Heram: Multi-Magnification Graph-Structured Whole Slide Image Representation

Ali Khajegili Mirabadi
School of Biomedical Engineering
University of British Columbia
`ali.mirabadi@ubc.ca`

Graham Archibald
School of Biomedical Engineering
University of British Columbia
`graham.archibald@ubc.ca`

Abstract

Ovarian carcinoma has the highest mortality of all female reproductive cancers with treatment being histotype-specific. Five common histotypes have been identified by pathologists via multi-magnification microscopic examination. However, there is only moderate interobserver agreement between general pathologists. At present, deep learning models have improved upon pathologist performance, yet they have not taken advantage of more nuanced tissue structure as well as the multi-magnification nature of the histopathologic images. Therefore, in this project, we introduce a graph data structure for representing inter- and intra-magnification information flow in the data, and use it to devise a GCN based architecture for further improving the performance of Deep Learning-based models. This model represents a more systematic approach as the introduced graph structure enables the capture of more comprehensive and interpretable histopathologic information relating to the underlying mechanisms of the disease. Performance was assessed on a set of 948 whole slide images corresponding to 485 patients. Our best performing model achieved a balanced accuracy of 85.7% at the patient-level along with a Cohen's Kappa of 0.86 and outperformed competing state-of-the-art deep learning architectures by a margin of at least 1.3%. The performance of our model suggests potential for use as a diagnostic adjunct for pathologists to aid in improving diagnostic accuracy in the clinical setting.[1]

1 Introduction

Projected to cause an estimated 12,810 deaths in the U.S in 2022, ovarian carcinoma is the deadliest cancer of the female reproductive system [17] and the fifth deadliest cancer for women overall. Ovarian epithelial cancer (carcinoma) occurs when epithelial cells from the ovaries or fallopian tubes mutate or change in a manner which allows them to grow uncontrollably within one of the ovaries, acquiring the potential to spread around the body [11]. It has five common histotypes, each with distinct molecular, genetic, cellular, and clinical attributes. In the U.S, high-grade serous carcinoma (HGSC) accounts for 70% of cases, clear cell ovarian carcinoma (CCOC) accounts for 12%, endometrioid (ENOC) for 11%, low-grade serous (LGSC) for 4%, and mucinous carcinoma (MUC) for 3%.

To diagnose cancer, pathologists generally use a variety of lenses in their inspection of a tissue sample under the microscope, switching between different magnifications as needed. They usually begin with low magnifications to identify regions of interest for making preliminary decisions before increasing magnifications to confirm or rule out those diagnoses[14].

---

[1] Both students contributed equally to the project: Ali designed research; Ali and Graham performed research; Graham analyzed data; and Both wrote the paper and prepared the slides.

Preprint. Under review.

Two of the main challenges associated with histotype classification performed by pathologists are diagnostic reproducibility and interobserver disagreement [5]. When diagnosing hematoxylin & eosin (H&E)-stained tissue sections, pathologists without current, gynecologic-specific training produce an interobserver agreement of only 0.54-0.67 Cohen's kappa [3] [12]. Additionally, the field of pathology is simultaneously facing challenges of increased workload as well as a shortage of pathologists creating demand for the rapid switch to digital pathology oriented workflows aiming to improve efficacy as well as efficiency [13].

Deep learning (DL) has demonstrated great potential for wide ranging clinical applications within the field of digital pathology in the last decade. One such application has been using DL as a tool for aiding pathologist diagnostic accuracy as well as efficiency in the context of cancer diagnoses. In this context, deep learning algorithms are frequently trained on whole slide images (WSIs) which are obtained by scanning a glass slide containing a tissue section via a whole slide scanner. H. Farahani et al. [4] improved upon the Cohen's kappa of general pathologists using deep convolutional neural networks from 0.54-0.67 in the case of pathologists up to 0.8134 using their DL-based models. In this project, we aim to further this progress by leveraging graph convolutional neural networks (GCNs) [10] with a multi-magnification component as they have been shown to capture more comprehensive and interpretable histopathologic information relating to the underlying mechanisms of a disease [1]. Ultimately, the closer an DL model follows the pathologist's routines, the more realistic and interpretable its results may become.

This research report is organized as follows: a short survey related works will be presented in section 2. Our method is explained on section 3. In section 4, the details of our experiments and results are reported. Finally, we conclude the research project and go over future directions in sections 5 and 6, respectively.

2 Related Work

2.1 Histopathologic tasks using GCN

Several recent works have attempted to leverage GCNs in a variety of histopathological classification contexts [1]. Some notable works include Guan et al. [6] which introduced a node-aligned GCN (NAGCN), incorporating local and global structural information to enrich the representation of WSIs. While this work achieved comparable performance with state-of-the-art methods, they did not include multi-magnification elements in their graph representation and were testing on relatively simple classification tasks.

2.2 Hierarchical Learning

Multiple efforts have been made to combine multi-magnification information in the context of histopathology. This has been explored within the context of vision transformers where Chen et al. [2] introduced their Hierarchical Image Pyramid Transformer (HIPT) architecture which uses two levels of magnification. While they outperformed state-of-the-art methods for cancer subtyping and survival prediction, they express major limitations related to the computational intractability of pre-training HIPT on commercial workstations. Zhou et al. [18] introduced the novel cell-graph convolutional neural network (CGC-Net) which produces graph representations of WSIs with edges denoting cellular interactions. They further present Adaptive GraphSage, a graph convolutional technique which combines multi-level features. While this work produces state-of-the-art performance, their classification task is one of *normal* vs *low-grade* vs *high-grade* gland differentiation whereas ours is a five histotype classification task with more nuanced structural differences.

2.3 KimiaNet

With recent progress in deep learning, deep features, i.e. high-level embeddings from a deep network, have advanced past handcrafted features and are considered the most robust sources for image representation. Unfortunately, pre-trained networks such as DenseNet draw their features from millions of non-medical and non-histopathological images. In this context, KimiaNet [15] was developed by employing the architecture of DenseNet [7] with four dense blocks trained on more than 240,000 histopathology images (patches). This work provides a densely connected network with weak labels which can serve as a feature extractor for more specific histopathological tasks.

2.4 Attention Based Models

2.4.1 DeepMIL

The first architecture we implemented as a competing state-of-the-art model was DeepMIL [8], a model that combines permutation invariant multiple instance learning (MIL) with an attention-based neural network. MIL is a type of supervised learning where labelled data (in our case, a WSI with a slide level label) is broken up into a "bag of instances" which is considered to be weakly labelled meaning that the patches are from one WSI but each patch is not individually labeled. DeepMIL uses attention-weighted patch instance feature vectors to compute bag-level features which are subseqeuntly classified using a fully connected layer. In our implementation, DeepMIL was trained using cross entropy loss as well as a learning rate of 0.001 for Adam ptimizer with a weight decay of 0.01.

2.4.2 VarMIL

Secondly, we compared our model to VarMIL [16], a model based on DeepMIL, focused on addressing it's limitations related to histopathology. This limitation is that the bag-level latent features of DeepMIL consider neither patch interactions nor high level features of the WSI. To address this shortcoming, VarMIL adds an additional attention weighted variance module to the architecture which represents the tissue heterogeneity of different patches within a WSI. We used the same settings as described above in DeepMIL for VarMIL.

3 Method

As mentioned earlier, most multi-magnification methods (to best of our knowledge, all of them) can be categorized under hierarchical learning [2] or ensemble learning [14], where the former tries to convey information from lower magnifications to higher magnifications, and the latter processes different magnifications independently and aggregates their information at the end of the process to represent the data. The issue with hierarchical learning is that it is not flexible with information flowing from higher magnifications to lower ones, whereas pathologists regularly increase and decrease magnification on the image when they are diagnosing challenging cases. As for ensemble learning, it can be implied that such methods are assuming that samples from different magnifications are $i.i.d$ (which is generally not true) because they feed into and train them with independent deep learning models, and then ensemble the output features of the models as the final representation for the data and as the input to a classifier. This way of unstructured learning might not capture inter- and intra- magnification feature interactions.

Therefore, as we know that valuable information is contained in different magnifications, and the information flow does not follow a fixed hierarchy for each patient and slide, we aim to first introduce a graph data structure for representing WSIs as graphs, then we use the graphs for training GCNs to represent their information in a structured paradigm to classify the five subtypes of ovarian carcinoma.

3.1 Heram

The multi-magnification nature of WSIs needs to be taught to a model, so that it can be utilized for learning and inference by the model. Here, we start by segmenting the tissue region using HistoQC [9], then extract n patches in each magnification of $5x$, $10x$, and $20x$ such that all patches are 1000×1000 pixels since the computational limitations of neural networks do not allow dimensional variation in their inputs. On the other hand, as the magnification doubles, the corresponding number of windows increases four fold. To address this challenge, higher magnification patches are at the center of the lower magnification patches yet with higher resolution and the same size. For example, the corresponding $10x$ patch to $5x$ is the centered window of $5x$, and the corresponding $20x$ patch to $10x$ is the centered window of $10x$ as illustrated in Fig. 1. This approach has been taken to mimic the exact hierarchical zooming procedure that a pathologist does by adjusting microscope lenses to reach the optimal focus (magnification).

Using this multi magnification patch extraction approach, we extract n patches for each magnification to represent a WSI with $3n$ patches. The coordinates of patches are chosen randomly in $20x$ and subsequently fixed for other magnification to follow the aforementioned patch extraction approach.

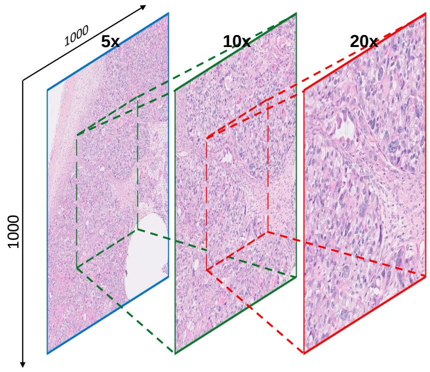

Figure 1: Going from low to high resolution with a fixed size, the area will be reduced, but more details are seen. From left to right, a given window of 1000×1000 pixels at $5x$ magnification in blue, then a $10x$ window of the same size in green, and finally a $20x$ window of the same size in red are shown.

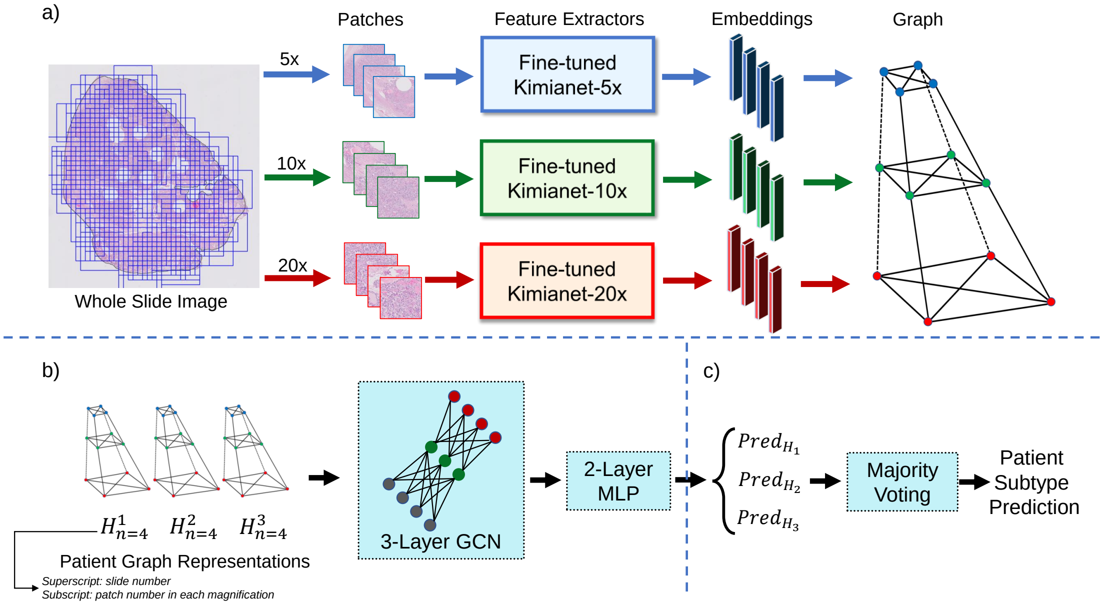

Figure 2: Overview of our workflow beginning with WSIs and outputting patient subtype predictions. **a)** shows the WSI being tiled into patches of varying magnification which are then embedded and assembled into a three layer multi-magnification graph. In **b)**, patient graph representations are used independently to feed into a three-layer GCN and subsequently a two-layer MLP. Finally, the predictions for each graph are amalgamated via majority voting in **c)**, producing the overall patient subtype prediction (For the sake of illustration, $n = 4$ is used to show the structure of Heram).

We then fine-tune KimiaNet for each magnification separately and design a graph H_n with $3n$ nodes, and $\frac{(3n+1)n}{2}$ edges, consisting of three levels: each a complete graph K_n. The graph H_n is a connected graph, meaning that there is a path for every node to transfer messages, so even in the worst case, the farthest node from a given node can be accessed through a path of length 3. Each node in the graph is an embedding of the last pooling layer of corresponding KimiaNet at that magnification as it is depicted in Fig.2-**a)**. Accordingly, a WSI can be represented by a multi-magnification graph.

Each magnification level in Heram is a complete graph to represent the inter-magnification relationship between patches. In addition, corresponding patches in different magnifications are connected together to represent intra-magnification relationship between $5x$, $10x$, and $20x$ magnifications patches, making a pyramidal structure (we decided to name Heram, which is how it is translated into Persian).

Due to clinical consideration, a patient could potentially have more than one slide, so it is essential to consider different slides of the patient when making the final subtype prediction. To do so, we represent each slide with a separated graph as depicted in Fig.2-**b)**. Considering the structure of Heram, a three-layer GCN can pass messages to every node in the graph. Thus, Heram has a three-layer GCN with embedding size of 128 while the input embedding is of size 1024 as in KimiaNet's last pooling layer. To calculate the final graph representation, average readout is adopted as part of the model. The final graph embedding is passed into a two-layer MLP with ReLU as the intermediate activation function. Heram's prediction for each slide of a given patient is calculated, and is fed into a majority voting module to determine the patient-level prediction as shown in Fig.2-**c)**.

4 Experiments

4.1 Data preparation and processing

We use a dataset of Ovarian Carcinoma consisting of 948 WSIs with five histotypes including HGSC (410), CCOC (167), ENOC (237), LGSC (69), and MUC (65). As there are no pathologist patch-level annotations for the dataset, and only patient level labels are available, this project is categorized as Weakly Supervised Learning because there is a label for WSIs for a given patient, but we do not have any label or annotation for patches.

The main component of processing the WSIs is to apply HistoQC to the dataset [9]. The reasons for this are twofold. Firstly, HistoQC masks the background region of the WSIs with the goal of only leaving the foreground which is comprised of the tissue region. This ensures that when we extract patches randomly from the slides, we are only obtaining tissue images for model input rather than white space and other background elements. Secondly, HistoQC removes artefacts and batch effects such as pen lines, which are commonly introduced during routine slide preparation, from the WSI.

Unit	Histotype					
	HGSC	CCOC	ENOC	LGSC	MUC	Total
Patients	200	95	114	34	42	485
Slides	410	167	237	69	65	948
Patches	123,000	50,100	71,100	20,700	19,500	284,400

Table 1: Overview of our dataset.

After cleaning the data with HistoQC, we extract patches randomly from the slides. For each slide, 100 patches are randomly extracted at each of $5x$, $10x$, and $20x$ magnification for a total of 300 patches per slide. Full details of the number of patients, slides, and subsequent patches for each histotype can be seen in Table 1.

4.2 Training the deep learning model

As outlined in Fig.2, after the patches are extracted, we use them to fine-tune KimiaNet to extract features for building the graph. Then we use the same training set as Kimianet for the graphs to train our model. The three layer GCN has an embedding size of 128 for all layers, a learning rate of 0.0001 for Adam optimizer, a weight decay of 0.01, and weighted cross entropy as loss function, also contains two dropout layers with $p = 0.15$ and $p = 0.05$. During training, we chose the best performing model in terms of Balanced Accuracy for the validation set over 100 epochs. For all training and testing, the GPU hardware used was either a GeForce GTX 3090 Ti (Nvidia) or a Tesla V100-SXM2-32GB (Nvidia) based on availability. Deep Graph Library (DGL), PyTorch, Numpy, SciPy, PyGeometric, and Scikit-Learn libraries have been used to perform the experiments.

4.3 Results & Visualization

To identify the model architecture with the best performance, we used three-fold cross-validation across ten random seeds and subsequently averaged performance across splits for each of the seeds.

Our results indicate that the highest performing model was Heram (Table 2), which achieved a mean patient-level Balanced Accuracy of 85.7 ± 0.39 %. In our case, Balanced Accuracy is the most crucial metric for assessing model performance due to the inherently unbalanced nature of the data. That is to say, a model could attain an Accuracy of 70% by simply predicting all cases as HGSC, though the resulting balanced accuracy would be 20%. In fact, this reasoning is likely the explanation for the sole instance of our model being outperformed which is in Accuracy Standard Deviation seen in Table 3.

The model 'GCN: $20x$ + *Majority Voting*' attained a superior Accuracy STD of 0.48%, but we should note that it's Balanced Accuracy STD is 5.02%, implying that it overly predicted more prevalent histotypes rather than learning to correctly classify less common histotypes, which is a crucial error in the clinical setting. Consequently, The multi-magnification structure of Heram not only improves performance for various metrics, but also helps to stabilize the model toward the objective function of Balanced Accuracy in comparison with the single magnification versions.

Model	Balanced Acc. (%)	Acc. (%)	F1 Score
DeepMIL: *Multi Magnification + Majority Voting*	83.9	85.9	0.845
DeepMIL: $5x$ + *Majority Voting*	84.4	86.0	0.846
DeepMIL: $10x$ + *Majority Voting*	83.7	85.2	0.845
DeepMIL: $20x$ + *Majority Voting*	83.3	86.0	0.839
VarMIL: *Multi Magnification + Majority Voting*	83.0	85.9	0.841
VarMIL: $5x$ + *Majority Voting*	84.0	85.2	0.839
VarMIL: $10x$ + *Majority Voting*	82.6	84.7	0.832
VarMIL: $20x$ + *Majority Voting*	81.9	85.6	0.828
Heram: *Majority Voting*	**85.7**	**86.3**	**0.858**

Table 2: Patient-level results comparing Heram, with competing state-of-the-art models, DeepMIL and VarMIL. All models presented in this table are selected based on optimal Balanced Accuracy in the validation set over 100 epochs. The values reported are average on three splits and ten different random seeds.

Even though we evaluated DeepMIL and VarMIL with multi-magnification data, the result in Table 2 shows that they can not capture multi-magnification information contained in WSIs as the $5x$ instance outperformed the *Multi Magnification* version. Contrasting this, as an ablation study, Heram demonstrates that a well-structured graph design can help the model to digest multi-magnification information. To clarify, for single magnification analysis, we feed the fully connected graph of all embeddings of that magnification to the same GCN model as used for Heram. For example, 'GCN: $5x$ + *Majority Voting*' in Table 3 means that it is the same GCN model as Heram except the input data is only $5x$ magnification embeddings.

Model	Balanced Acc. (%)	Acc. (%)	F1 Score
GCN: $5x$ + *Majority Voting*	85.0	84.5	0.840
GCN: $10x$ + *Majority Voting*	83.3	84.0	0.836
GCN: $20x$ + *Majority Voting*	81.5	85.2	0.824
Heram: Majority Voting	**85.7**	**86.3**	**0.858**
Model	Balanced Acc. STD (%)	Acc. STD (%)	F1 Score STD (%)
GCN: $5x$ + *Majority Voting*	0.63	0.93	1.50
GCN: $10x$ + *Majority Voting*	0.49	1.68	0.96
GCN: $20x$ + *Majority Voting*	5.02	**0.48**	3.47
Heram: *Majority Voting*	**0.39**	1.57	**0.65**

Table 3: Ablation study:comparing Heram, with the same GCN at three different magnifications: $5x$, $10x$, and $20x$ classifying at the patient-level. All models presented in this table are selected based on optimal Balanced Accuracy in the validation set over 100 epochs. The values reported are average on three splits and ten different random seeds.

Model	Patient-level Cohen's Kappa
Pathologists	0.54-0.67
H.Farahani *et. al*	0.8134
Heram (*Avg. on 10 Random Seeds*)	**0.8395**
Heram (*Best Performing Model Among 10 Random Seeds*)	**0.8599**

Table 4: Comparing Cohen's Kappa for our model and previous works (For Heram, the values reported in the table are the mean value from three splits.

From table 4, we see that Heram has improved upon Farahani *et al.* [4] in terms of the patient-level Cohen's Kappa. This indicates that the introduced multi-magnification graph data structure alleviates interobserver disagreement.

Figure 3 shows the confusion matrix associated with our best performing model selected from ten random seeds. From this matrix, we are able to visualize the model's performance, as well as it's shortcomings in relation to classification. It is worth noting that the most common failure mode for this model is classifying EC as HGSC and *vice-versa*. Notably, this is a clinically explainable failure mode as these histotypes of ovarian carcinoma often exhibit morphological mimicry, that is, presenting features on H&E slides which mimic each others histotype with respect to architecture and cytological features. It is also worth noting that Heram performs well on rare histotypes such as MUC and LGSC, which shows the model is not skewed towards the most abundant classes (histotypes).

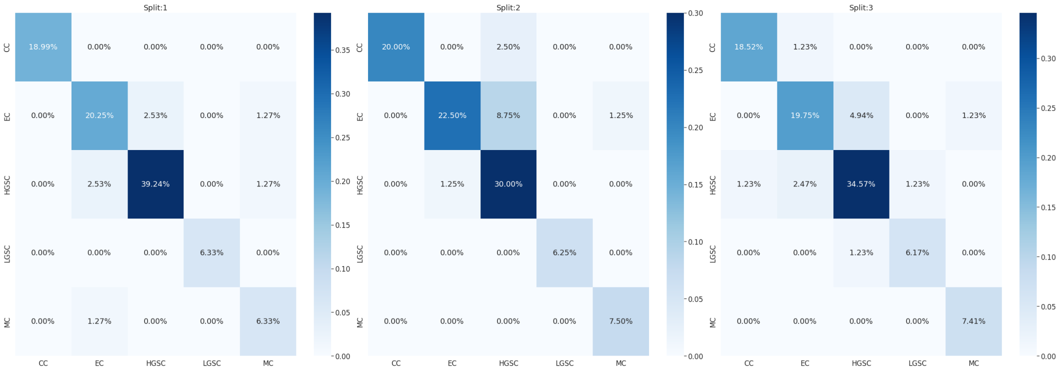

Figure 3: Confusion matrix showing how our best performing model classified histotypes at the patient-level for three splits.

5 Conclusion

Our main objective in this research project was to develop a Deep learning-based model for improving ovarian carcinoma histotype diagnosis by using the most intuitive property of WSIs: multi-magnification. We created a graph-based architecture based on the workflow of clinical pathologists and evaluated their performance on a dataset composed of 948 WSIs. Our chosen metrics for comparing models are Balanced Accuracy, Accuracy, and F1-score. Our proposed model achieved a mean patient-level balanced accuracy of 85.7% outperforming DeepMIL and VarMIL. Heram, which yields high performance on histopathology slides, advances the progress towards the implementation of structure-based deep learning tools as a diagnostic adjunct for pathologists in diagnosing ovarian carcinoma histotypes. Moreover, not only can it capture multi-magnification details contained in WSIs, but it is also more stable in comparison to single magnification methods.

At present, the 'gold standard' for clinical ovarian carcinoma histotype diagnoses is the expert pathologist diagnosis which uses H&E stained tissues samples inspected at a variety of magnification levels for identifying select structural and morphological markers. This remains a challenging task, both for reasons of time and resource consumption as well as difficulty for even highly specalized, gynaecological trained pathologists to confidently differentiate certain histotypes. While the model we present cannot, and should not replace clinical pathologists, we posit that the performance is at a level that it could be implemented into practice due to it's ability to rapidly formulate a diagnosis with a high degree of diagnostic concordance.

We conclude that Heram, which incorporates graph-structured learning as well as multi-magnification information, provides performance at a level which suggests it can act as a valuable clinical adjunct for informing histotype diagnosis of ovarian carcinoma.

6 Future Works

Heram has shown state-of-the-art performance and needs to be further investigated to identify its other potential strengths and limitations that were not discussed in this manuscript. Therefore, we plan to assess the generalizability of the model by testing it on two completely different cancer datasets, a Bladder cancer dataset and a Colorectal cancer dataset. The number of patches (embeddings) for each magnification, n, is also a hyperparameter of the model, and we are going to design further experiments to elucidate its role in the model's functionality. The main reason we extracted patches randomly across different regions of the given WSI is that we did not intend to impose prior knowledge to the model. That is to say, different cancers have distinct morphological characteristics, and we hope the model can understand this since we extracted patches randomly, yet we need to test the hypothesis that the model can learn morphological patterns of cancers and distinguish between them. To this end, we are planing to add a comprehensive analysis on the two aforementioned datasets.

References

[1] David Ahmedt-Aristizabal, Mohammad Ali Armin, Simon Denman, Clinton Fookes, and Lars Petersson. A survey on graph-based deep learning for computational histopathology. *CoRR*, abs/2107.00272, 2021.

[2] Richard J. Chen, Chengkuan Chen, Yicong Li, Tiffany Y. Chen, Andrew D. Trister, Rahul G. Krishnan, and Faisal Mahmood. Scaling vision transformers to gigapixel images via hierarchical self-supervised learning, 2022.

[3] Jacob Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46, 1960.

[4] Hossein Farahani, Jeffrey Boschman, David Farnell, Amirali Darbandsari, Allen Zhang, Pouya Ahmadvand, Steven J. M. Jones, David Huntsman, Martin Köbel, C. Blake Gilks, Naveena Singh, and Ali Bashashati. Deep learning-based histotype diagnosis of ovarian carcinoma whole-slide pathology images. *Modern Pathology*, Sep 2022.

[5] C Blake Gilks, Esther Oliva, and Robert A Soslow. Poor interobserver reproducibility in the diagnosis of high-grade endometrial carcinoma. *Am. J. Surg. Pathol.*, 37(6):874–881, June 2013.

[6] Yonghang Guan, Jun Zhang, Kuan Tian, Sen Yang, Pei Dong, Jinxi Xiang, Wei Yang, Junzhou Huang, Yuyao Zhang, and Xiao Han. Node-aligned graph convolutional network for whole-slide image representation and classification. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 18791–18801, 2022.

[7] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks, 2016.

[8] Maximilian Ilse, Jakub M. Tomczak, and Max Welling. Attention-based deep multiple instance learning, 2018.

[9] Andrew Janowczyk, Ren Zuo, Hannah Gilmore, Michael Feldman, and Anant Madabhushi. Histoqc: An open-source quality control tool for digital pathology slides. *JCO Clinical Cancer Informatics*, (3):1–7, 2019. PMID: 30990737.

[10] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks, 2016.

[11] Martin Köbel, Steve E Kalloger, Niki Boyd, Steven McKinney, Erika Mehl, Chana Palmer, Samuel Leung, Nathan J Bowen, Diana N Ionescu, Ashish Rajput, Leah M Prentice, Dianne Miller, Jennifer Santos, Kenneth Swenerton, C Blake Gilks, and David Huntsman. Ovarian carcinoma subtypes are different diseases: implications for biomarker studies. *PLoS Med.*, 5(12):e232, December 2008.

[12] Martin Köbel, Steve E Kalloger, Sandra Lee, Máire A Duggan, Linda E Kelemen, Leah Prentice, Kimberly R Kalli, Brooke L Fridley, Daniel W Visscher, Gary L Keeney, Robert A Vierkant, Julie M Cunningham, Christine Chow, Roberta B Ness, Kirsten Moysich, Robert Edwards, Francesmary Modugno, Clareann Bunker, Eva L Wozniak, Elizabeth Benjamin, Simon A Gayther, Aleksandra Gentry-Maharaj, Usha Menon, C Blake Gilks, David G Huntsman, Susan J Ramus, Ellen L Goode, and Ovarian Tumor Tissue Analysis consortium. Biomarker-based ovarian carcinoma typing: a histologic investigation in the ovarian tumor tissue analysis consortium. *Cancer Epidemiol. Biomarkers Prev.*, 22(10):1677–1686, October 2013.

[13] Giovanni Lujan, Jennifer C Quigley, Douglas Hartman, Anil Parwani, Brian Roehmholdt, Bryan Van Meter, Orly Ardon, Matthew G Hanna, Dan Kelly, Chelsea Sowards, Michael Montalto, Marilyn Bui, Mark D Zarella, Victoria LaRosa, Gerard Slootweg, Juan Antonio Retamero, Mark C Lloyd, James Madory, and Doug Bowman. Dissecting the business case for adoption and implementation of digital pathology: A white paper from the digital pathology association. *J. Pathol. Inform.*, 12(1):17, April 2021.

[14] Maral Rasoolijaberi, Morteza Babaei, Abtin Riasatian, Sobhan Hemati, Parsa Ashrafi, Ricardo Gonzalez, and Hamid R. Tizhoosh. Multi-magnification image search in digital pathology. *IEEE Journal of Biomedical and Health Informatics*, 26(9):4611–4622, 2022.

[15] Abtin Riasatian, Morteza Babaie, Danial Maleki, Shivam Kalra, Mojtaba Valipour, Sobhan Hemati, Manit Zaveri, Amir Safarpoor, Sobhan Shafiei, Mehdi Afshari, Maral Rasoolijaberi, Milad Sikaroudi, Mohd Adnan, Sultaan Shah, Charles Choi, Savvas Damaskinos, Clinton JV Campbell, Phedias Diamandis, Liron Pantanowitz, Hany Kashani, Ali Ghodsi, and H. R. Tizhoosh. Fine-tuning and training of densenet for histopathology image representation using tcga diagnostic slides, 2021.

[16] Yoni Schirris, Efstratios Gavves, Iris Nederlof, Hugo Mark Horlings, and Jonas Teuwen. Deepsmile: Self-supervised heterogeneity-aware multiple instance learning for dna damage response defect classification directly from hamp;e whole-slide images, 2021.

[17] Rebecca L. Siegel, Kimberly D. Miller, Hannah E. Fuchs, and Ahmedin Jemal. Cancer statistics, 2021. *CA: A Cancer Journal for Clinicians*, 71(1):7–33, 2021.

[18] Yanning Zhou, Simon Graham, Navid Alemi Koohbanani, Muhammad Shaban, Pheng-Ann Heng, and Nasir Rajpoot. Cgc-net: Cell graph convolutional network for grading of colorectal cancer histology images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, Oct 2019.