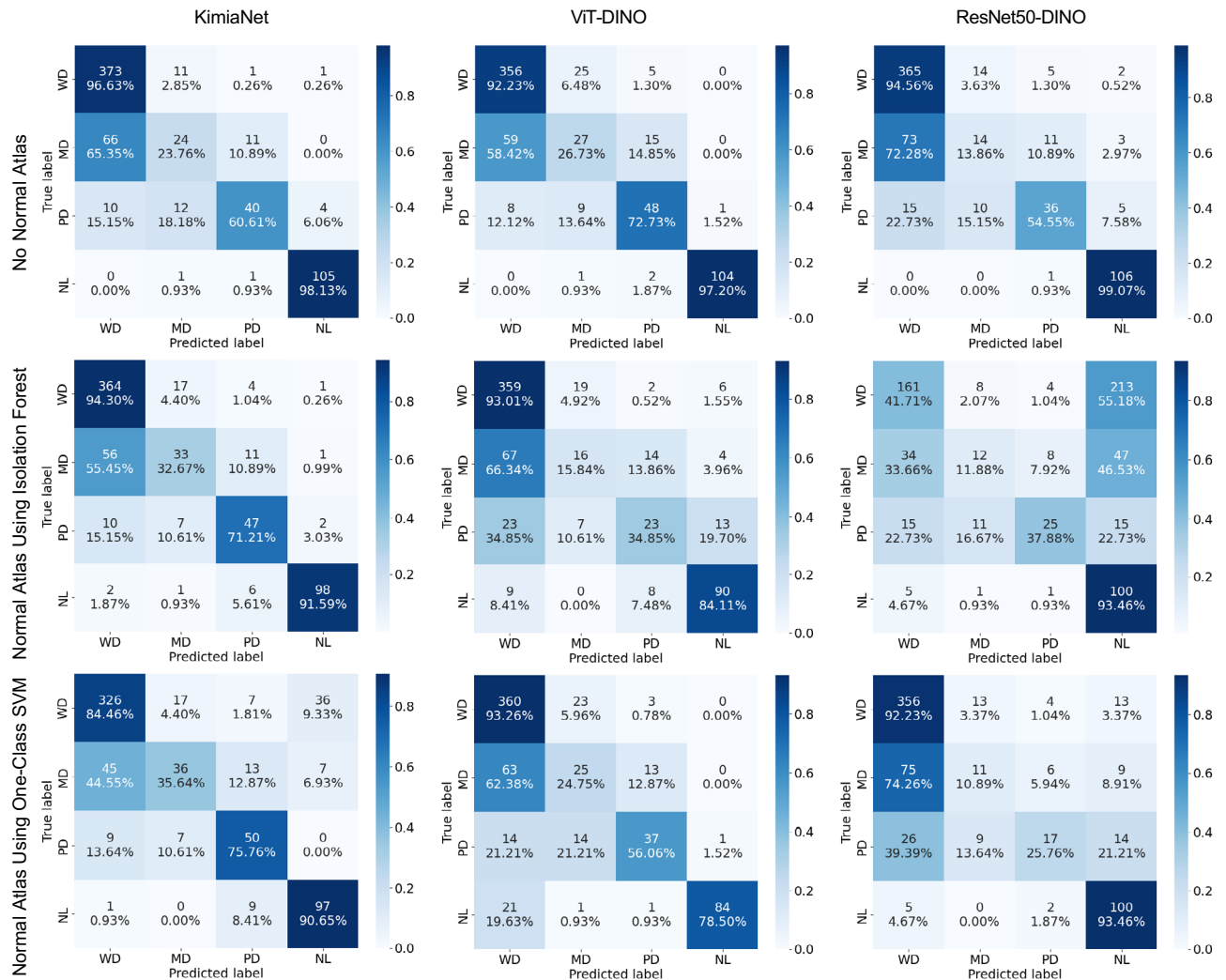


Figure 6. Indexing and searching confusion matrices for leave-one-patient-out validation for top-5 consensus predictions in the skin dataset. Columns represent the indexing and search using deep features from KimiaNet, DINO-trained vision transformer, and DINO-trained ResNet50, respectively. Rows represent results in three experimental setups using no normal atlas, using a normal atlas with Isolation Forest, and using a normal atlas with one-class SVM, respectively.

Figs. 4 and 5. These findings prove that patch-level classification performance is an indicator of how a normal atlas would perform in an indexing and search pipeline. This calls for the need to validate the performance of the normal atlas at the patch level before applying it in the pipeline.

There are some limitations to this study. As previously stated, the indexing and search performance for some cases was suboptimal before applying the normal atlas. Examples of such cases include moderately versus differentiated cases of cSCC and lobular versus ductal carcinoma of the breast. We speculate that this is an active research area involving WSI representation. Improving the performance of WSI search engines and pathology-specific deep networks is not in the scope of this study. We selected the currently established pipeline of Yottixel as the state-of-the-art search engine for our experiments.

Looking toward future research, other studies have shown the role of vision transformers in the area of anomaly detection^{30–32}. These approaches mainly rely on using the reconstruction score of a previously trained vision transformer encoder and decoder to extinguish normal versus abnormal image features. This approach hypothesizes that a coupled encoder and decoder trained only on images of normal phenomena would struggle to reconstruct an image of an ab-normal instance. This approach holds promise for training unsupervised anomaly detection models in medical imaging and should be further investigated in future studies.

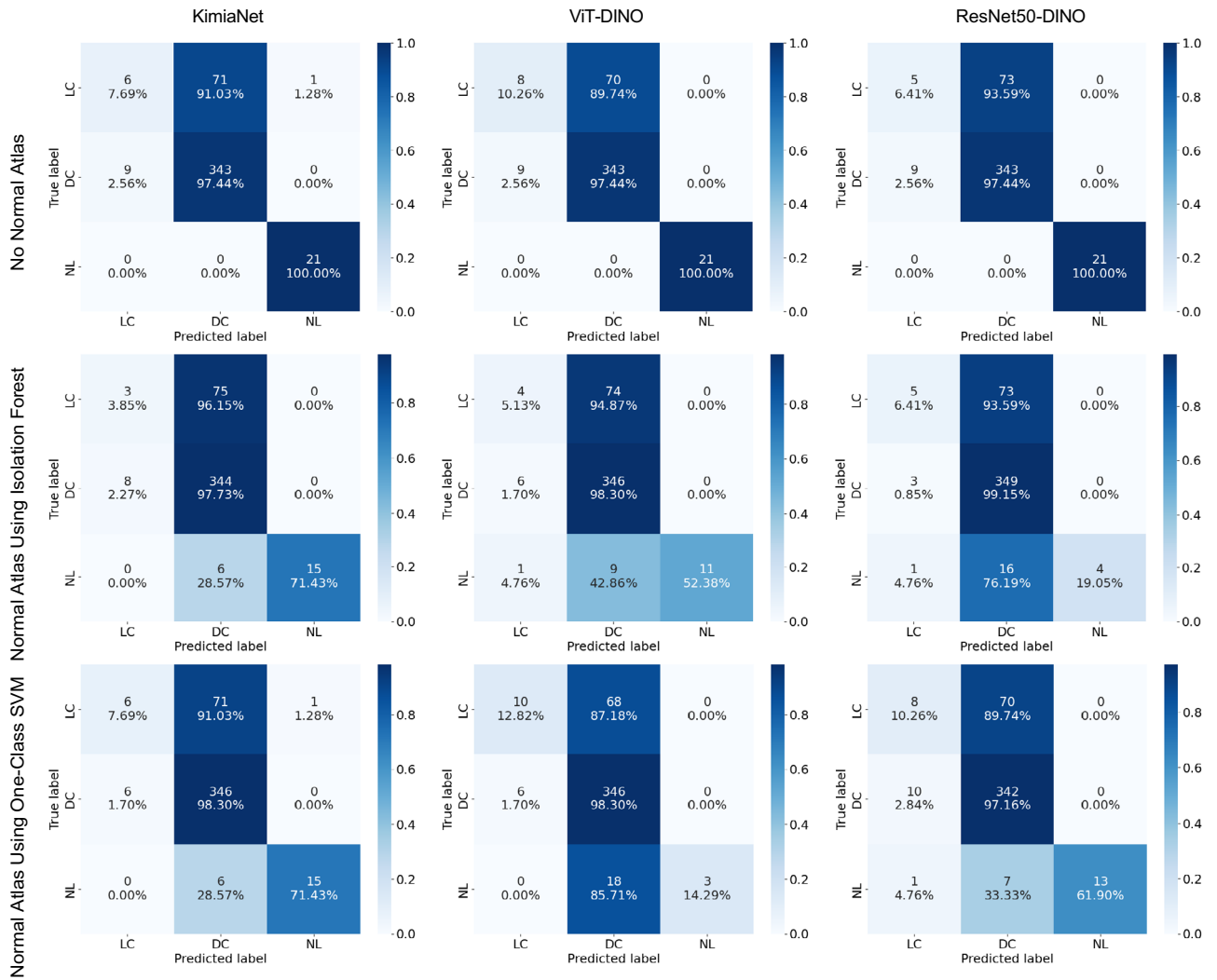


Figure 7. Indexing and searching confusion matrices for leave-one-patient-out validation for top-5 consensus predictions in the breast dataset. Columns represent the indexing and search using deep features from KimiaNet, DINO-trained vision transformer, and DINO-trained ResNet50, respectively. Rows represent results in three experimental setups using no normal atlas, using a normal atlas with Isolation Forest, and using a normal atlas with one-class SVM, respectively.

	Experiment setup	Average per WSI	No normal Atlas	Normal Atlas via Isolation Forest	Normal Atlas via one-class SVM
Skin	KimiaNet	Patches	66.4	33 (↓ 50%)	21.6 (↓ 67%)
		Total time	3112	2853 (↓ 8%)	2502 (↓ 20%)
	ViT DINO	Patches	66.5	11.7 (↓ 82%)	33.8 (↓ 49%)
		Total time	2929	1489 (↓ 49%)	3309 (↑ 13%)
	ResNet50 DINO	Patches	66.4	13.4 (↓ 80%)	9.4 (↓ 86%)
		Total Time	3872	845 (↓ 78%)	2433 (↓ 37%)
Breast	KimiaNet	Patches	248.5	216.5 (↓ 13%)	199.1 (↓ 20%)
		Total time	8957	6995 (↓ 22%)	5840 (↓ 35%)
	ViT DINO	Patches	248.5	132.8 (↓ 47%)	219.3 (↓ 12%)
		Total time	6628	2463 (↓ 63%)	5293 (↓ 20%)
	ResNet50 DINO	Patches	248.4	144.9 (↓ 42%)	134.6 (↓ 41%)
		Total time	17,932	6390 (↓ 64%)	5591 (↓ 69%)

Table 1. Average number of patches and total time per WSI in three different experimental setups including no normal atlas, normal atlas with Isolation Forest, and normal atlas with one-class SVM for skin and breast WSIs. For each setup, the numbers have been reported for deep features obtained from all three networks. The percentage and arrow in the parenthesis indicate the increase or reduction in each experiment compared to the base configuration where no normal atlas was used. All times are reported in milliseconds.

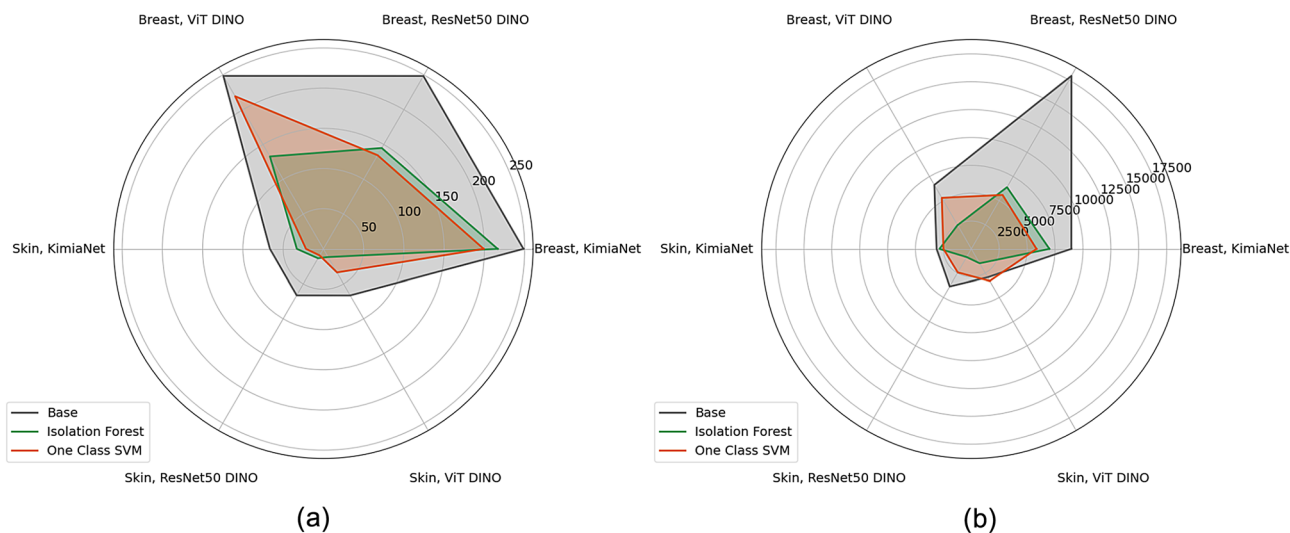


Figure 8. Average number of patches (a) and search time (b) per WSI. Measurements are reported for the base configuration of not using a normal atlas, using Isolation Forest and one-class SVM. For each configuration, the measurements are also categorized based on the dataset and network used.

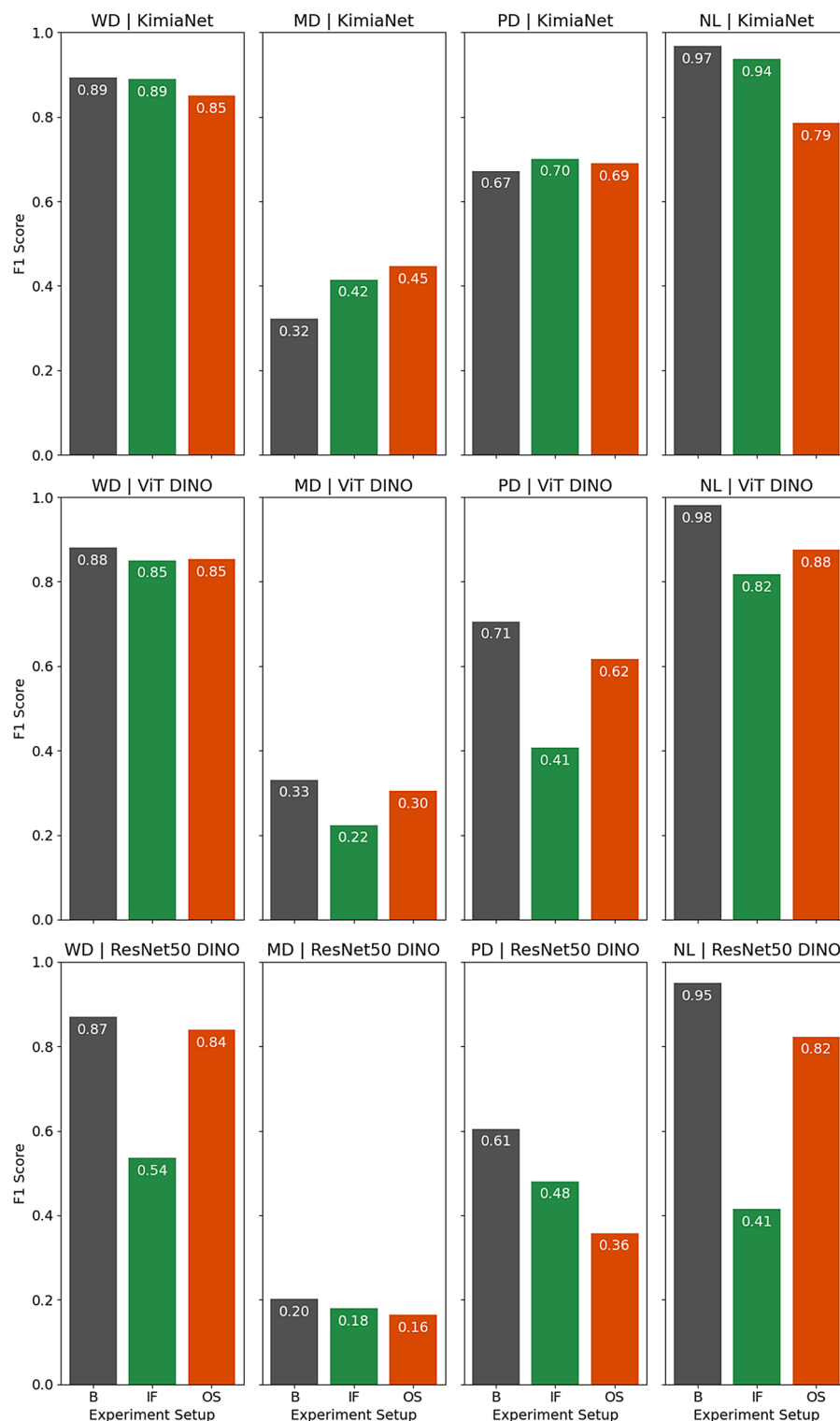


Figure 9. Indexing and search F1 score for leave-one-patient-out validation for top-5 consensus predictions in skin dataset plotted for each class, deep network, and experimental setup separately (*WD* well-differentiated, *MD* moderately differentiated, *PD* poorly differentiated, *NL* normal, *B* base configuration of using no atlas, *IF* using normal atlas based on Isolation Forest, *OS* using normal atlas based on one-class SVM).

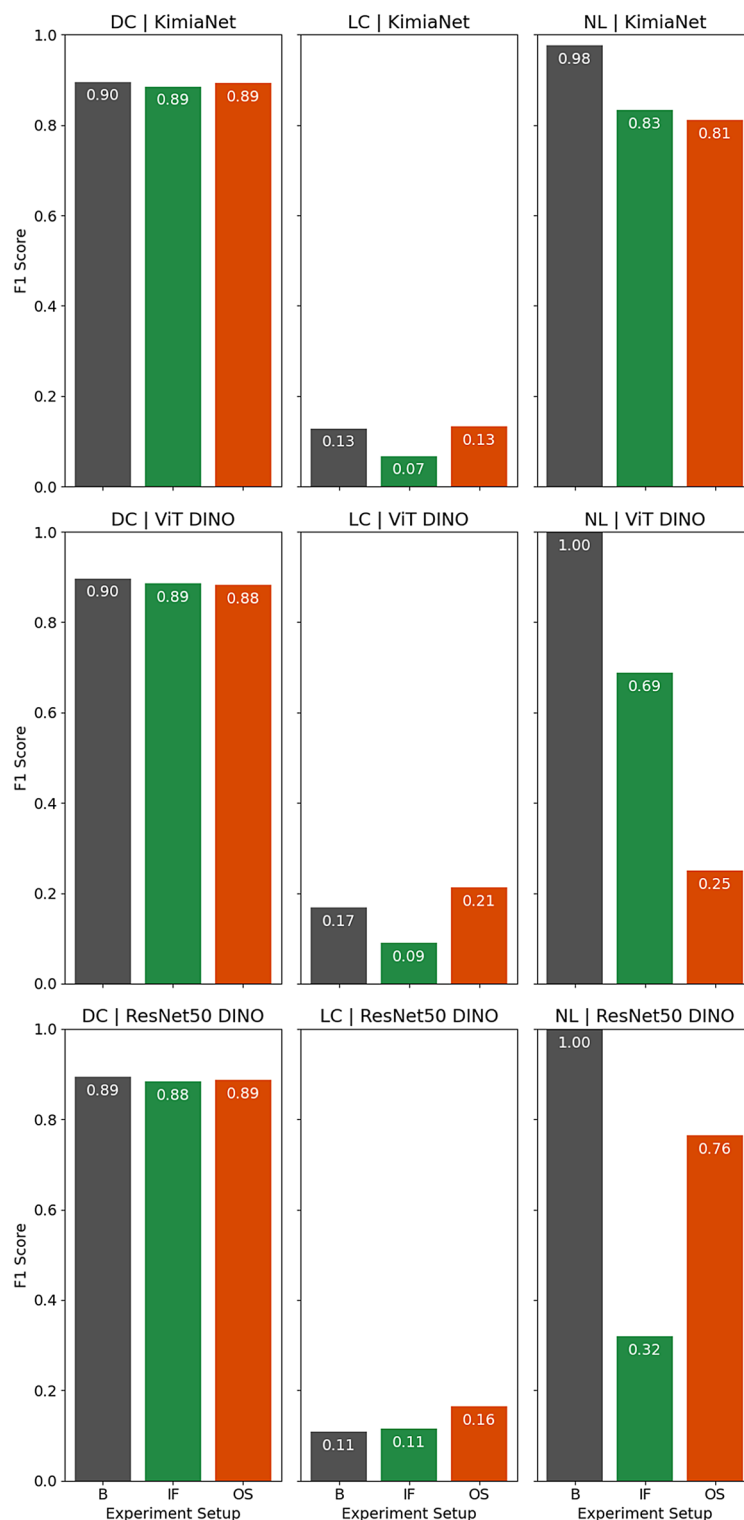


Figure 10. Indexing and search F1 score leave-one-patient-out validation for top-5 consensus predictions in breast dataset plotted for each class, deep network, and experimental setup separately (*DC* infiltrating ductal carcinoma, *LC* lobular carcinoma, *NL* normal, *B* base configuration of using no atlas, *IF* using normal atlas based on Isolation Forest, *OS* using normal atlas based on one-class SVM).

Data availability

The datasets generated during and/or analyzed during the current study are not publicly available due to privacy concerns and ethical considerations. Access to the data can be requested by qualified researchers and institutions for purposes of replication, verification, and further research. Interested parties may request access to the data by contacting the corresponding author at Tizhoosh.Hamid@mayo.edu or the institution at Mayo Clinic, 200 First St. SW Rochester, MN 55905. Requests for data access will be subject to review and approval to ensure compliance with all relevant ethical and legal standards regarding patient data privacy and confidentiality.

Received: 11 October 2023; Accepted: 13 February 2024

Published online: 16 February 2024

References

- Chen, X., Zheng, B. & Liu, H. Optical and digital microscopic imaging techniques and applications in pathology. *Anal. Cell. Pathol.* **34**, 150563. <https://doi.org/10.3233/ACP-2011-0006> (2011).
- Pantanowitz, L. *et al.* Twenty years of digital pathology: An overview of the road travelled, what is on the horizon, and the emergence of vendor-neutral archives. *J. Pathol. Inform.* **9**, 40. https://doi.org/10.4103/jpi.jpi_69_18 (2018).
- Abels, E. *et al.* Computational pathology definitions, best practices, and recommendations for regulatory guidance: A white paper from the digital pathology association. *J. Pathol.* **249**, 286–294 (2019).
- Patel, A. *et al.* Contemporary whole slide imaging devices and their applications within the modern pathology department: A selected hardware review. *J. Pathol. Inf.* **12**, 50. https://doi.org/10.4103/jpi.jpi_66_21 (2021).
- Shrestha, A. & Mahmood, A. Review of deep learning algorithms and architectures. *IEEE Access* **7**, 53040–53065. <https://doi.org/10.1109/ACCESS.2019.2912200> (2019).
- Kalra, S. *et al.* Pan-cancer diagnostic consensus through searching archival histopathology images using artificial intelligence. *NPJ Digit. Med.* <https://doi.org/10.1038/s41746-020-0238-2> (2019).
- Kalra, S. *et al.* Yottixel—An image search engine for large archives of histopathology whole slide images. *Med. Image Anal.* **65**, 101757. <https://doi.org/10.1016/j.media.2020.101757> (2020).
- Gabriele, C. *et al.* Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat. Med.* <https://doi.org/10.1038/s41591-019-0508-1> (2019).
- Shafique, A. *et al.* A preliminary investigation into search and matching for tumor discrimination in world health organization breast taxonomy using deep networks. *Mod. Pathol.* **37**, 100381 (2024).
- Lowe, D. G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **60**, 91–110. <https://doi.org/10.1023/B:VISI.0000029664.99615.94> (2004).
- Babenko, A., Slesarev, A., Chigorin, A. & Lempitsky, V. Neural codes for image retrieval. In *Computer Vision—ECCV 2014* (eds Fleet, D. *et al.*) 584–599 (Springer International Publishing, 2014).
- Alzu'bi, A., Amira, A. & Ramzan, N. Content-based image retrieval with compact deep convolutional features. *Neurocomputing* **249**, 95–105. <https://doi.org/10.1016/j.neucom.2017.03.072> (2017).
- Riasatian, A. *et al.* Fine-tuning and training of densenet for histopathology image representation using tcga diagnostic slides. *Med. Image Anal.* **70**, 102032 (2021).
- Dosovitskiy, A. *et al.* An image is worth 16x16 words: Transformers for image recognition at scale. Preprint at <https://arxiv.org/quant-ph/2010.11929> (2020).
- Caron, M. *et al.* Emerging properties in self-supervised vision transformers. Preprint at <https://arxiv.org/quant-ph/2104.14294> (2021).
- Dimitriou, N., Arandjelovic', O. & Caie, P. D. Deep learning for whole slide image analysis: An overview. *Front. Medicine* **6**, DOI: <https://doi.org/10.3389/fmed.2019.00264> (2019).
- Wang, X. *et al.* Retccl: Clustering-guided contrastive learning for whole-slide image retrieval. *Med. Image Anal.* **83**, 102645 (2023).
- Sikaroudi, M., Afshari, M., Shafique, A., Kalra, S. & Tizhoosh, H. Comments on fast and scalable search of whole-slide images via self-supervised deep learning. Preprint at <https://arxiv.org/quant-ph/2304.08297> (2023).
- Perera, P., Oza, P. & Patel, V. M. One-class classification: A survey. Preprint at ArXiv abs/2101.03064 (2021).
- Liu, F. T., Ting, K. M. & Zhou, Z.-H. Isolation forest. In *2008 Eighth IEEE International Conference on Data Mining* (ed. Liu, F. T.) 413–422 (IEEE, 2008).
- Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J. & Williamson, R. C. Estimating the support of a high-dimensional distribution. *Neural Comput.* **13**, 1443–1471. <https://doi.org/10.1162/089976601750264965> (2001).
- Chang, C. & Lin, C. Training v-support vector classifiers: Theory and algorithms. *Neural Comput.* **13**, 2119–2147. <https://doi.org/10.1162/089976601750399335> (2001).
- Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
- Hemati, S., Kalra, S., Babaie, M. & Tizhoosh, H. R. Learning binary and sparse permutation-invariant representations for fast and memory efficient whole slide image search. *Comput. Biol. Med.* **162**, 107026. <https://doi.org/10.1016/j.compbiomed.2023.107026> (2023).
- Tommasino, C., Merolla, F., Russo, C., Staibano, S. & Rinaldi, A. M. Histopathological image deep feature representation for cbir in smart pacs. *J. Digit. Imaging* **09**, 09. <https://doi.org/10.1007/s10278-023-00832-x> (2023).
- Fashi, P. A., Hemati, S., Babaie, M., Gonzalez, R. & Tizhoosh, H. R. A self-supervised contrastive learning approach for whole slide image representation in digital pathology. *J. Pathol. Inform.* **13**, 100133. <https://doi.org/10.1016/j.jpi.2022.100133> (2022).
- Hu, D. *et al.* Informative retrieval framework for histopathology whole slides images based on deep hashing network. *2020 IEEE 17th Int. Symp. Biomed. Imaging (ISBI)* 244–248 (2020).
- Sikaroudi, M., Afshari, M., Shafique, A., Kalra, S. & Tizhoosh, H. Comments on fast and scalable search of whole-slide images via self-supervised deep learning. Preprint at <https://arxiv.org/quant-ph/2304.08297> (2023).
- Jaber, M. I. *et al.* Automated adeno/squamous-cell nslcl classification from diagnostic slide images: A deep-learning framework utilizing cell-density maps. *Cancer Res. Conf. Am. Assoc. for Cancer Res. Annu. Meet.* **79**, DOI: <https://doi.org/10.1158/1538-7445.2019>.
- Mishra, P., Verk, R., Fornasier, D., Piciarelli, C. & Foresti, G. L. Vt-adl: A vision transformer network for image anomaly detection and localization. In: *2021 IEEE 30th International Symposium on Industrial Electronics (ISIE)*, 01–06, DOI: <https://doi.org/10.1109/ISIE45552.2021.9576231> (2021).
- Pirnay, J. & Chai, K. Inpainting transformer for anomaly detection. In *Image Analysis and Processing—ICIAP 2022* (eds Sclaroff, S. *et al.*) 394–406 (Springer International Publishing, 2022).
- Lee, Y. & Kang, P. Anovit: Unsupervised anomaly detection and localization with vision transformer-based encoder-decoder. *IEEE Access* **10**, 46717–46724. <https://doi.org/10.1109/ACCESS.2022.3171559> (2022).

Acknowledgements

Some figures were created with BioRender.com.

Author contributions

P.N. implemented and validated the approach, contributed to the algorithmic design, and wrote the manuscript. A.A. and G.A. contributed to the experiments. A.R.M., N.I.C., D.H.M., P.Z., S.Y., and J.J.G. collected and analyzed data and provided guidance. H.R.T. conceptually designed the framework, provided algorithmic guidance, analyzed results, and contributed to writing the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-54489-9>.

Correspondence and requests for materials should be addressed to H.R.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024