Domain Adaptation for Unsupervised Cancer Detection: An application for skin Whole Slides Images from an interhospital dataset

Natalia P. García-de-la-Puente*[0009−0009−9704−9102], Miguel López-Pérez*[0000−0003−2965−0624], Laëtitia Launet[0000−0003−4230−0987], and Valery Naranjo[0000−0002−0181−3412]

Instituto Universitario de Investigación en Tecnología Centrada en el Ser Humano, Universitat Politècnica de València, Spain
{napegar,mlopper3,lmlaunet,vnaranjo}@upv.es

Abstract. Skin cancer diagnosis relies on assessing the histopathological appearance of skin cells and the patterns of epithelial skin tissue architecture. Despite recent advancements in deep learning for automating skin cancer detection, two main challenges persist for their clinical deployment. (1) Deep learning models only recognize the classes trained on, giving arbitrary predictions for rare or unknown diseases. (2) The generalization across healthcare institutions, as variations arising from diverse scanners and staining procedures, increase the task complexity. We propose a novel Domain Adaptation method for Unsupervised cancer Detection (DAUD) using whole slide images to address these concerns. Our method consists of an autoencoder-based model with stochastic latent variables that reflect each institution's features. We have validated DAUD in a real-world dataset from two different hospitals. In addition, we utilized an external dataset to evaluate the capability for out-of-distribution detection. DAUD demonstrates comparable or superior performance to the state-of-the-art methods for anomaly detection. https://github.com/cvblab/DAUD-MICCAI2024

Keywords: Unsupervised detection · Skin Cancer · Histopathology.

1 Introduction

Skin cancer is one of the most prevalent cancers worldwide, with approximately one in three cancer cases diagnosed globally [4]. Despite the high mortality risk associated with skin cancer, early detection has the potential of elevating its survival rate to 99% [25]. However, the current gold standard for its detection, histopathological image analysis, is particularly error-prone and time-consuming. This burden can be eased while improving the accuracy with computer-aided diagnosis (CAD) methods, which typically leverage deep learning models to analyze histology via whole slide images (WSIs). The adoption of such methods

* The first two authors contributed equally.

for classifying WSIs has increased recently due to their outstanding capabilities for detecting several cancer [18,27,10].

While deep learning methods have achieved notable results through continuous improvement on specific tasks, there are still some common drawbacks to achieving reliable and useful CAD methods [13]. First, a CAD method will only recognize the patterns of the subset utilized to train the model. For example, most studies have focused on specific types of skin cancer, e.g., spitzoid tumors [5], melanoma [14], or spindle cell cutaneous neoplasms [3]. Under this scenario, other pathologies, which suppose a concept shift, are regarded as erroneous predictions. This is known in the literature as out-of-distribution (OoD) samples and is frequently addressed using supervised methods [16], which is unfeasible where malign samples are scarce and difficult to obtain. Second, utilized datasets for skin cancer detection may lack representative characteristics observed in daily practice [1]. WSIs may suffer from different sources of variation (also called, a covariate shift), such as those derived from different scanners or staining techniques [11]. These differences are even higher when attempting to generalize to different hospitals or institutions with different protocols. Some studies aimed to diminish the color variations across centers. However, distinctive centers' features remain [26,21]. Figure 1 depicts these two issues.

Unsupervised anomaly detection has emerged due to its promising results for detecting a concept shift [2,7]. In histopathology, anomaly detection has been mainly applied for detecting anomalous cells or patches [30,17]. In this study, we tailored this framework for WSI malignancy detection when considering benign samples as normal and malign or unseen samples as anomalies upon which pathologists would focus in the clinic. In this context, domain adaptation becomes crucial to reduce the covariate shift because it allows training models with unseen data in an external or target domain by acquiring knowledge from a related source domain [22]. Thus, their integration into cancer detection makes them a solution for the joint use of interhospital datasets while applying to different diseases. Despite the promise of domain adaptation for anomaly detection [28], it is a novel research area yet to be explored in histopathology.

To address the abovementioned issues, we propose a novel Domain Adaptation method for Unsupervised cancer Detection (DAUD) of WSIs of malignant and unknown samples. DAUD's main characteristic is employing an autoencoder to learn benign WSIs (i.e., detecting benign observed samples independent of the hospital/institution). Due to its formulation, it detects as an anomaly, malignant and OoD WSIs. We hypothesize that both suppose a concept shift, so the reconstruction error must be higher. This formulation can ensure that non-benign samples (malign and unknown regardless of being benign or malignant) are raised by the CAD system for review by the pathologist. This way, we ensure which samples are benign and do not require further doctor review. Furthermore, DAUD reduces the covariate shift between centers by encoding the sources of variations of the WSIs. Specifically, we utilize stochastic latent variables for this purpose. These variables represent the inter-hospital differences. As being represented with random variables, they can capture the intra-variability typical

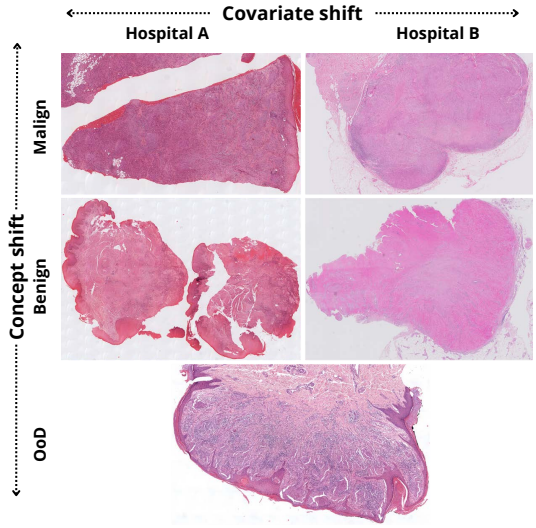

Fig. 1. Visualization of Hosp. A and B samples. Malignant leiomyosarcoma (first row), benign leiomyoma (second row), and OoD spitzoid (third row).

of hospitals. This model is a relevant step toward generalizing cancer detection across institutions. Our main contributions are as follows.

1. A novel domain adaptation method for unsupervised detection was developed called DAUD, which is robust to covariate shifts, while detecting the concept shift.
2. We conducted experiments in a real-world cutaneous spindle cell (CSC) neoplasm dataset. These prevalent skin tumors are one of the most challenging for expert pathologists. We considered CSC neoplasms from two different hospitals, where the covariate shift is large.
3. We also validated the proposed method in an external dataset of spitzoid tumors to assess the generalization capability for OoD samples.

2 Methods

The proposed method (Figure 2) consists of a self-supervised feature extraction step and a domain adaptation unsupervised cancer detection model, called DAUD. We divided this pipeline into two independent stages to obtain more representative features independent of the domain.

2.1 Problem formulation

Given a histopathological dataset from a hospital i composed of benign tumoral biopsies, we extracted a d-dimensional embedding from each WSI (as described

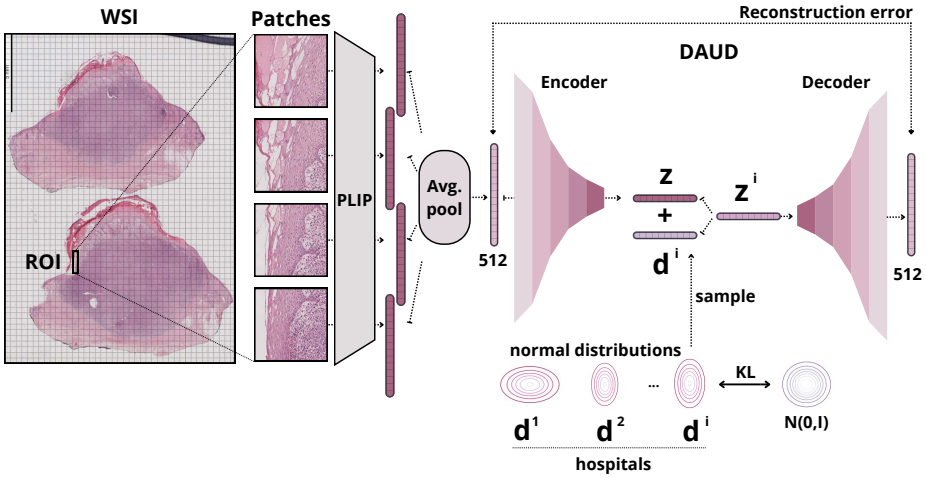

Fig. 2. Proposed framework. First, we obtain a global descriptor of the Whole-Slide Image (WSI) through Region of Interest (ROI) detection, PLIP feature extractor, and average pooling. Then, we applied our DAUD model, which disentangles in the latent space the concept features $\mathbf{z}$ and the institution features $\mathbf{d}$.

in Section 2.2). Then, we denoted the feature matrix as $\mathbf{X}^i \in \mathbb{R}^{N_i \times d}$, where N_i is the total number of WSIs in that hospital. In this context, our objective was to, for a given number of different hospitals I, learn the distribution of benign samples of each hospital for a specific type of tumor and learn a function $f^i(\cdot)$ that can predict if a new sample $\mathbf{x}_*^i$ is an anomaly. We considered anomalies to include malign biopsies and other kinds of tumors, regardless of the benignity (i.e., any class unobserved by the model). All of these cases are considered a concept shift since they represent new classes not observed during training.

2.2 DAUD: Domain Adaptation for Unsupervised Detection

Feature extraction. We used PLIP [9] as the feature extractor, with an embedding dimension equal to 512. We applied it to each patch and then averaged the features to obtain a global descriptor for each WSI.

Architecture. The main idea is to learn the benign features of our training dataset across the I institutions with the autoencoder. We designed an autoencoder structure using f_ϕ and f_θ as the *encoder* (parameterized by ϕ) and the *decoder* (parameterized by θ), respectively. First, we obtained a latent representation of the input $\mathbf{x}$ using the deterministic transformation $f_\phi(\mathbf{x}) = \mathbf{z}$. Then, we introduced a stochastic variable in the latent space that encodes each hospital's information $\mathbf{d}$. This variable follows a multivariate Gaussian distribution given the specific hospital $\mathbf{d}|i \sim \mathcal{N}(\mu^i, \mathbf{\Sigma}^i)$. We assumed that $\mathbf{\Sigma}^i$ is diagonal. We also impose a prior centered isotropic multivariate Gaussian distribution on the

domain variables, $p(\mathbf{d}) = \mathcal{N}(\mathbf{0}, \mathbf{I})$. Note that the distribution $p(\mathbf{d}|i)$ encodes the variability and intrinsic properties of the domain, while $\mathbf{z}$ retains the concept features of the sample. We define $\mathbf{z}^i$ as the random variable that encodes both sources of information, and we can approximate this distribution by a Gaussian distribution given by $\mathbf{z}^i|\mathbf{d}, \mathbf{z}, i \sim \mathcal{N}(\mathbf{z}^i|\mathbf{z} + \mu^i, \mathbf{\Sigma}^i)$. Finally, we aimed to decode this compressed representation given the institution. The probabilistic decoder gives the following observation model $p_\theta(\mathbf{x}|\mathbf{z}^i) = \mathcal{N}(\mathbf{x}|f_\theta(\mathbf{z}^i), \mathbf{I})$. Since the posterior of $\mathbf{z}^i$ is intractable, we utilize variational inference to compute an approximate posterior $q_\phi(\cdot)$.

Training scheme. The evidence lower bound (ELBO) of our model for a data point $\mathbf{x}$ from institution i is given by:

$$\text{ELBO}(\theta, \phi, \mu^i, \mathbf{\Sigma}^i; \mathbf{x}, i) = \mathbb{E}_{q_\phi(\mathbf{z}^i|\mathbf{x})}\left[\log p_\theta(\mathbf{x}|\mathbf{z}^i)\right] - \text{KL}(q_\phi(\mathbf{d}|i)||p(\mathbf{d})). \quad (1)$$

where $\text{KL}(q_\phi(\mathbf{d}|i)||p(\mathbf{d}))$ is the Kullback-Leibler (KL) divergence between the approximate posterior $q_\phi(\mathbf{d}|i)$ and the prior $p(\mathbf{d})$. Note that the approximate posterior $q_\phi(\mathbf{z}^i|\mathbf{d}, \mathbf{z}, i)$ is obtained from the the encoder and the approximate posterior $q_\phi(\mathbf{d}|i)$. The first term in Eq. 1 corresponds to the reconstruction error (i.e., mean square error). The second term in Eq. 1 uses KL divergence to encourage fidelity to the prior distribution. This term can be regarded as a regularizer. The training objective is to maximize the ELBO given the observational data, and the model can be optimized with respect to the variational ϕ, μ^i, $\mathbf{\Sigma}^i$ and generative parameters θ.

Reparametrization trick. To obtain samples from the distribution $q_\phi(\mathbf{z}^i|\mathbf{d}, \mathbf{z}, i)$, we utilized the reparametrization trick. We sampled from the posterior using $\mathbf{z}^{(i,l)} = g_\phi(\mathbf{x}, \epsilon^l) = \mathbf{z} + \mu^i + \sigma^i * \epsilon^l$, where $\epsilon^l \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. With *, we denote the element-wise product. We can compute an estimator of the ELBO using L Monte Carlo samples:

$$\mathbb{E}_{q_\phi(\mathbf{z}^i|\mathbf{x}_*)}\left[\log p_\theta(\mathbf{x}|\mathbf{z}^i)\right] \approx \frac{1}{L}\sum_{l=1}^{L} \log p_\theta(\mathbf{x}_*|\mathbf{z}_*{}^{(i,l)}). \quad (2)$$

Prediction and anomaly score. For a new test sample $\mathbf{x}_*$ from institution i, we estimate the likelihood of this sample with the Monte Carlo estimation given by Eq. 2. If the institution i is unobserved during the training, we assume that $q_\phi(\mathbf{d}|i) = p(\mathbf{d})$. We sampled from the prior of $\mathbf{d}$ because we did not learned a posterior for that institution. Once the likelihood was estimated, we use this value as the anomalous score to distinguish anomalous/OoD samples, this can also be thought of as the mean square error.

3 Experiments and results

3.1 Experimental settings

AI4SkIN dataset. This is an in-house dataset consisting of 239 slides from the Hospital Universitario San Cecilio de Granada (called Hosp. A hereafter) and 369

Table 1. Average AUC values for 10 independent runs. We also report the 0.95-CI interval. We depict the values for testing benign vs malignant (anomaly) for Hosp. A and B, and benign vs unknown (OoD). We report the average AUCs.

	Hosp. A	Hosp. B	OoD dataset
VAE [12]	0.6575 ± 0.0014	0.8199 ± 0.0038	0.6591 ± 0.0005
β-VAE [8]	$\mathbf{0.6630 \pm 0.0046}$	0.8395 ± 0.0136	0.6627 ± 0.0064
AnoGAN [24]	0.6490 ± 0.0467	0.7753 ± 0.0466	0.5205 ± 0.0889
DeepSVDD [23]	0.6226 ± 0.0338	0.7931 ± 0.0301	0.6533 ± 0.0329
ECOD [15]	0.6033 ± 0.0	0.7512 ± 0.0	0.5710 ± 0.0
DAUD (ours)	0.6421 ± 0.0147	$\mathbf{0.8974 \pm 0.0063}$	$\mathbf{0.8199 \pm 0.0070}$

slides from the Hospital Clínico Universitario de Valencia (Hosp. B hereafter). This dataset includes seven types of cutaneous spindle cell neoplasms. We utilized the same preparations steps and details as in [6]. Each institution scanned H&E WSIs at a magnification of 40× using a standardized protocol and a single scanner, a Philips Ultra-Fast scanner (for Hosp. A), and a Roche Ventana iScan HT scanner (for Hosp. B). Then, they were downsampled to 10× and cropped into overlapping patches of 512×512. We used an ROI detector to extract the relevant patches from each WSI [3]. We utilized 70/30 for the train/test split, and ensured a balanced set for testing. The dataset was normalized using the Macenko method [19].

SOPHIE dataset. This is a publicly available dataset [20], which comprises 61 WSIs of skin spitzoid tumors from the Department of Anatomical Pathology of the Hospital Clínico Universitario de Valencia. Although this dataset also originates from Hosp. B, it may have different characteristics, and a covariate shift with respect to the AI4SkIN dataset as assumed. The slides were classified into benign and malignant. However, we utilized all of them as OoD samples for external validation in this study. Notably, spitzoid tumors are not fusiform as CSC neoplasms, thus supposing a concept shift despite both originating from the skin. The samples were processed utilizing the same procedure.

Metrics. We report the area under the curve (AUC) for each method. We ran all experiments ten times and reported the average and the 95% CI.

Implementation details. We implemented our proposed models in PyTorch. We employed the Adam optimizer with a learning rate of 10^{-3}, weight decay of 10^{-5} and a mini-batch size of 32. We trained our models during 200 epochs. We utilized the popular library PyOD for running state-of-the-art methods of anomaly detection [29].

3.2 Performance evaluation

Malignant detection. We compared our model with five state-of-the-art anomaly detection methods to evaluate its performance. We validated these methods with the in-house AI4SkIN dataset to detect malignant WSIs. These methods are unsupervised and learn to detect malignant samples from a benign training set.

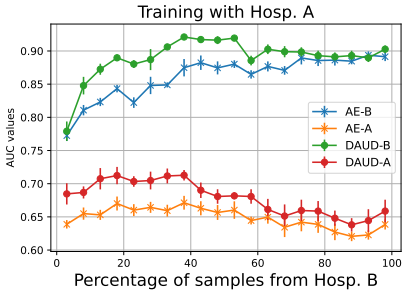

Fig. 3. Training with Hosp. A and adding samples from Hosp. B to the training set.

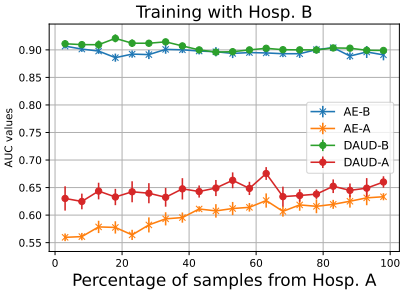

Fig. 4. Training with Hosp. B and adding samples from Hosp. A to the training set.

All the methods utilize the feature vector extracted with PLIP. Table 1 shows the quantitative results of this comparison, demonstrating that DAUD outperformed the rest when evaluated in Hosp. B, and was competitive when evaluated in Hosp. A. Reconstruction methods based on autoencoders, such as VAE, perform better than other widely used methods, such as ECOD, DeepSVDD, and AnoGAN.

Out-of-Distribution detection. The models were trained with benign samples of CSC neoplasms and then tested with benign and malignant CSC samples and the samples of the external SOPHIE dataset with spitzoid tumors . The goal was to distinguish between in-distribution and out-of-distribution samples. In this case, DAUD outperformed all the methods, as shown in Table 1.

3.3 Ablation study

We evaluated the role of the latent variable **d** in our domain adaptation model. We used an autoencoder (w/o domain latent variables) with the same architecture as our proposed model. For a more insightful experiment, we tested the generalization capability of the models for the other hospital. Figure 3 depicts these results for training with Hosp. A with a different percentage of samples added from Hosp. B (x-axis). Figure 4 shows the results for Hosp. B. DAUD consistently outperformed the autoencoder across different percentages of external data added to the training set. This advantage is more significant when we only have access to a small amount of data from the other institution in the training set. DAUD is more robust to the covariate shift. The relative improvement achieved nearly a 6% relative increase from 0.85 to 0.9 for Hosp. A (see Figure 3) and a 16% relative increase from 0.55 to 0.64 for Hosp. B (see Figure 4).

3.4 Visualizations

We gained further insights into our model by visualizing the latent variables. Figure 5 shows that the posterior distribution of **d** for the two first dimensions

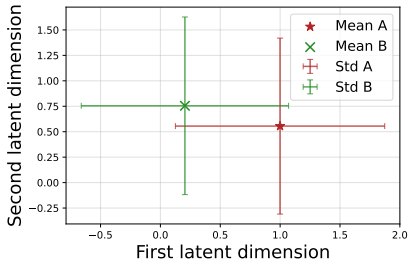

Fig. 5. Visualization of the latent domain variables for Hosp. A and B.

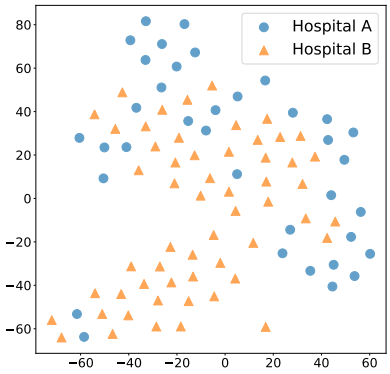

Fig. 6. AE (w/o domain latent variables).

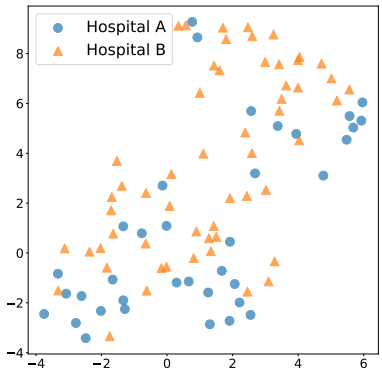

Fig. 7. DAUD latent space.

is shifted, reflecting the difference between hospitals. Furthermore, Figure 6 and 7 show the learned latent representations of the autoencoder and the DAUD with a t-SNE. We show that the distribution across the center is more separated in the autoencoder, while DAUD can make both distributions much closer.

4 Limitations

In this study, we did not distinguish between malignant and OoD but grouped both concepts as non-benign, as a group that needs review by a doctor. In the future, a more sophisticated anomaly score should be adopted to make the clinical application more efficient. Clinical knowledge can be introduced to measure the magnitude of the concept shift. The main objective has been to provide a model that can discriminate benign WSI so doctors can focus on malign and dubious cases. DAUD is an initial step towards reliable CAD systems for skin cancer detection, which also alleviates the burden on pathologists.

5 Conclusion and future work

In this paper, we proposed a novel domain adaption method called DAUD for unsupervised detection of malignant and unknown WSIs, demonstrating similar or superior performance to state-of-the-art anomaly detection methods. Additionally, using latent representations of each center enabled a better generalization with few samples from the other institution. Our method also outperformed the competitors for OoD detection in the external dataset. Future work should focus on further validation with more varied datasets and exploring more heterogeneous diseases.

Acknowledgments. The work of N. P. García de la Puente was supported by the grant PID2022-140189OB-C21 funded by MICIU/AEI/10.13039/ 501100011033 ERDF/UE and FSE+. The work of M. López-Pérez was supported by the grant JDC2022-048318-I funded by MICIU/AEI/10.13039/501100011033 and the "European Union NextGenerationEU/PRTR". This work was also supported by the project PID2022-140189OB-C21 (ASSIST) funded by MICIU/AEI/10.13039/501100011033 and by "FEDER, EU".

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Abels, E., Pantanowitz, L., Aeffner, F., Zarella, M.D., van der Laak, J., Bui, M.M., Vemuri, V.N., Parwani, A.V., Gibbs, J., Agosto-Arroyo, E., et al.: Computational pathology definitions, best practices, and recommendations for regulatory guidance: a white paper from the digital pathology association. The Journal of pathology **249**(3), 286–294 (2019)
2. Ahmad, S., Lavin, A., Purdy, S., Agha, Z.: Unsupervised real-time anomaly detection for streaming data. Neurocomputing **262**, 134–147 (2017)
3. del Amor, R., Colomer, A., Morales, S., Pulgarín-Ospina, C., Terradez, L., Aneiros-Fernandez, J., Naranjo, V.: A self-contrastive learning framework for skin cancer detection using histological images. In: 2022 IEEE International Conference on Image Processing (ICIP). pp. 2291–2295 (2022)
4. Apalla, Z., Lallas, A., Sotiriou, E., Lazaridou, E., Ioannides, D.: Epidemiological trends in skin cancer. Dermatology practical & conceptual **7**(2), 1–6 (2017)
5. Del Amor, R., Launet, L., Colomer, A., Moscardó, A., Mosquera-Zamudio, A., Monteagudo, C., Naranjo, V.: An attention-based weakly supervised framework for spitzoid melanocytic lesion diagnosis in whole slide images. Artificial Intelligence in Medicine **121**, 102197 (2021)
6. Del Amor, R., Pérez-Cano, J., López-Pérez, M., Terradez, L., Aneiros-Fernandez, J., Morales, S., Mateos, J., Molina, R., Naranjo, V.: Annotation protocol and crowdsourcing multiple instance learning classification of skin histological images: The CR-AI4SkIN dataset. Artificial Intelligence in Medicine **145**, 102686 (2023)
7. Han, D., Wang, Z., Chen, W., Wang, K., Yu, R., Wang, S., Zhang, H., Wang, Z., Jin, M., Yang, J., et al.: Anomaly detection in the open world: Normality shift detection, explanation, and adaptation. In: 30th Annual Network and Distributed System Security Symposium (NDSS) (2023)

8. Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., Mohamed, S., Lerchner, A.: beta-vae: Learning basic visual concepts with a constrained variational framework. In: International Conference on Learning Representations (ICLR) (2016)
9. Huang, Z., Bianchi, F., Yuksekgonul, M., Montine, T.J., Zou, J.: A visual–language foundation model for pathology image analysis using medical twitter. Nature Medicine pp. 1–10 (2023)
10. Jaber, M.I., Song, B., Taylor, C., Vaske, C.J., Benz, S.C., Rabizadeh, S., Soon-Shiong, P., Szeto, C.W.: A deep learning image-based intrinsic molecular subtype classifier of breast tumors reveals tumor heterogeneity that may affect survival. Breast Cancer Research **22**, 1–10 (2020)
11. Kanwal, N., Pérez-Bueno, F., Schmidt, A., Engan, K., Molina, R.: The devil is in the details: Whole slide image acquisition and processing for artifacts detection, color variation, and data augmentation: A review. IEEE Access **10**, 58821–58844 (2022)
12. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. In: International Conference on Learning Representations (ICLR) (2014)
13. Van der Laak, J., Litjens, G., Ciompi, F.: Deep learning in histopathology: the path to the clinic. Nature medicine **27**(5), 775–784 (2021)
14. Li, M., Abe, M., Nakano, S., Tsuneki, M.: Deep learning approach to classify cutaneous melanoma in a whole slide image. Cancers **15**(6), 1907 (2023)
15. Li, Z., Zhao, Y., Hu, X., Botta, N., Ionescu, C., Chen, G.: Ecod: Unsupervised outlier detection using empirical cumulative distribution functions. IEEE Transactions on Knowledge and Data Engineering (2022)
16. Linmans, J., Elfwing, S., van der Laak, J., Litjens, G.: Predictive uncertainty estimation for out-of-distribution detection in digital pathology. Medical Image Analysis **83**, 102655 (2023)
17. Linmans, J., Raya, G., van der Laak, J., Litjens, G.: Diffusion models for out-of-distribution detection in digital pathology. Medical Image Analysis **93**, 103088 (2024)
18. Lu, M.Y., Williamson, D.F., Chen, T.Y., Chen, R.J., Barbieri, M., Mahmood, F.: Data-efficient and weakly supervised computational pathology on whole-slide images. Nature biomedical engineering **5**(6), 555–570 (2021)
19. Macenko, M., Niethammer, M., Borland, D., Woosley, J.T., Guan, X., Schmitt, C., Thomas, N.E.: A method for normalizing histology slides for quantitative analysis. In: 2009 IEEE international symposium on biomedical imaging: from nano to macro (ISBI). pp. 1107–1110 (2009)
20. Mosquera-Zamudio, A., Launet, L., Del Amor, R., Moscardó, A., Colomer, A., Naranjo, V., Monteagudo, C.: A spitzoid tumor dataset with clinical metadata and whole slide images for deep learning models. Scientific Data **10**(1), 704 (2023)
21. Pérez-Bueno, F., Serra, J.G., Vega, M., Mateos, J., Molina, R., Katsaggelos, A.K.: Bayesian k-svd for h and e blind color deconvolution. applications to stain normalization, data augmentation and cancer classification. Computerized Medical Imaging and Graphics **97**, 102048 (2022)
22. Redko, I., Morvant, E., Habrard, A., Sebban, M., Bennani, Y.: Advances in domain adaptation theory. Elsevier (2019)
23. Ruff, L., Vandermeulen, R., Goernitz, N., Deecke, L., Siddiqui, S.A., Binder, A., Müller, E., Kloft, M.: Deep one-class classification. In: International Conference on Machine Learning (ICML). pp. 4393–4402. PMLR (2018)

24. Schlegl, T., Seeböck, P., Waldstein, S.M., Schmidt-Erfurth, U., Langs, G.: Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In: International conference on information processing in medical imaging. pp. 146–157. Springer (2017)
25. Shah, M.: Lrnet: Skin cancer classification using low-resolution images. In: 2021 International Conference on Communication information and Computing Technology. pp. 1–5 (2021)
26. Tellez, D., Litjens, G., Bándi, P., Bulten, W., Bokhorst, J.M., Ciompi, F., Van Der Laak, J.: Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. Medical Image Analysis **58**, 101544 (2019)
27. Xiang, T., Song, Y., Zhang, C., Liu, D., Chen, M., Zhang, F., Huang, H., O'Donnell, L., Cai, W.: Dsnet: A dual-stream framework for weakly-supervised gigapixel pathology image analysis. IEEE Transactions on Medical Imaging **41**(8), 2180–2190 (2022)
28. Yang, Z., Soltani, I., Darve, E.: Anomaly detection with domain adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2957–2966 (2023)
29. Zhao, Y., Nasrullah, Z., Li, Z.: Pyod: A python toolbox for scalable outlier detection. Journal of Machine Learning Research **20**(96), 1–7 (2019)
30. Zingman, I., Stierstorfer, B., Lempp, C., Heinemann, F.: Learning image representations for anomaly detection: application to discovery of histological alterations in drug development. Medical Image Analysis **92**, 103067 (2024)