Semi-Supervised Nuclear Pleomorphism Analysis from Sparse Annotations

Šimon Ukuš*
Supervised by: Vanda Benešová†

Faculty of Informatics and Information Technologies
Slovak University of Technology
Bratislava, Slovakia

Abstract

Breast cancer grading via the Nottingham Grading System relies on the evaluation of nuclear pleomorphism, but automating this process is often hindered by real-world data constraints. In many practical clinical settings, detailed pixel-level annotation of nuclei is expensive and time-consuming. As a result, annotations may be sparse, imprecise, or limited to coarse geometric markers such as ellipses, leading to missing labels and weak supervision. This paper proposes a segmentation-driven framework to overcome these challenges. We first utilize a patch-based segmentation approach leveraging Cellpose-SAM, combined with an Intersection over Union (IoU) stitching strategy. This pipeline extracts precise nuclei instances from Whole Slide Images. To align these segmentation masks with imprecise expert annotations, we introduce a centroid-based label matching algorithm that mitigates localization errors. Finally, to resolve the general issue of missing labels within the regions, we introduce a semi-supervised classification framework. A Siamese neural network trained with triplet loss learns a discriminative embedding space from the matched nuclei, leveraging domain-specific histopathology pretraining. Unlabeled instances are subsequently assigned classes using nearest-centroid assignment in the learned embedding space. The proposed pipeline enables the translation of rough localization cues into fully segmented and classified nuclei instances, providing a structured dataset suitable for downstream nuclear pleomorphism assessment.

Keywords: Computer Vision, Nuclear pleomorphism, Weak supervision, Semi-supervised learning

1 Introduction

Breast cancer is the most commonly diagnosed cancer in women worldwide and is one of the leading causes of cancer-related deaths (Rakha et al., 2010 [12]). It is responsible for one in six cancer deaths among women and accounts for one in four cancer cases (Sung et al., 2021 [15]). Timely diagnosis remains crucial, and paired with suitable treatments it can significantly reduce the rate of breast cancer deaths (Wang, 2017 [16]).

There are many breast cancer prognostic factors, of which the most well-known include breast cancer and tumor size, lymph node stage, histological grade and lympho-vascular invasion (Peiris et al., 2017 [10]). Peiris et al. further explain that out of these features, the ones that provide the strongest prognostic determining factors are lymph node stage, tumor size and histological grade. For the histological grade, the most adopted system for grading breast cancer is the Nottingham Grading System (NGS). It addresses 3 criteria, which are measures that determine the degree of differentiation, or how well the tumor resembles normal breast epithelial cells (Rakha et al., 2010 [12]). The 3 criteria, which are considered during the assessment of the histological grade using the NGS are: the formation of tubular structures, mitotic activity, and nuclear pleomorphism [5]. Among these, nuclear pleomorphism assesses the differences in appearance of tumor nuclei in comparison with normal epithelium nuclei. The grade is given based on the degree of difference in size, shape and overall structure of the nuclei. As noted by Mercan et al. [8], nuclear pleomorphism shows low inter-observer agreement among the NGS criteria. This reduced agreement likely stems from the subjective nature of the criterion, where grading is based on visual assessment of deviations in nuclear structure rather than on a well-defined quantitative measure.

Automating nuclear pleomorphism analysis from histopathological images has the potential to support pathologists by providing quantitative and reproducible measurements. However, developing such automated systems is challenging due to the nature of available annotations. In practical clinical settings, obtaining precise pixel-level annotations of individual nuclei is expensive and time-consuming. As a result, annotations are often sparse, imprecise, and incomplete, frequently provided only as coarse geometric markers (e.g., ellipses) that cover only a subset of nuclei within a region. This leads to a weakly supervised learning scenario where many instances remain

*xukus@stuba.sk
†vanda_benesova@stuba.sk

unlabeled and conventional supervised pipelines are difficult to apply.

In this work, we address these challenges by proposing a segmentation-driven pipeline designed to extract structured instance-level information from sparse and imprecisely annotated histopathological data. Our approach first applies a patch-based instance segmentation strategy using the Cellpose-SAM model [9] to obtain precise nuclei masks from selected regions of Whole Slide Images (WSI). Because segmentation is performed independently on overlapping patches, an Intersection-over-Union-based (IoU-based) stitching procedure is used to construct a consistent instance map of the entire region. The resulting segmentation instances are then associated with expert-provided annotations through a centroid-based matching strategy that is robust to the imprecision of elliptical annotation shapes.

Although this process allows labels to be assigned to a subset of nuclei, a large fraction of instances remains unlabeled due to the sparsity of the original annotations. To address this issue, we employ a semi-supervised classification approach. A Siamese neural network [1] trained with triplet loss [13] learns a discriminative embedding space for nuclei appearance using the matched labeled instances. Unlabeled nuclei are subsequently assigned class labels through clustering in the learned embedding space, enabling the construction of an expanded labeled dataset within the annotated region.

The proposed pipeline demonstrates how rough expert annotations can be transformed into structured instance-level data suitable for downstream analysis of nuclear pleomorphism. However, beyond this practical contribution, our study also highlights fundamental challenges that arise when learning from weak and imprecise annotations in histopathology.

In particular, we show that while semi-supervised metric learning can produce well-structured embedding spaces within a single annotated region, it fails to generalize across regions due to inconsistencies in annotation semantics. The expert-defined classes often encode multiple overlapping morphological characteristics within a single label, making class boundaries inherently ambiguous. This suggests that the problem is rooted not only in limited data, but also in label formulation.

Based on these findings, we further explore a reformulation of the task in terms of decomposed morphological attributes, moving from single-label classification toward a structured, multi-stage representation of nuclear characteristics. This perspective provides a foundation for more robust and interpretable learning from weakly annotated histopathological data.

2 Related Work

Several studies have investigated the use of deep learning methods for automating histological cancer grading. Kim et al. [6] proposed a method for predicting Nottingham Grading System scores using a foundation model combined with multiple instance learning (MIL). Their approach leverages weak slide-level labels and achieves strong performance (AUROC of 0.835), demonstrating the effectiveness of MIL-based strategies for histopathological grading tasks.

Another lower-level approach of automating the NGS grading is to train a model to evaluate individual NGS criterion. That is what Mercan et al. [8] did in their work. They evaluated histology slides for nuclear pleomorphism. For this task they used data from the Radboud University Medical Center. They were working with 39 WSIs in total from which they extracted 125 regions of interest with tumor, for which they used epithelial cell detection network (not part of this paper). They collected annotations from multiple pathologists and formulated the problem as regression on a continuous pleomorphism scale. Using a DenseNet-based model trained on image patches, they achieved low prediction error relative to the grading range.

Weakly supervised learning approaches have also been explored to reduce the need for expensive pixel-level annotations. For example, Qu et al. [11] proposed a weakly supervised framework for nuclei segmentation that relies only partial point annotations indicating a subset of nuclei locations. Their method consists of two stages. A nuclei detection model is trained in the first stage using partially labeled data. An extended Gaussian mask is used to represent annotated nuclei locations, and a self-training strategy with background propagation is used to utilize unlabeled regions and reduce false positive detections. In the second stage, the detected nuclei locations are used to generate coarse labels that guide the training of a segmentation model in a weakly supervised manner. Experimental results on two nuclei segmentation datasets - a synthesized Lung Cancer dataset and a publicly available Multi-Organ dataset - demonstrated that the proposed approach can achieve performance comparable to fully supervised methods while requiring substantially less annotation effort.

Although the primary goal of this work is nuclei segmentation, it demonstrates the effectiveness of weakly supervised learning strategies for histopathology image analysis when precise annotations are scarce. In contrast to this approach, the method proposed in this work does not aim to learn a segmentation model but instead relies on an existing cell segmentation model and focuses on learning from imprecise labels derived from higher-level annotations.

3 Dataset

In this work we use a dataset of histological images provided by the Faculty of Medicine at Comenius University in Bratislava (LFUK). The dataset consists of WSIs of breast tissue with annotations provided by 3 professional

pathologists. six unique WSIs are provided with detailed annotations focused specifically on the nuclear pleomorphism criterion. In these annotations, a selected region within the WSI was examined and individual nuclei were marked using ellipses. In total, eight annotated regions across six WSIs are provided, with nuclei marked using elliptical annotations by expert pathologists. Table 1 shows classes that were used to describe atypical nuclei. Additionally, reference cells, such as lymphocytes, are annotated to provide a baseline for size comparison.

Table 1: Annotated instance classes describing atypical nuclei

Name translated from original Slovak
Multinucleated cell
Large irregular nucleus
Multiple nucleoli
Large nucleolus
Large nucleus
Multiple irregular nucleoli
Irregular nucleus
Large nucleus with numerous nucleoli
Hyperchromatic nucleus
Vesicular nucleus
Large irregular hyperchromatic nucleus
Mitosis
Hyperchromatic irregular nucleus

A significant challenge of this work is in the analysis of the instance annotations. The experiments we conducted were directly shaped by the weaknesses of these annotations. Two issues in particular contributed to the proposed approach: the scarcity of annotated data and the roughness of the annotations themselves.

The annotations are provided as bounding ellipses rather than precise segmentation masks, and a further examination shows that these ellipses are often only an approximate indication of the intended nucleus. In many instances, the annotated region, outlined by an ellipse, covers not only the intended nucleus but also fragments of neighboring cells, tissue artifacts, or other background structures. In other cases, the ellipse only partially overlaps the target nucleus, leaving noticeable portions of it outside the area of the ellipse. As a result, the annotations cannot be treated as pixel-accurate ground truth but rather as rough guides to where relevant nuclei are expected to be. This creates a need for a targeted pipeline to preprocess the tissue region and extract nuclei level information (like segmentation masks or detection boxes) and only after that use these annotations to assign an atypic class to the nucleus.

The second challenge of this dataset is the data scarcity. Although there are in total 77 WSIs, only six of them are provided with a nuclei-level ellipse annotated region. This by itself might be problematic depending on the actual count of annotated nuclei. However, a more significant challenge is that the annotated region itself is not fully annotated — the annotations are not exhaustive and leave quite a lot of unannotated nuclei in the selected region for annotation. Because the annotations are sparse, supervised learning becomes more challenging. A model can identify a nucleus correctly, but still get penalized simply because the annotation is not there, and thus is forced to learn to treat it as a background.

Together, these constraints prevent the dataset from being directly used in a conventional supervised training pipeline. Instead these annotations can only serve as approximate localization cues. This motivated the development of a pipeline designed to interpret and refine the rough annotations rather than relying on them as precise nuclear boundaries or as a fully reliable ground truth.

4 Proposed Solution

Our proposed solution utilizes a separate patch-based segmentation pipeline and a Siamese based embedding space-based classification of the unclassified segmented nuclei. The overview of the whole pipeline of the final solution is provided as a graphical abstract in Figure 1.

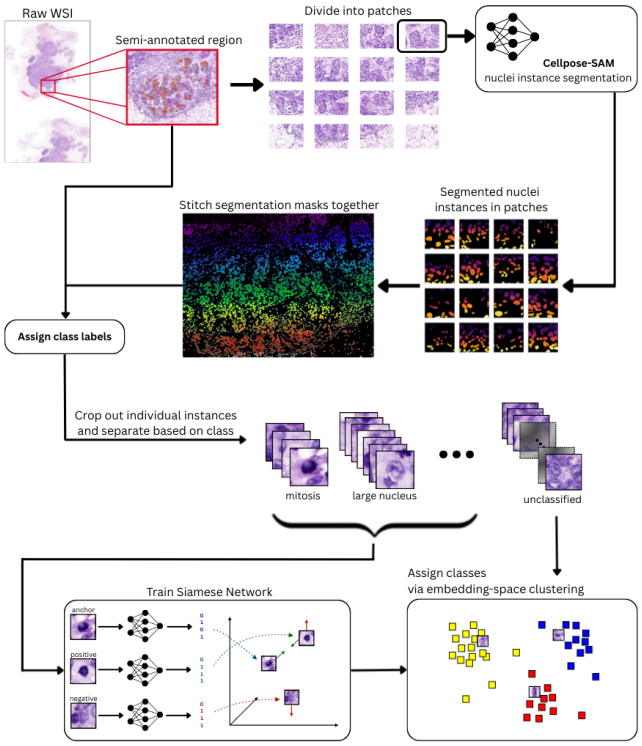


Figure 1: Graphical abstract showing the overall pipeline of the proposed solution.

4.1 Patch-Based Segmentation

Our solution begins with a patch-based segmentation pipeline for which we utilize an existing segmentation

model – Cellpose-SAM [9]. It is a cell and nucleus segmentation model that integrates the pre-trained Vision Transformer (ViT-L) backbone of the Segment Anything Model (SAM) [7] into the Cellpose framework. Described by its authors as a model "with superhuman generalization" that exceeds "inter-human agreement and approaches the human-consensus bound".

We run the Cellpose-SAM model inside of our segmentation pipeline on patches of size 224×224 with the default parameters as they are used in examples provided by the Cellpose-SAM authors at `https://github.com/MouseLand/cellpose`. The segmentation pipeline processes selected region (resolution of approximately 3000×3000) in which the expert annotations were provided. The region is divided into patches of size 224×224 with a stride of 192×192, which creates a slight overlap of 32 pixels. Each patch is segmented using the Cellpose-SAM model and the segmentation masks are stitched back together. To achieve a reliable nucleus segmentation mask a stitching strategy had to be designed to work around merging the duplicated instance masks of the overlapping regions of the neighboring patches. We implemented an IoU-based matching stitching strategy. This stitching strategy assigns consistent global nuclei IDs by comparing each patch-level nucleus to already-stitched nuclei in overlapping regions and merging them when their IoU exceeds a threshold ($\tau = 0.5$), otherwise, it creates a new instance.

With this segmentation pipeline we were able to get a fully segmented region of a WSI. The instance segmentation mask is illustrated in the Figure 2

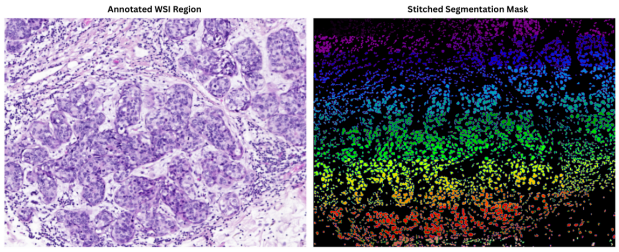


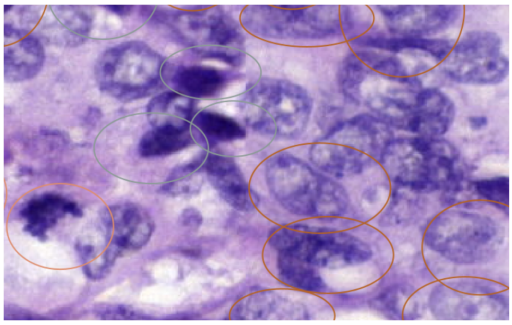
Figure 2: Side-by-side visualization of a region in a WSI (on the left) and a corresponding segmentation mask (on the right) as the output of our segmentation pipeline. The segmentation mask visualizes individual instances using distinct colors sampled from a continuous color spectrum.

4.2 Label Matching

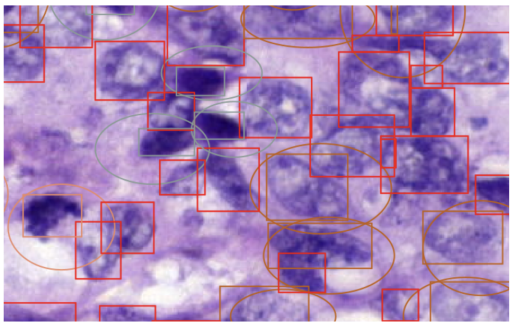
After obtaining precise segmentation masks of the nuclei instances we were able to assign a class name to a subset of these instances from the expert-annotated labels. The challenge here was to match the mask of a nucleus to its intended expert annotation, which was made difficult due to the imprecise nature of the expert annotations. These are provided as elliptical shapes that do not guarantee that the whole area of the nucleus is within that ellipse and on the other hand, in some cases it also encompasses other neighboring nuclei or their parts.

To achieve reliable mapping of expert-annotated labels to their intended nuclei despite the imprecision of the annotations, we implemented an algorithm that selects, from all segmentation delineations intersecting a given expert-provided elliptical annotation, the delineation that is closest based on centroid distance. This centroid-based matching was chosen over overlap-based matching such as intersection area or IoU, as these approaches may incorrectly prefer small nuclei that are fully contained within a large annotation ellipse intended for a larger nucleus, where the true target nucleus may not be fully contained within the annotation. The proposed approach produced visually consistent results as observed during expert evaluation and most consistent matches based on qualitative inspection.

Figure 3 shows a selected annotated region in a WSI and illustrates the expert annotations (Figure 3a) side-by-side with the matched and unmatched segmentation annotations (Figure 3b) overlaid on top of the same WSI region.



(a) Expert annotations overlaid on a WSI region.



(b) Expert annotations and segmentation annotations overlaid on a WSI region.

Figure 3: Result of expert-provided label matching to segmented instances in a selected region. The top image shows only the expert annotations. The bottom image additionally shows the segmentation instances (rectangles). Instances matched to an expert annotation are colored according to the corresponding expert label, while unmatched instances are shown in red.

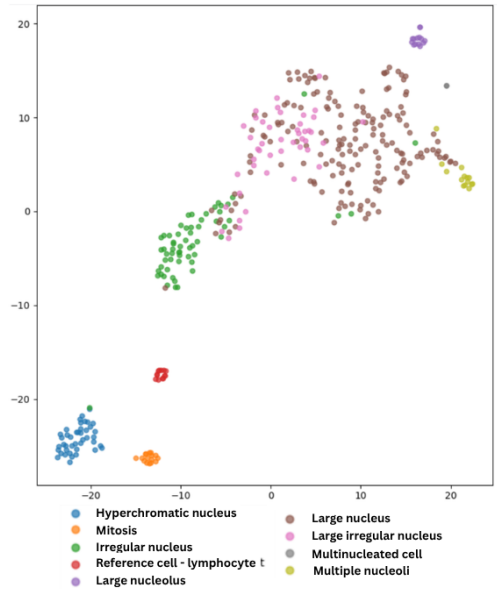
4.3 Semi-Supervised Classification

The label-matching procedure allowed us to assign class labels to a subset of the segmented nuclei, however, a large portion of instances remained unlabeled. To assign class information to these remaining instances, we employed a neural network learning-based approach.

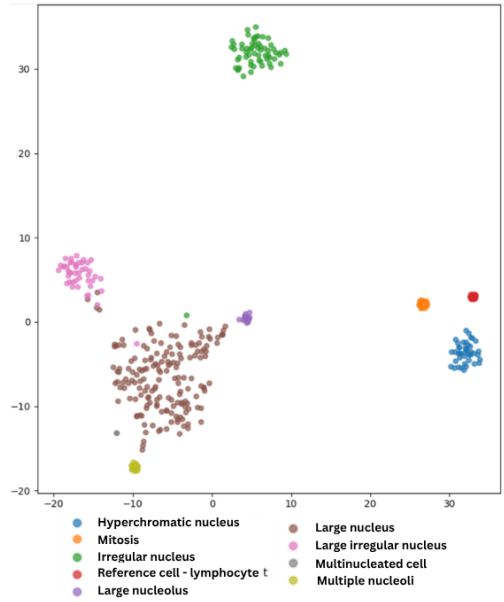
Specifically, we utilized a Siamese neural network trained with a triplet loss to learn an embedding space for the expert-provided classes. This enabled us to assign class labels to the majority of previously unclassified instances. All segmented nuclei were cropped into individual images of uniform dimensions, specified by the size of the largest nucleus with an added margin. These crops were grouped according to their assigned class labels, with an additional "unclassified" category containing instances that could not be matched to an expert annotation in the previous step. The network backbone was based on a ResNet architecture. We initially utilized a ResNet pretrained on ImageNet [4], which produced reasonable results, however, replacing it with a ResNet pretrained on histopathology images [3] led to improved class cluster separation, which we observed in the embedding space visualizations. Figure 4 compares the embedding distributions obtained with ImageNet and histopathology-pretrained backbones. Based on this comparison, the histopathology- pretrained model was selected and subsequently used to assign class labels to instances from the unclassified category.

To classify the remaining unlabeled nuclei, we first calculated a representative centroid for each expert-defined class by averaging the normalized embeddings of its labeled instances. Unlabeled instances were then mapped into the same embedding space. Class assignment was performed by identifying the class centroid with the highest cosine similarity to the unlabeled instance. To ensure high-confidence assignments, we applied a similarity threshold of 0.6, with which we observed stable behavior. This value was selected as a reasonable cutoff in the absence of reliable validation data, given the presence of noisy and incomplete labels. Instances falling below this threshold were categorized as 'unknown'.

Training Details The Siamese network was trained using the Adam optimizer with a learning rate of 1×10^{-4} and cosine annealing. Training was performed for 20 epochs with a batch size of 32. Triplets were sampled randomly, and the model was optimized using the standard triplet margin loss ($m = 1.0$). The embedding dimension was set to 128. Data augmentation included flips, rotations, and brightness/contrast adjustments. All models were initialized from pretrained weights (ImageNet or histopathology-specific) and fine-tuned on nuclei crops.



(a) Embeddings using an ImageNet-pretrained backbone.



(b) Embeddings using a WSI-pretrained backbone.

Figure 4: Comparison of embedding spaces learned by the Siamese network using different backbone initializations. Illustrating the effect of generic ImageNet pretraining versus domain-specific WSI pretraining.

5 Results

The following section presents both qualitative and quantitative evaluations of the proposed pipeline. The qualitative analysis examines the visual outputs of the segmentation and classification stages, while the quantitative analysis assesses the structure of the learned embedding space.

The presented results are based on a single annotated WSI region, comprising approximately 300 expert-annotated nuclei instances used for Siamese network training and label assignment, and approximately 4,500 additionally segmented instances classified by the pipeline. Qualitative and quantitative evaluations are performed on this region.

5.1 Qualitative Results

The proposed method was applied to the histological images in the dataset. Visual inspection by the authors of the generated predictions indicates that the model is able to identify relevant cellular structures and produce outputs consistent with the expected nuclear morphology. Figure 5 shows representative examples of the model predictions.

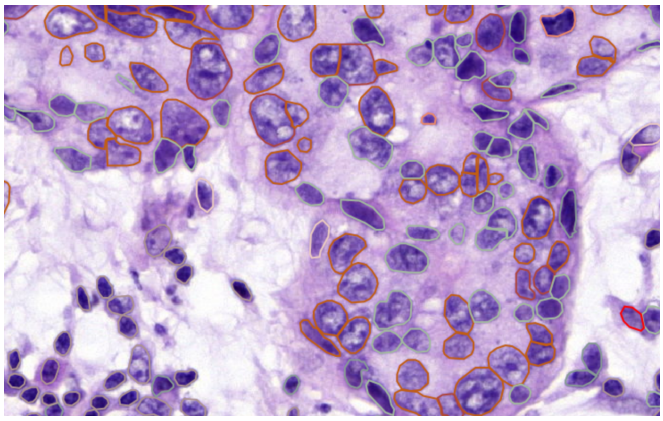


Figure 5: Example of segmented and classified (distinct colors represent classes) nuclei instances in a WSI region.

5.2 Quantitative Results

To evaluate the quality of the learned representation, we analyzed the embedding space distribution of the previously unclassified nuclei after label assignment. For the evaluation of the structure of the embedding space, we use several clustering metrics. The Silhouette score measures how well-separated clusters are by measuring both inter-cluster separation and intra-cluster cohesion [2], with higher values indicating better separation. The Davies-Bouldin (DB) index evaluates intra-cluster similarity relative to inter-cluster differences [14], where lower values are better. The Calinski-Harabasz (CH) score is a ratio of inter-cluster variance and the overall intra-cluster variance [17], with higher values indicating better-defined clusters. Finally, 5-NN consistency measures how often the nearest neighbors share the same class label, reflecting local structure in the embedding space.

The metrics in Table 2 represent the structural integrity of the clusters formed in the embedding space. The high 5-NN consistency (0.98) and Silhouette score (0.63) for the Histopathology-pretrained model indicate that the Siamese network successfully projected the unlabeled data into a highly structured space where instances form tight, distinct groups. Interestingly, the ImageNet model achieves a higher Calinski-Harabasz (CH) score of 1259.35. This difference is probably due to the sensitivity of the CH metric to global variance and absolute distance from the data centroid. The ImageNet model may produce a more expansive global distribution, whereas the Histopath model creates more compact clusters that, while being closer together in the high-dimensional space, are more discriminative. This suggests that domain-specific pretraining captures more consistent biological features, leading to more reliable semi-supervised labeling.

Pretraining	Silh. ↑	DB ↓	CH ↑	5-NN ↑
ImageNet	0.29	0.79	**1259.35**	0.86
Histopathology	**0.63**	**0.41**	432.53	**0.98**

Table 2: Embedding evaluation: Silh.=Silhouette, DB=Davies-Bouldin, CH=Calinski-Harabasz. 5-NN = 5-nearest-neighbor accuracy.

5.3 Expert Evaluation

The outputs of the proposed pipeline were qualitatively reviewed with an expert pathologist in an informal walk-through session. The expert was shown representative regions of a whole-slide image, including the segmented nuclei instances and their assigned class labels, and assessed whether the predicted labels appeared consistent with the observable nuclear morphology. While this review was not a systematic per-nucleus evaluation, the expert indicated that the segmentation correctly delineates individual nuclei and that the label assignments appear plausible given the visual characteristics of the structures. A limitation of this evaluation is its informal nature. A structured validation study, in which pathologists systematically grade annotated regions, is planned as future work.

6 Limitations and Further Analysis

We conducted an extended experiment to assess the generalizability of the proposed pipeline. We trained the Siamese network on annotations from four WSI regions and evaluated on a reserved fifth region. This experiment revealed significant overfitting, where the network failed to generalize across regions even when aggressive data augmentation was applied.

Key finding: The primary limitation does not stem solely from data scarcity, but from the structure of the annotation labels themselves. The expert-defined classes encode multiple overlapping morphological characteristics within a single label (e.g., *Large irregular nucleus* combines both size and shape irregularity). As a result, the label space is inherently inconsistent: two nuclei that share one characteristic but differ in another may be assigned to different classes, or the same class may be assigned to nuclei with different visual appearances. This ambiguity likely contributes to the poor generalization observed across regions.

To address this limitation, we explored a reformulation of the classification task based on decomposing expert labels into a set of independent morphological attributes. Specifically, we designed a two-stage classification pipeline. In the first stage, nuclei are assigned to broad categories (reference, mitotic, pleomorphic). In the second stage, pleomorphic nuclei are further described using a set of binary attributes capturing characteristics such as size, shape irregularity, chromatin pattern, and nucleoli presence.

Initial experiments with this formulation were conducted using a DINOv2-based backbone (Phikon-v2) with LoRA adaptation and a 4-channel input (RGB + segmentation mask). Although this approach improved interpretability and aligned better with the underlying annotation structure, it introduced new challenges. While the initial broad category classification achieved a promising accuracy of 0.830, the multi-label characteristic prediction struggled, yielding an Exact Match Ratio of only 0.1265. In particular, extreme class imbalance across attributes significantly hindered learning, even when applying weighted sampling, focal loss, and strong data augmentation. Additionally, correlations between characteristics in the dataset led to biased predictions.

These findings indicate that future progress depends not only on model design but also on the restructuring of label representations and the mitigation of dataset bias.

7 Conclusion

In this paper, we presented a semi-supervised framework designed to transform sparse and imprecise clinical annotations into structured data for automated nuclear pleomorphism analysis. By integrating the Cellpose-SAM model with a Siamese-based pipeline, rough localization cues in the form of elliptical annotations can be converted into detailed nuclei instances with associated semantic information.

Our experiments show that domain-specific histopathology pretraining substantially improves the quality of the learned embedding space, leading to more compact and better-separated clusters compared to ImageNet-based initialization. These results indicate that domain-specific representations capture more consistent morphological features and provide a stronger foundation for semi-supervised learning in histopathology.

At the same time, our analysis highlights a fundamental limitation that extends beyond the proposed method. We observe that the expert-defined labels encode multiple overlapping morphological characteristics within a single class, resulting in an inherently ambiguous and inconsistent label space. This ambiguity limits the ability of metric learning approaches to generalize across regions, even when the embedding space is locally well-structured.

Overall, the proposed framework demonstrates a practical approach for leveraging weak and imprecise annotations to extract structured instance-level data. More importantly, our findings suggest that future progress in automated nuclear pleomorphism assessment depends not only on model design, but also on the formulation of label representations to better reflect the underlying morphological characteristics.

8 Future Work

The findings of this work suggest that improving generalization requires both better label representations and more balanced data distributions. While the proposed attribute-based formulation provides a more consistent representation of nuclear morphology, its effectiveness is currently limited by severe class imbalance and correlations between features.

To address particularly the inter-feature correlation, future work will pivot from a single multi-label model to a series of independent binary classifiers — one for each pleomorphic characteristic. By assigning a separate LoRA configuration to each characteristic, we aim to isolate the learning process, prevent artificial dependencies, and provide a more robust, interpretable nuclear pleomorphy characterisation.

Next direction for future work is to address the limited amount of data and the overall class imbalance. This includes exploring strategies, such as targeted data collection, synthetic data generation, or more advanced reweighting and sampling techniques.

Another important step is conducting a structured validation study in collaboration with expert pathologists. This would involve systematic annotation and evaluation at the nucleus level, enabling the use of standard supervised metrics and providing a more rigorous assessment of clinical relevance.

References

[1] Jane Bromley, Isabelle Guyon, Yann LeCun, Eduard Säckinger, and Roopak Shah. Signature verification using a "siamese" time delay neural network. In *Proceedings of the 7th International Conference on Neural Information Processing Systems*, NIPS'93, page

737–744, San Francisco, CA, USA, 1993. Morgan Kaufmann Publishers Inc.

[2] Giulia Bruno, Tania Cerquitelli, Silvia Chiusano, and Xin Xiao. A clustering-based approach to analyse examinations for diabetic patients. In *2014 IEEE International Conference on Healthcare Informatics*, pages 45–50, 2014.

[3] Ozan Ciga, Tony Xu, and Anne Louise Martel. Self supervised contrastive learning for digital histopathology. *Machine Learning with Applications*, 7:100198, 2022.

[4] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255. IEEE, 2009.

[5] Jamie DePolo. Breast cancer grades (nottingham grade). `https://www.breastcancer.org/pathology-report/breast-cancer-grades`, August 2023. Accessed: 2026-03-06.

[6] Jun Seo Kim, Jeong Hoon Lee, Yousung Yeon, Doyeon An, Seok Jun Kim, Myung-Giun Noh, and Suehyun Lee. Predicting nottingham grade in breast cancer digital pathology using a foundation model. *Breast Cancer Res*, 27(1):58, April 2025.

[7] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. Segment anything, 2023.

[8] Caner Mercan, Maschenka Balkenhol, Roberto Salgado, Mark Sherman, Philippe Vielh, Willem Vreuls, António Polónia, Hugo M Horlings, Wilko Weichert, Jodi M Carter, Peter Bult, Matthias Christgen, Carsten Denkert, Koen van de Vijver, John-Melle Bokhorst, Jeroen van der Laak, and Francesco Ciompi. Deep learning for fully-automated nuclear pleomorphism scoring in breast cancer. *NPJ Breast Cancer*, 8(1):120, November 2022.

[9] Marius Pachitariu, Michael Rariden, and Carsen Stringer. Cellpose-sam: superhuman generalization for cellular segmentation. *bioRxiv*, 2025.

[10] H Peiris, L Mudduwa, N Thalagala, and K Jayatialake. The value of nottingham grade in breast cancer re-visited in the sri lankan setting. *Malays J Pathol*, 39(2):141–148, August 2017.

[11] Hui Qu, Pengxiang Wu, Qiaoying Huang, Jingru Yi, Zhennan Yan, Kang Li, Gregory M. Riedlinger, Subhajyoti De, Shaoting Zhang, and Dimitris N. Metaxas. Weakly supervised deep nuclei segmentation using partial points annotation in histopathology images. *IEEE Transactions on Medical Imaging*, 39(11):3655–3666, November 2020.

[12] Emad A Rakha, Jorge S Reis-Filho, Frederick Baehner, David J Dabbs, Thomas Decker, Vincenzo Eusebi, Stephen B Fox, Shu Ichihara, Jocelyne Jacquemier, Sunil R Lakhani, José Palacios, Andrea L Richardson, Stuart J Schnitt, Fernando C Schmitt, Puay-Hoon Tan, Gary M Tse, Sunil Badve, and Ian O Ellis. Breast cancer prognostic classification in the molecular era: the role of histological grade. *Breast Cancer Res.*, 12(4):207, July 2010.

[13] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, page 815–823. IEEE, June 2015.

[14] Akhilesh Kumar Singh, Shantanu Mittal, Prashant Malhotra, and Yash Vardhan Srivastava. Clustering evaluation by davies-bouldin index(dbi) in cereal data using k-means. In *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)*, pages 306–310, 2020.

[15] Hyuna Sung, Jacques Ferlay, Rebecca L. Siegel, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal, and Freddie Bray. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 71(3):209–249, 2021.

[16] Lulu Wang. Early Diagnosis of Breast Cancer. *Sensors*, 17(7):1572, July 2017. PubMedID: 28678153.

[17] Szymon Łukasik, Piotr A. Kowalski, Małgorzata Charytanowicz, and Piotr Kulczycki. Clustering using flower pollination algorithm and calinski-harabasz index. In *2016 IEEE Congress on Evolutionary Computation (CEC)*, pages 2724–2728, 2016.