Am J Cancer Res 2024;14(7):3513-3522
www.ajcr.us /ISSN:2156-6976/ajcr0157402

Original Article

Interpretable deep learning model to predict lymph node metastasis in early gastric cancer using whole slide images

You-Na Sung[1*], Hyeseong Lee[2*], Eunsu Kim[2], Woon Yong Jung[3], Jin-Hee Sohn[4], Yoo Jin Lee[1], Bora Keum[5], Sangjeong Ahn[1,6,7], Sung Hak Lee[2]

[1]Department of Pathology, Korea University Anam Hospital, College of Medicine, Korea University, Seoul, South Korea; [2]Department of Hospital Pathology, Seoul St. Mary's Hospital, College of Medicine, The Catholic University, Seoul, South Korea; [3]Department of Pathology, Hanyang University Guri Hospital, College of Medicine, Hanyang University, Guri, South Korea; [4]Department of Pathology, Samkwang Medical Laboratories, Seoul, South Korea; [5]Division of Gastroenterology and Hepatology, Department of Internal Medicine, Korea University Anam Hospital, College of Medicine, Korea University, Seoul, South Korea; [6]Artificial Intelligence Center, Korea University Anam Hospital, College of Medicine, Korea University, Seoul, South Korea; [7]Department of Medical Informatics, College of Medicine, Korea University, Seoul, South Korea. [*]Equal contributors.

Received April 24, 2024; Accepted June 24, 2024; Epub July 15, 2024; Published July 30, 2024

Abstract: In early gastric cancer (EGC), the presence of lymph node metastasis (LNM) is a crucial factor for determining the treatment options. Endoscopic resection is used for treatment of EGC with minimal risk of LNM. However, owing to the lack of definitive criteria for identifying patients who require additional surgery, some patients undergo unnecessary additional surgery. Considering that histopathologic patterns are significant factor for predicting lymph node metastasis in gastric cancer, we aimed to develop a machine learning algorithm which can predict LNM status using hematoxylin and eosin (H&E)-stained images. The images were obtained from several institutions. Our pipeline comprised two sequential approaches including a feature extractor and a risk classifier. For the feature extractor, a segmentation network (DeepLabV3+) was trained on 243 WSIs across three datasets to differentiate each histological subtype. The risk classifier was trained with XGBoost using 70 morphological features inferred from the trained feature extractor. The trained segmentation network, the feature extractor, achieved high performance, with pixel accuracies of 0.9348 and 0.8939 for the internal and external datasets in patch level, respectively. The risk classifier achieved an overall AUC of 0.75 in predicting LNM status. Remarkably, one of the datasets also showed a promising result with an AUC of 0.92. This is the first multi-institution study to develop machine learning algorithm for predicting LNM status in patients with EGC using H&E-stained histopathology images. Our findings have the potential to improve the selection of patients who require surgery among those with EGC showing high-risk histological features.

Keywords: Gastric cancer, lymph node metastasis, machine learning, deep learning, endoscopic resection, computational pathology

Introduction

Gastric cancer is the fifth most common cancer, and third most common cause of cancer-related mortality worldwide. It has a high incidence rate in Central and East Asia [1, 2]. In Korea and Japan, which are some of the hotspot countries for gastric cancer, the development of population-based screening programs has enabled early detection and treatment [3, 4]. For early gastric cancers (EGC) with a minimal risk of lymph node metastasis (LNM), the preferred treatment is endoscopic curative resection, such as endoscopic mucosal resection or endoscopic submucosal dissection [5, 6]. Endoscopic resection is superior to surgical resection for improving patient quality of life and reduce the risks associated with surgery. However, it should be performed only when LNM is absent. If LNM is suspected, additional gastrectomy with lymphadenectomy is necessary [7, 8]. The JGCA (Japanese Gastric Cancer

https://doi.org/10.62347/RJBH6076

Association) guidelines recommend the eCura scoring system to categorize patients at risk after endoscopic resection (e.g., of patients with eCura C-2 have certain LNM risk) [6]. Currently, there is no definite criteria for identifying patients who may require additional surgery after endoscopic resection. Hence, though only about 15% of patients with EGC have LNM, some of these patients undergo unnecessary additional surgery [9, 10]. Despite the use of various imaging modalities to aid in deciding the treatment strategy, this situation continues to persist [11-14]. Considering the frequency of LNM in EGC, additional surgical treatment may not be effective, but the potential risk of LNM cannot be disregarded.

Recently, as artificial intelligence has revolutionized the medical field, several studies have been conducted to predict LNM using computational analyses [8, 15-21]. Most of these studies are based on radiologic images [16-18], clinicopathologic features [8, 19, 21], or both [15, 20]. Histological imaging provides a higher density of morphological information for the primary tumor than other clinicopathological parameters or radiological images. Considering that the risk of LNM is determined by the primary tumor, it is necessary to apply advanced computational analysis to identify the intricate and diverse morphological patterns (which cannot be seen by naked human eyes) with hematoxylin and eosin (H&E)-stained slides.

Several studies have predicted LNM using computational pathology [22-27]. However, most previous studies have used histological images of advanced cancers, for an understanding of cancer-related information. There is currently only one study [23] that has reported the prediction of LNM in early-stage cancer with relatively limited cancer-related image information. Moreover, no study has employed deep learning techniques using histological images to predict LNM in gastric cancer. In our previous study, we successfully classified histological subtypes of gastric cancer using deep learning [28]. Considering the histopathological patterns as a significant factor for predicting lymph node metastasis in gastric cancer, we aimed to provide an enhanced treatment guide for EGC patients with LNM. To achieve this goal, we obtained H&E-stained histological images from multiple institutions and extracted morphological features from the slides. Subsequently, machine learning was applied to predict the LNM status.

Materials and methods

Study design

We conducted LNM classification using 1,237 whole slide image (WSI) datasets obtained from five different institutions (HGH, Hanyang University Guri Hospital; KBSMC, Kangbuk Samsung Medical Center; SS, Seoul St. Mary's Hospital; ISH, International St. Mary's Hospital; KUMC, Korea University Medical Center). Our approach consisted of three main steps: tumor segmentation, tumor morphology feature extraction, and lymph node metastasis classification. The workflow of the proposed method is illustrated in **Figure 1**.

Data collection and image preprocessing

Ethical approval was obtained from the ethics committees of the respective institutions (HGH, 2020-09-002; KBSMC, 2021-06-030; SS, KC20RISI0329; ISH, IS21TISI0033; KUMC, 2020AN0385). Representative slide sections of surgical specimen from EGC cases were selected and scanned with Aperio, Mirax, Phillips and Hamamatsu at X40 equivalent magnification (0.25 μm per pixel) and digitalized into svs, mrxs, tiff and ndpi format. ASAP (Automated Slide Analysis Platform) was used for pixel-based annotation of differentiated tumors, undifferentiated tumors, and normal tissues. The annotation was performed by two gastrointestinal pathologists (S.A., S.H.L.) and two general pathologists (Y.-N.S., Y.J.L), and in case of disagreement consensus was reached through discussion.

The WSIs obtained from the five different institutions exhibited color differences owing to batch effects. Therefore, we examined the color distribution of patches generated from the WSIs from each institution, and the results are shown in Supplementary Figure 1. Considering these color differences, we divided the internal datasets for tumor segmentation into HGH, KBSMC, and SS, and the external datasets into ISH and KUMC.

A total of 13,538 patches were generated by the five institutions at a magnification of 10×,

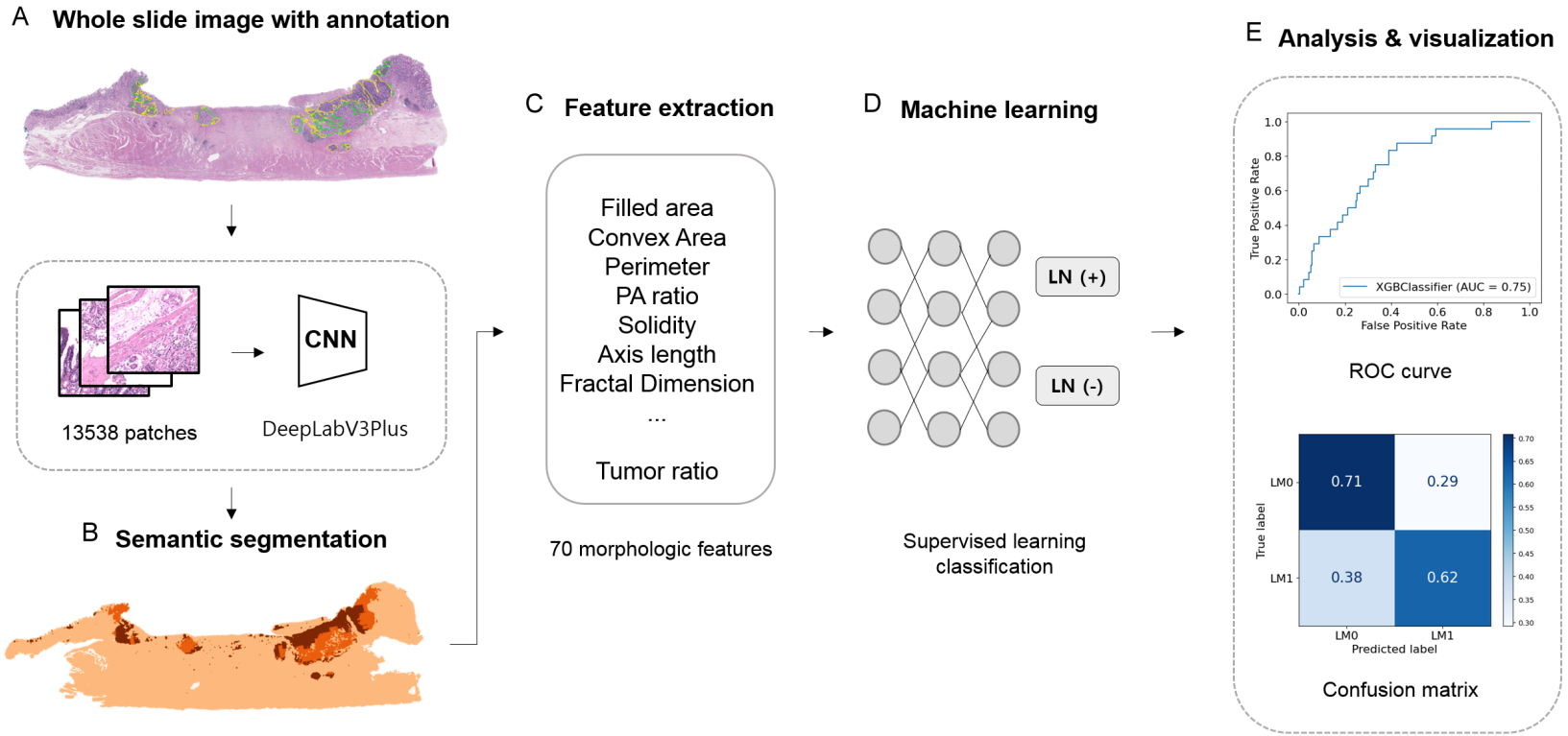


Figure 1. Overview of study design. A. Tumor regions on whole slide image were identified and annotated as regions of interest. B. Semantic segmentation was performed into three types: differentiated tumor, undifferentiated tumor, and normal. C. 70 morphologic features (see Supplementary Materials for details) were computed. D. Morphologic features were processed to predict lymph node metastasis using machine learning. E. Lymph node metastasis prediction performance was visualized by ROC curves and confusion matrix.

Table 1. Patch-wise data scheme trained for tumor subtype segmentation

	HGH	KBSMC	SS	ISH	KUMC	All
Training set	2,692	2,361	2,156	0	0	7,209
Validation set	770	675	616	0	0	2,061
Test set	384	337	308	1,017	2,222	4,268
Total	3,846	3,373	3,080	1,017	2,222	13,538

Table 2. Slide-wise data scheme trained for LNM classifier

	Train set		Test set		Total
	LNM (-)	LNM (+)	LNM (-)	LNM (+)	
HGH	202	29	51	7	289
KBSMC	196	16	50	4	266
SS	161	18	41	4	224
ISH	48	4	13	1	66
KUMC	280	33	71	8	392
Total	887	100	226	24	1,237

with a patch size of 512 × 512 pixels. Details of the patches are summarized in **Table 1**.

Development of multi-class segmentation network for the subtype of gastric cancer

For tumor segmentation, we used SE-ResNeXt101-32x4d as the encoder model in DeepLabV3Plus. The four classes included normal tissues, differentiated tumors, undifferentiated tumors, and background. Due to class imbalance, we employed an imbalance sampler to achieve a 1:1:1 ratio for normal, differentiated, and undifferentiated tumor patches. During training, the batch size was set to 64, and we used Dice loss with a 1e-4 learning rate and the CosineAnnealingLR scheduler. Augmentation techniques were applied for training, including location augmentation (horizontal flip, vertical flip, shift scale rotate), transform augmentation (elastic transform, grid distortion, optical distortion), and color augmentation (random brightness contrast). Before training, normalization was conducted by dividing the patches by the maximum pixel value of 255.

For all 1,237 WSIs, we used a trained tumor segmentation model to infer the tumor masks at the patch-level, and the images were processed using a sliding window approach to obtain slide-level results. For patch-level inference, we compared each patch with the corresponding tissue mask patch, and only performed inferences on patches where the tissue ratio was greater than 0.3. We did not apply morphology operations (e.g., opening or closing) as a postprocessing method to prevent distortion of tumor morphology.

Extraction of interpretable morphologic features

We extracted over 70 morphological features related to differentiated and undifferentiated tumor regions from slide-level masks of the 1,237 WSIs. These features were quantified and analyzed in the total tumor and main tumor regions. The total tumor region includes all the segmentation results, whereas the main tumor region comprises the largest ones. We extracted eight quantitative and five qualitative features related to the characteristics of the tumor area from the main tumor region because we expected that the main tumor regions would exhibit the most distinct qualitative features. Detailed information regarding these features is summarized in Supplementary Table 1.

Development of machine learning model to predict lymph node metastasis using the interpretable morphologic features

We stratified the 1,237 WSIs acquired for LNM classification based on the institution they were collected from, and then split them into training and test datasets in an 8:2 ratio. This ensured that the occurrence of lymph node metastatic cases in each institution was distributed proportionally in both datasets. As a result, the training dataset consisted of 987 WSIs and the test dataset consisted of 250 WSIs. The distribution of slides and lymph node metastasis cases at each institution is shown in **Table 2**.

Results

Robust deep learning network for tumor subtype segmentation on internal and external datasets

The inferences for differentiated/undifferentiated tumors and normal tissues are shown in **Figure 2** (slide-wise) and Supplementary Figure 2 (patch-wise). The performance of the segmentation network at the patch and slide levels is presented in **Table 3**. The mean pixel accuracy of the external cohort, which was not used

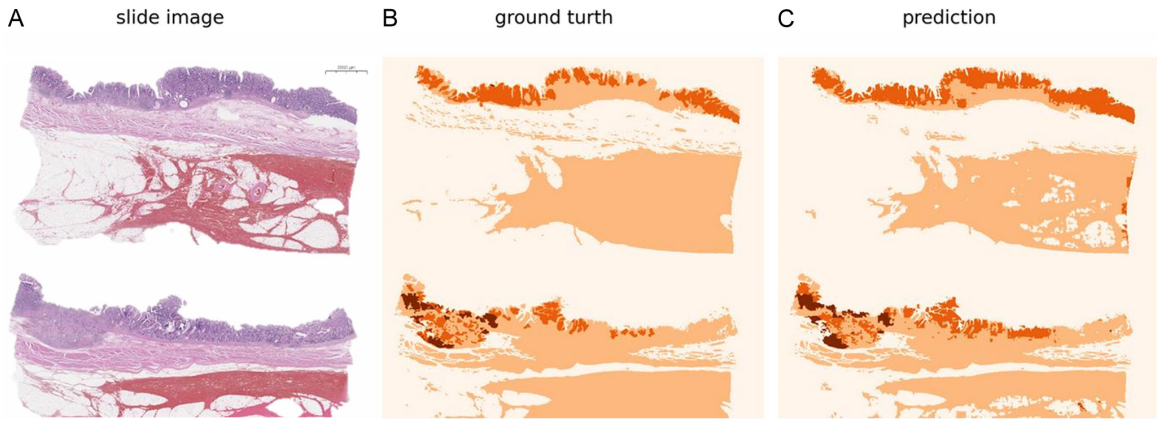


Figure 2. Results of segmentation network for tumor subtype in slide level. H&E-stained slides (A) alongside corresponding visualizations of tissue-type ground truth (B) and prediction (C) in external dataset. Slide regions are classified into three tissue types: differentiated tumor (dark orange), undifferentiated tumor (brown), or normal (light orange).

Table 3. Performance of tumor subtype segmentation across path-wise and slide-wise

Datasets	mIoU	F1-Score	mPA	Recall	Precision
Patch - internal (1,029)	0.77	0.8696	0.9348	0.8694	0.8698
Patch - external (3,239)	0.6551	0.7878	0.8939	0.7875	0.7881
Slide - external (39)	0.6152	0.6701	0.9877	0.7803	0.6663

mPA = mean Pixel Accuracy.

for training, was as high as 0.8939 and 0.9877 at the patch and slide levels, respectively. The performance of the segmentation network for differentiated and undifferentiated tumors is shown in Supplementary Table 2. Overall, the segmentation performance for differentiated tumors was significantly better than that for undifferentiated tumors, likely because of the higher number of patches available for differentiated tumors. The slide-level inference results for both tumor types were low; however, this was because the areas occupied by differentiated and undifferentiated tumors were very small on the entire slide.

Introspection of segmentation network by visualization of the learned representation

To verify that the encoder of the segmentation model was well-trained, the latent vectors of the patch-level test datasets through the encoder were mapped to UMAP (Uniform Manifold approximation and Projection for visualization) [29]. **Figure 3A** shows a scatter plot of UMAP embedding colored according to the tissue type. Representative patches with similar features, according to the encoder, were clustered closely in distinct regions, proving the discriminant performance of the model. Among the undifferentiated tumors, some appeared to be integrated into the differentiated tumor cluster. Those outside of the distribution patches were reviewed by pathologists and confirmed to be differentiated tumors (**Figure 3B**).

Interpretable morphologic features can predict lymph node metastasis in early gastric cancer

The basic characteristics according to each cohort are shown in Supplementary Tables 3 and 4. And **Table 4** shows the results of lymph node metastasis prediction using XGBoost. The model was trained using 70 morphological features obtained from 987 masks and tested on 250 cases. The feature importance plot used in this model is shown in Supplementary Figure 3. Of the 70 morphologic features included in the model, the top 5 ranked features were 'filled area of total tumor', 'area of total tumor', 'equivalent diameter of total tumor', 'perimeter of total tumor', and 'perimeter of undifferentiated' in sequential order. The overall AUC approached 0.75, whereas ISH, SS, and HGH achieved AUC above 0.8. Remarkably, despite not being used

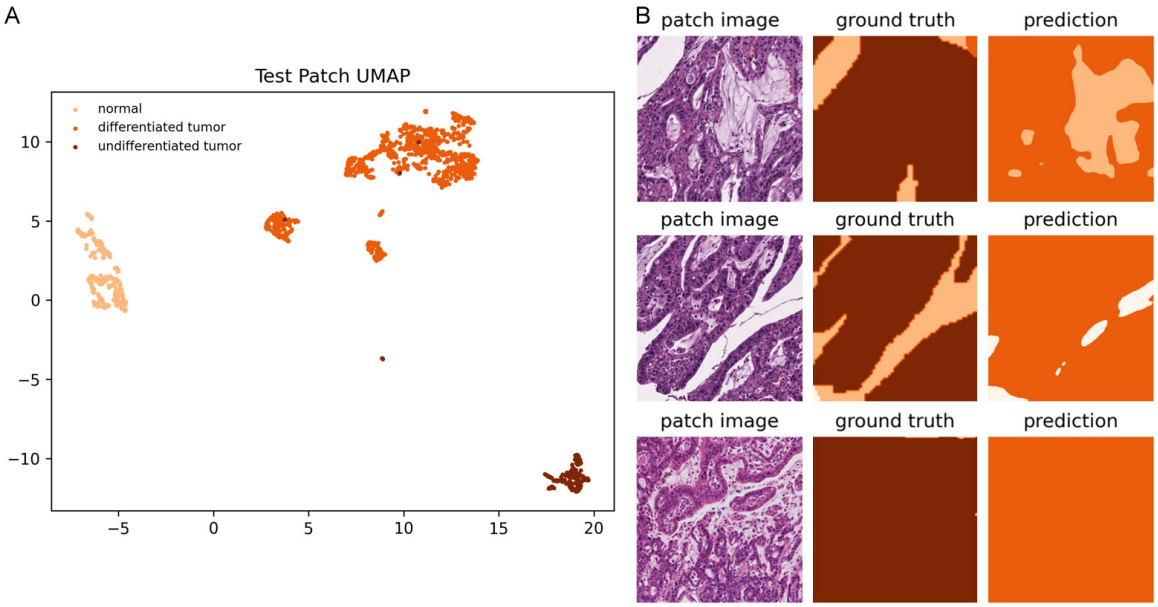


Figure 3. Visualization of representation learning and actual segmentation results for differentiated tumor area of UMAP. A. Uniform Manifold Approximation and Projection (UMAP) visualization showing disentangled representation in differentiated/undifferentiated tumor and normal tissue. B. H&E-stained slides alongside corresponding visualizations of tissue-type ground truth and prediction (differentiated tumor, dark orange; undifferentiated tumor, brown, MMP=1).

Table 4. Performance of XGBoost for predicting lymph node metastasis

	HGH	KBSMC	SS	ISH	KUMC	Total
AUC (± CI)	0.8011 (± 0.0058)	0.6450 (± 0.0202)	0.8293 (± 0.0064)	0.9231 (± nan)	0.7007 (± 0.0050)	0.7487 (± 0.0021)
Accuracy	0.59	0.87	0.73	0.86	0.62	0.7
Weighted Precision	0.87	0.85	0.9	0.95	0.86	0.87
Weighted Recall	0.59	0.87	0.73	0.86	0.62	0.7
Weighted F1-Score	0.66	0.86	0.79	0.89	0.7	0.76

in segmentation training, the ISH showed a notable AUC of 0.92. However, the performance of KBSMC external dataset was slightly lower at approximately 0.65. The ROC curves and confusion matrix for each institution and the overall institution are shown in **Figure 4** and Supplementary Figure 4. The results of the same detailed statistical analyses using SGD and LGB are presented in Supplementary Tables 5 and 6, respectively. The corresponding ROC curves are shown in Supplementary Figures 5 and 6.

Clinicopathologic analysis of false negative results

Nine false-negative slides were obtained from the XGBoost model. Pathologists reviewed the pathology reports and inference masks. In five of the nine false-negative cases, poorly differentiated tumor cells were scattered diffusely as individual cells in the stroma (Supplementary Figure 7). In those cases, although the analysis was performed using a representative section, the segmentation model often predicted suboptimal results. This could be interpreted as a limitation of data labeling in supervised learning because the annotation is not pixel-based. In addition, there were two cases in which the images used for the analysis were not representative. Although the initial pathological diagnosis indicated the depth of SM3 invasion or the presence of lymphovascular invasion, the corresponding area was not included in the image used for analysis (Supplementary Figure 8).

Discussion

Predicting lymph node metastasis status in patients with EGC is very important for determining appropriate treatment. In this study, we

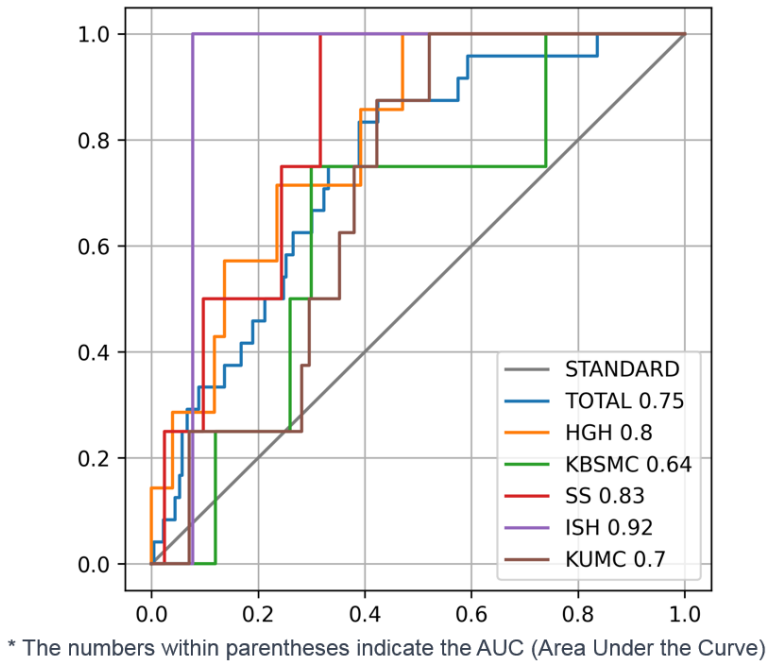


Figure 4. The ROC curve for predicting LNM using XGBoost.

developed an LNM prediction model that is morphologically interpretable using whole-slide images of primary tumors. The model yielded satisfactory predictions with an AUC of 0.8 for the internal dataset. Although its performance was reduced to an AUC of 0.65 in one of the external datasets, it was as high as 0.92 in another external dataset. These variations suggest that the model learns by extracting biologically relevant features from HE images rather than random correlations between artifacts and LNM in the training and test cohorts. Various studies have attempted to predict LNM from primary tumor histology using deep learning algorithms in other cancers including the colorectal [23, 26], bladder [24], prostate [25], breast [30], and melanoma [22]. However, to the best of our knowledge, this is the first study to demonstrate the potential of deep learning to predict LNM status from routine H&E-stained slides of primary EGC.

Current guidelines for the endoscopic treatment of EGC include histological differentiation, tumor size, depth of invasion, and the presence of an ulcer. This indicates that a patient's prognosis and the likelihood of lymph node metastasis, can be influenced by the histological phenotype of the cancer. Several studies have shown that an undifferentiated histological pattern is a risk factor for LNM [31-35]. Based on the assumption that the ratio and distribution pattern of tumor differentiation in the slide images might affect LNM, we annotated tumors into differentiated and undifferentiated subtypes. Subsequently, a multiclass segmentation network was developed using the annotated image, and a LNM prediction model was developed by extracting morphological features from each area of the tumor.

This paper proposes a novel approach for developing a prediction model. Unlike previous studies on other organs that used scoring heat maps for LNM prediction [24-26, 30], we utilized morphological features extracted from differentiated, undifferentiated, and combined tumor areas. Consequently, it was possible to inversely estimate the crucial features used in the prediction model. Specifically, among the 70 morphological features incorporated into the model, more of the features associated with the undifferentiated tumor area (perimeter, number of regions, and filled area) were included in the top 10 features as opposed to the differentiated tumor area. Furthermore, the fact that features representing the quantity of a tumor have been identified as meaningful indicators has biological plausibility considering previous studies showing that undifferentiated histologic type is a risk factor for LNM.

This study was performed in a large multicenter cohort, and the generalizability of the model was confirmed through external validation. Furthermore, the model demonstrated remarkable prediction accuracy (AUC of 0.8), even in the absence of clinical or radiological information other than the histological features of the primary tumor. This is comparable to the AUC value of 0.836 obtained by the modified eCura system [36]. This is important because our method overcomes the limitations of the classical scoring system, which has limited accuracy and consistency for variable measurement.

However, this study has some limitations. Firstly, as a multimodal study, the AUC for predicting LNM differed between cohort (**Figure 4**). Specifically, the ISH cohort exhibited a high AUC of 0.92. Besides differences in tumor size and submucosal invasion among the basic characteristics (Supplementary Table 5), the confusion matrix (Supplementary Figure 4) suggests that the proportion effect may have influenced the results due to the smaller number of true LNM cases in the ISH cohort. Comparing the color distribution (Supplementary Figure 1) between the cohort with a relatively larger number of cases and a high AUC (SS: 0.83) and the cohorts with lower AUCs (KUMC: 0.7, KBSMC: 0.64), it is evident that the SS cohort has a much more favorable color distribution. This indicates that data disparity due to color distribution is a crucial factor affecting LNM prediction results, highlighting the need to overcome data disparity through the pre-training strategy for robust feature extraction [37]. Secondly, although a good accuracy for predicting LNM was achieved using the deep learning model, there is a drawback inherent to supervised learning stemming from the labor-intensive nature of the image annotation process. In addition, it has constraints in identifying new image biomarkers because it relies on predefined targets. To overcome the limitations of supervised learning, automatic feature extraction from images should be attempted using a recent medical foundation model trained on unlabeled data [38, 39]. Although pathological images containing high density information were used, a comprehensive analysis requires multimodal data that integrates AI with various data types, including histopathological images, tabular, and textual information from pathology reports. This comprehensive approach will enhance the model's accuracy and robustness [40, 41]. To achieve this, cross-attention model, which has been shown to effectively integrate multimodal data for improved performance in predictive tasks is necessary [42-44]. Additionally, model lighting and extensive external validation are required to ensure feasibility in clinical practice.

Conclusion

In conclusion, we developed a prediction model for identifying patients who might be at a higher risk of developing LNM in EGC based on digital H&E slides of primary gastric cancer. The model is not subject to intra- or inter-observer variations, which enables its application in various institutions. Ultimately, this model could aid in reducing the overtreatment of patients with EGC who do not need surgery, while ensuring appropriate treatment for those who need surgery.

Acknowledgements

This research was supported by a Korea University grant (project no. 02310961, 30%), a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: RS-2021-KH113146, 40%), and ICT Creative Consilience Program through the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (IITP-2024-2020-0-01819, 30%).

Disclosure of conflict of interest

None.

Address correspondence to: Dr. Sangjeong Ahn, Department of Pathology, Korea University Anam Hospital, College of Medicine, Korea University, 145, Anam-ro, Seongbuk-gu, Seoul 02841, South Korea. E-mail: vanitasahn@gmail.com; Dr. Sung Hak Lee, Department of Hospital Pathology, Seoul St. Mary's Hospital, College of Medicine, The Catholic University, 222, Banpo-daero, Seocho-gu, Seoul 06591, South Korea. E-mail: hakjjang@catholic.ac.kr

References

[1] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA and Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018; 68: 394-424.

[2] Smyth EC, Nilsson M, Grabsch HI, van Grieken NC and Lordick F. Gastric cancer. Lancet 2020; 396: 635-648.

[3] Jun JK, Choi KS, Lee HY, Suh M, Park B, Song SH, Jung KW, Lee CW, Choi IJ, Park EC and Lee D. Effectiveness of the Korean National Cancer Screening Program in reducing gastric cancer mortality. Gastroenterology 2017; 152: 1319-1328, e1317.

[4] Miyamoto A, Kuriyama S, Nishino Y, Tsubono Y, Nakaya N, Ohmori K, Kurashima K, Shibuya D

and Tsuji I. Lower risk of death from gastric cancer among participants of gastric cancer screening in Japan: a population-based cohort study. Prev Med 2007; 44: 12-19.
[5] de Martel C, Forman D and Plummer M. Gastric cancer: epidemiology and risk factors. Gastroenterol Clin North Am 2013; 42: 219-240.
[6] Hatta W, Gotoda T, Koike T and Masamune A. History and future perspectives in Japanese guidelines for endoscopic resection of early gastric cancer. Dig Endosc 2020; 32: 180-190.
[7] Saka M, Katai H, Fukagawa T, Nijjar R and Sano T. Recurrence in early gastric cancer with lymph node metastasis. Gastric Cancer 2008; 11: 214-218.
[8] Zhu H, Wang G, Zheng J, Zhu H, Huang J, Luo E, Hu X, Wei Y, Wang C, Xu A and He X. Preoperative prediction for lymph node metastasis in early gastric cancer by interpretable machine learning models: a multicenter study. Surgery 2022; 171: 1543-1551.
[9] Roviello F, Rossi S, Marrelli D, Pedrazzani C, Corso G, Vindigni C, Morgagni P, Saragoni L, de Manzoni G and Tomezzoli A. Number of lymph node metastases and its prognostic significance in early gastric cancer: a multicenter Italian study. J Surg Oncol 2006; 94: 275-280; discussion 274.
[10] Chen J, Zhao G and Wang Y. Analysis of lymph node metastasis in early gastric cancer: a single institutional experience from China. World J Surg Oncol 2020; 18: 57.
[11] Kwee RM and Kwee TC. Imaging in assessing lymph node status in gastric cancer. Gastric Cancer 2009; 12: 6-22.
[12] Kim HJ, Kim AY, Oh ST, Kim JS, Kim KW, Kim PN, Lee MG and Ha HK. Gastric cancer staging at multi-detector row CT gastrography: comparison of transverse and volumetric CT scanning. Radiology 2005; 236: 879-885.
[13] Chen J, Cheong JH, Yun MJ, Kim J, Lim JS, Hyung WJ and Noh SH. Improvement in preoperative staging of gastric adenocarcinoma with positron emission tomography. Cancer 2005; 103: 2383-2390.
[14] Seevaratnam R, Cardoso R, McGregor C, Lourenco L, Mahar A, Sutradhar R, Law C, Paszat L and Coburn N. How useful is preoperative imaging for tumor, node, metastasis (TNM) staging of gastric cancer? A meta-analysis. Gastric Cancer 2012; 15 Suppl 1: S3-18.
[15] Li Y, Xie F, Xiong Q, Lei H and Feng P. Machine learning for lymph node metastasis prediction of in patients with gastric cancer: a systematic review and meta-analysis. Front Oncol 2022; 12: 946038.
[16] Dong D, Fang MJ, Tang L, Shan XH, Gao JB, Giganti F, Wang RP, Chen X, Wang XX, Palumbo D, Fu J, Li WC, Li J, Zhong LZ, De Cobelli F, Ji JF, Liu ZY and Tian J. Deep learning radiomic nomogram can predict the number of lymph node metastasis in locally advanced gastric cancer: an international multicenter study. Ann Oncol 2020; 31: 912-920.
[17] Li J, Dong D, Fang M, Wang R, Tian J, Li H and Gao J. Dual-energy CT-based deep learning radiomics can improve lymph node metastasis risk prediction for gastric cancer. Eur Radiol 2020; 30: 2324-2333.
[18] Jin C, Jiang Y, Yu H, Wang W, Li B, Chen C, Yuan Q, Hu Y, Xu Y, Zhou Z, Li G and Li R. Deep learning analysis of the primary tumour and the prediction of lymph node metastases in gastric cancer. Br J Surg 2021; 108: 542-549.
[19] Zhou CM, Wang Y, Ye HT, Yan S, Ji M, Liu P and Yang JJ. Machine learning predicts lymph node metastasis of poorly differentiated-type intramucosal gastric cancer. Sci Rep 2021; 11: 1300.
[20] Wei X, Yan XJ, Guo YY, Zhang J, Wang GR, Fayyaz A and Yu J. Machine learning-based gray-level co-occurrence matrix signature for predicting lymph node metastasis in undifferentiated-type early gastric cancer. World J Gastroenterol 2022; 28: 5338-5350.
[21] Tian H, Ning Z, Zong Z, Liu J, Hu C, Ying H and Li H. Application of machine learning algorithms to predict lymph node metastasis in early gastric cancer. Front Med (Lausanne) 2022; 8: 759013.
[22] Brinker TJ, Kiehl L, Schmitt M, Jutzi TB, Krieghoff-Henning EI, Krahl D, Kutzner H, Gholam P, Haferkamp S, Klode J, Schadendorf D, Hekler A, Frohling S, Kather JN, Haggenmuller S, von Kalle C, Heppt M, Hilke F, Ghoreschi K, Tiemann M, Wehkamp U, Hauschild A, Weichenthal M and Utikal JS. Deep learning approach to predict sentinel lymph node status directly from routine histology of primary melanoma tumours. Eur J Cancer 2021; 154: 227-234.
[23] Brockmoeller S, Echle A, Ghaffari Laleh N, Eiholm S, Malmstrom ML, Plato Kuhlmann T, Levic K, Grabsch HI, West NP, Saldanha OL, Kouvidi K, Bono A, Heij LR, Brinker TJ, Gogenur I, Quirke P and Kather JN. Deep learning identifies inflamed fat as a risk factor for lymph node metastasis in early colorectal cancer. J Pathol 2022; 256: 269-281.
[24] Harmon SA, Sanford TH, Brown GT, Yang C, Mehralivand S, Jacob JM, Valera VA, Shih JH, Agarwal PK, Choyke PL and Turkbey B. Multiresolution application of artificial intelligence in digital pathology for prediction of positive lymph nodes from primary tumors in bladder cancer. JCO Clin Cancer Inform 2020; 4: 367-382.
[25] Wessels F, Schmitt M, Krieghoff-Henning E, Jutzi T, Worst TS, Waldbillig F, Neuberger M, Maron RC, Steeg M, Gaiser T, Hekler A, Utikal

JS, von Kalle C, Frohling S, Michel MS, Nuhn P and Brinker TJ. Deep learning approach to predict lymph node metastasis directly from primary tumour histology in prostate cancer. BJU Int 2021; 128: 352-360.

[26] Kiehl L, Kuntz S, Hohn J, Jutzi T, Krieghoff-Henning E, Kather JN, Holland-Letz T, Kopp-Schneider A, Chang-Claude J, Brobeil A, von Kalle C, Frohling S, Alwers E, Brenner H, Hoffmeister M and Brinker TJ. Deep learning can predict lymph node status directly from histology in colorectal cancer. Eur J Cancer 2021; 157: 464-473.

[27] Chen S, Xiang J, Wang X, Zhang J, Yang S, Yang W, Zheng J and Han X. Deep learning-based pathology signature could reveal lymph node status and act as a novel prognostic marker across multiple cancer types. Br J Cancer 2023; 129: 46-53.

[28] Jang HJ, Song IH and Lee SH. Deep learning for automatic subclassification of gastric carcinoma using whole-slide histopathology images. Cancers (Basel) 2021; 13: 3811.

[29] McInnes L, Healy J, Saul N and Großberger L. UMAP: uniform manifold approximation and projection. J Open Source Softw 2018; 3: 861.

[30] Xu F, Zhu C, Tang W, Wang Y, Zhang Y, Li J, Jiang H, Shi Z, Liu J and Jin M. Predicting axillary lymph node metastasis in early breast cancer using deep learning on primary tumor biopsy slides. Front Oncol 2021; 11: 759007.

[31] Miyahara K, Hatta W, Nakagawa M, Oyama T, Kawata N, Takahashi A, Yoshifuku Y, Hoteya S, Hirano M, Esaki M, Matsuda M, Ohnita K, Shimoda R, Yoshida M, Dohi O, Takada J, Tanaka K, Yamada S, Tsuji T, Ito H, Aoyagi H and Shimosegawa T. The role of an undifferentiated component in submucosal invasion and submucosal invasion depth after endoscopic submucosal dissection for early gastric cancer. Digestion 2018; 98: 161-168.

[32] Choi KK, Bae JM, Kim SM, Sohn TS, Noh JH, Lee JH, Choi MG and Kim S. The risk of lymph node metastases in 3951 surgically resected mucosal gastric cancers: implications for endoscopic resection. Gastrointest Endosc 2016; 83: 896-901.

[33] Yamao T, Shirao K, Ono H, Kondo H, Saito D, Yamaguchi H, Sasako M, Sano T, Ochiai A and Yoshida S. Risk factors for lymph node metastasis from intramucosal gastric carcinoma. Cancer 1996; 77: 602-606.

[34] Gotoda T, Yanagisawa A, Sasako M, Ono H, Nakanishi Y, Shimoda T and Kato Y. Incidence of lymph node metastasis from early gastric cancer: estimation with a large number of cases at two large centers. Gastric Cancer 2000; 3: 219-225.

[35] Chung JW, Jung HY, Choi KD, Song HJ, Lee GH, Jang SJ, Park YS, Yook JH, Oh ST, Kim BS and Kim JH. Extended indication of endoscopic resection for mucosal early gastric cancer: analysis of a single center experience. J Gastroenterol Hepatol 2011; 26: 884-887.

[36] Nagao K, Ebi M, Shimura T, Yamada T, Hirata Y, Iwai T, Ozeki T, Ohashi W, Sugiyama T, Yamaguchi Y, Adachi K, Izawa S, Funaki Y, Ogasawara N, Sasaki M, Kataoka H and Kasugai K. The modified eCura system for identifying high-risk lymph node metastasis in patients with early gastric cancer resected by endoscopic submucosal dissection. Gastroenterology Insights 2022; 13: 60-67.

[37] Wölflein G, Ferber D, Meneghetti AR, El Nahhas OS, Truhn D, Carrero ZI, Harrison DJ, Arandjelović O and Kather JN. A good feature extractor is all you need for weakly supervised learning in histopathology. arXiv preprint arXiv:2311.11772 2023.

[38] Huang Z, Bianchi F, Yuksekgonul M, Montine T and Zou J. Leveraging medical Twitter to build a visual-language foundation model for pathology AI. bioRxiv 2023; 2023.2003.2029.534834.

[39] Moor M, Banerjee O, Abad ZSH, Krumholz HM, Leskovec J, Topol EJ and Rajpurkar P. Foundation models for generalist medical artificial intelligence. Nature 2023; 616: 259-265.

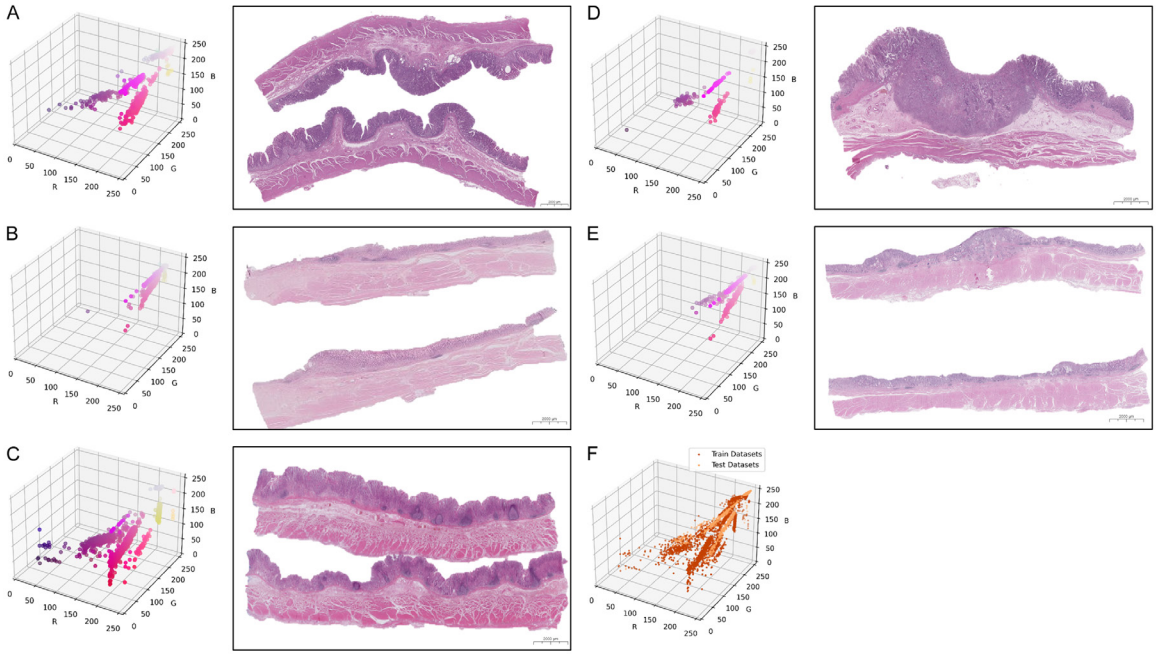
[40] Lipkova J, Chen RJ, Chen B, Lu MY, Barbieri M, Shao D, Vaidya AJ, Chen C, Zhuang L, Williamson DFK, Shaban M, Chen TY and Mahmood F. Artificial intelligence for multimodal data integration in oncology. Cancer Cell 2022; 40: 1095-1110.

[41] Acosta JN, Falcone GJ, Rajpurkar P and Topol EJ. Multimodal biomedical AI. Nat Med 2022; 28: 1773-1784.

[42] Wei X, Zhang T, Li Y, Zhang Y and Wu F. Multi-modality cross attention network for image and sentence matching. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020. pp. 10941-10950.

[43] Huang G, Wu C, Li M, Chang X, Chen L, Sun Y, Zhao S, Liang X and Lin L. Predicting genetic mutation from whole slide images via biomedical-linguistic knowledge enhanced multi-label classification. arXiv preprint arXiv:2406.02990 2024.

[44] Tan H and Bansal M. Lxmert: learning cross-modality encoder representations from transformers. arXiv preprint arXiv:1908.07490 2019.

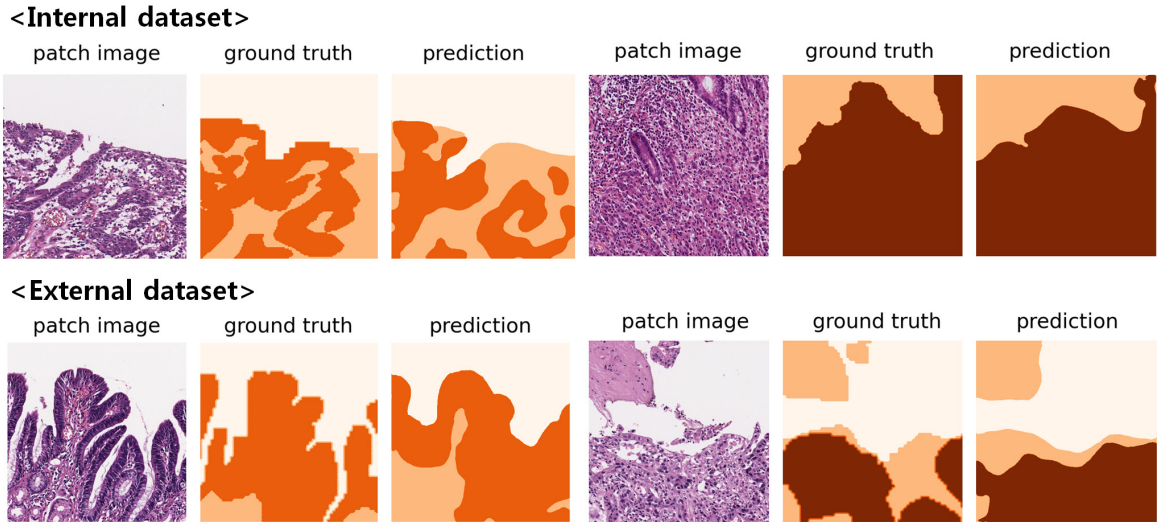


Supplementary Figure 1. Color distribution analysis graph and representative matched H&E image in (A) HGH, (B) KBSMC, (C) SS, (D) ISH, and (E) KUMC and (F) total cohort color distribution.

Supplementary Table 1. Lists of interpretable morphologic features

No.	Feature	Descri ption
1	tot_tumor1_ratio	Proportion of tumor1 areas in total tumor areas
2	tot_tumor2_ratio	Proportion of tumor2 areas in total tumor areas
3	main_ tumor1_ratio	Proportion of tumor1 areas in the 5 largest tumor areas
4	main_tumor2_ratio	Proportion of tumor2 areas in the 5 largest tumor areas
5	tot_all num_region	Number of total tumor regions
6	tot_all_area	Area of total tumor
7	tot_all_convex_area	Convex area of total tumor
8	tot_all filled_area	Filled area of total tumor
9	tot_all_perimeter	Perimeter of total tumor
10	tot_all_equiv_diameter	Equiv diameter of total tumor
11	tot_all_euler_number	Euler number of total tumor
12	tot_all_mj_axis_length	Major xis length of total tumor
13	tot_all_mi_axis_length	Minor xis length of total tumor
14	tot_tumor1_num_region	Number of tumor1 regions
15	tot_tumor1_area	Area of tumor1
16	tot_tumor1_convex_area	Convex area of tumor1
17	tot_tumor1_flled_area	Filled area of tumor1
18	tot_tumor1_perimeter	Perimeter of tumor1
19	tot_tumor1_equiv_diameter	Equiv diameter of tumor1
20	tot_tumor1_euler_number	Euler number of tumor1
21	tot_tumorl_mj_axis_length	Major axis length of tumorl
22	tot_tumor1_mi_axis_length	Minor axis length of tumor1
23	tot_tumor2_num_region	Number of tumor2 regions
24	tot_tumor2_area	Area of tumor2

25	tot_tumor2_convex_area	Convex area of tumor2
26	tot_tumor2_filled_area	Filled area of tumor2
27	tot_tumor2_perimeter	Perimeter of tumor2
28	tot_tumor2_equiv_diameter	Equiv diameter of tumor2
29	tot_tumor2_euler_number	Euler number of tumor2
30	tot_tumor2_mj_axis_length	Major axis leng th of tumor2
31	tot_tumor2_mi_axis_length	Minor axis length of tumor2
32	main_all_area	Area of the 5 biggest tumor
33	main_all_convex_area	Convex area of the 5 largest tumor
34	main_all_flleld_area	Filled area of the 5 largest tumor
35	main_all_perimeter	Alll perimeter of the 5 largest tumor
36	main_all_equiv_diameter	Equiv diameter of the 5 largest tumor
37	main_all_euler_number	Euler number of the 5 largest tumor
38	main_all_mj_axis_length	Major axis length of the 5 largest tumor
39	main_all_mi_axis_length	Major axis length of the 5 largest tumor
40	main_all_eccentricity	All eccentricity of the 5 largest tumor
41	main_all_extent	Al| extent of the 5 largest tumor
42	main_all_solidity	All solidity of the 5 largest tumor
43	main_all_pa_ratio	Pa ratio of the 5 largest tumor
44	main_all_fractal_dimension	Fractal dimension of 5 largest tumor
45	main_tumor1_area	Area of the the 3 largest tumor1
46	main_tumor1_convex_area	Convex area of the the 3 largest tumor1
47	main_tumor1_filled_area	Filled area of the the 3 largest tumor1
48	main_tumor1_perimeter	Perimeter of the the 3 largest tumor1
49	main_tumor1_equiv_diameter	Equiv diameter of the 3 largest tumor1
50	main_tumor1_euler_number	Euler number of the 3 largest tumor1
51	main_tumor1_mj_axis_length	Major axis length of the 3 largest tumor1
52	main_tumor1_mi_axis_length	Minor axis length of the 3 largest tumor1
53	main_tumor1_all_eccentricity	All eccentricity of the 3 largest tumor1
54	main_tumor1_all_extent	All extent of the 3 largest tumor1
55	main_tumor1_all_solidity	All solidity of the 3 largest tumor1
56	main_tumor1_all_pa_ratio	All pa ratio of the 3 largest tumor1
57	main_tumor1_all_fractal_dimension	All fractal dimension of the 3 largest tumor1
58	main_tumor2_area	Area of the 3 largest tumor2
59	main_tumor2_convex_area	Convex area of the 3 largest tumor2
60	main_tumor2_filled_area	Filled area of the 3 largest tumor2
61	main_tumor2_perimeter	Perimeter of the 3 largest tumor2
62	main_tumor2_equiv_diameter	Equiv diameter of the 3 largest tumor2
63	main_tumor2_euler_number	Euler number of the the 3 largest tumor2
64	main_tumor2_mj_axis_length	Major axis length of the 3 largest tumor2
65	main_tumor2_mi_axis_length	Minor axis length of the 3 largest tumor2
66	main_tumor2_eccentricity	All eccentricity of the 3 largest tumor2
67	main_tumor2_extent	All extent of the 3 largest tumor2
68	main_tumor2_solidity	All solidity of the 3 largest tumor2
69	main_tumor2_pa_ratio	All pa ratio of the 3 largest tumor2
70	main_tumor2_fractal_dimension	All fractal dimension of the 3 largest tumor2.



Supplementary Figure 2. Results of segmentation network for tumor subtype in patch level. H&E-stained slides alongside corresponding visualizations of tissue-type ground truth and prediction in internal and external dataset (MMP=1). Slide regions are classified into three tissue types: differentiated tumor (dark orange), undifferentiated tumor (brown), or normal (light orange).

Supplementary Table 2. Performance of tumor subtype segmentation

Datasets	IoU	F1-Score	PA	Recall	Precision
Differentiated					
patch - internal (1,029)	0.7999	0.8881	0.9339	0.8842	0.8933
patch - external (3,239)	0.5659	0.6764	0.8526	0.71	0.683
slide -external (39)	0.296	0.3718	0.9909	0.5694	0.3419
Undifferentiated					
patch - internal (1029)	0.5595	0.6976	0.9468	0.7442	0.726
patch - external (3,239)	0.2952	0.3813	0.9234	0.4481	0.5309
slide -external (39)	0.2707	0.365	0.9944	0.6261	0.3571

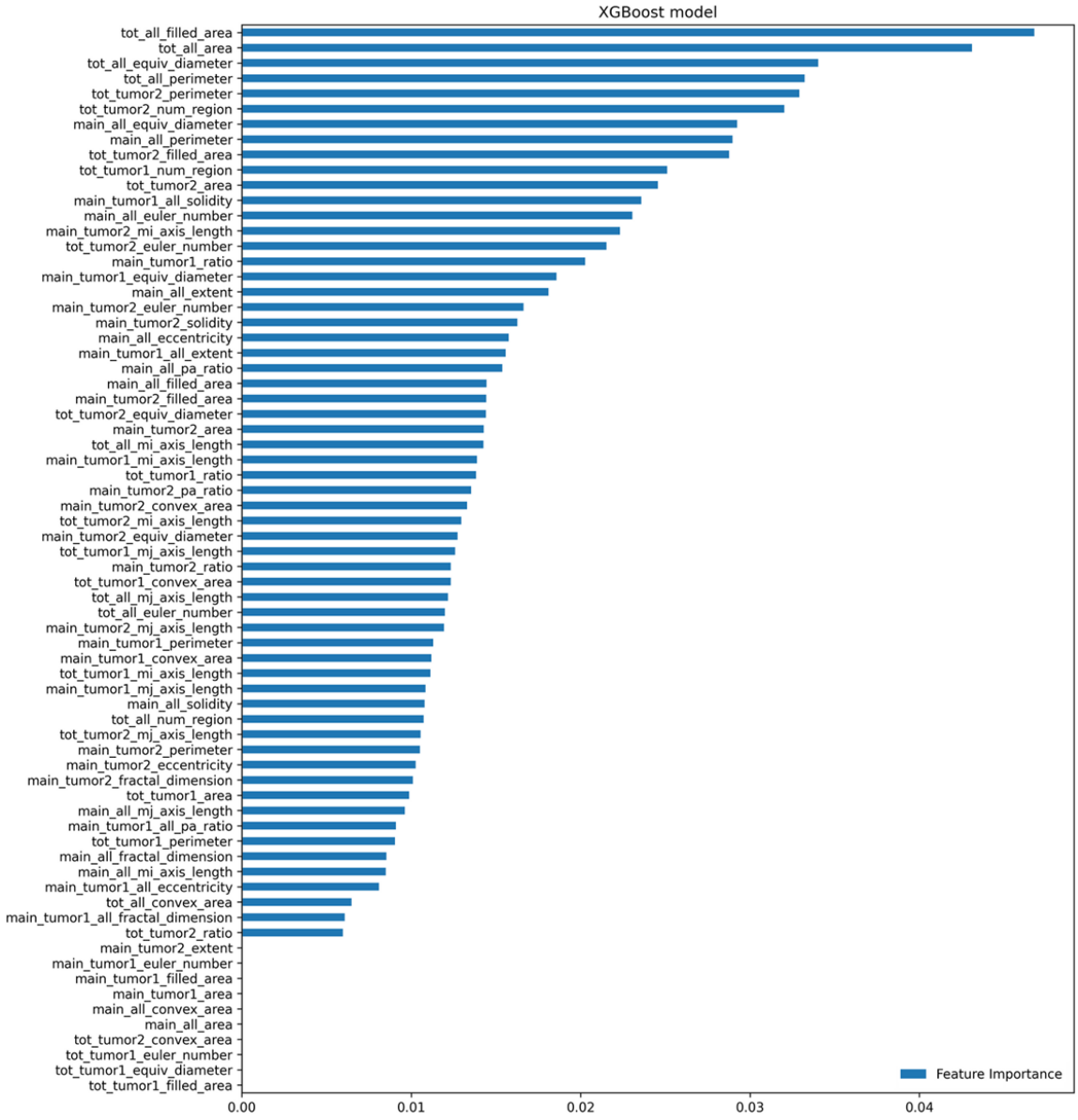
PA = Pixel Accuracy.

Supplementary Table 3. Basic characteristics according to cohort

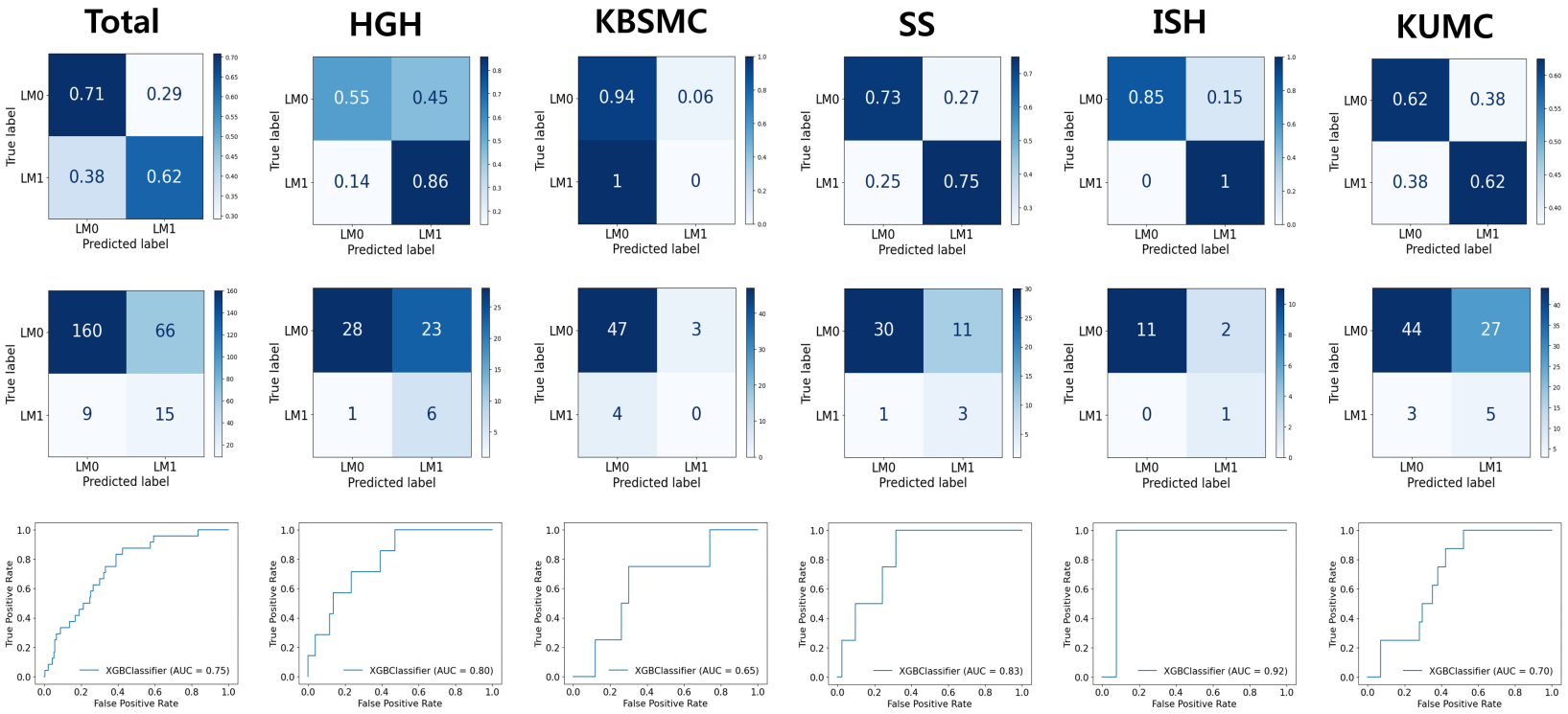
Characteristics	HGH (289)	ISH (66)	KUMC (392)	KBSMC (266)	SS (224)	*p-value*
Size (mm)	27.6	23	30.8	33.4	29.5	0.00035
Submucosal invasion						0.00003
Present	123 (47.1%)	15 (22.7%)	203 (51.8%)	102 (38.3%)	103 (46.0%)	
Absent	138 (52.9%)	51 (77.3%)	189 (48.2%)	164 (61.7%)	121 (54.0%)	
Missing	28	0	0	0	0	
Perineural invasion						0.14286
Present	11 (4.2%)	3 (4.5%)	5 (1.3%)	6 (2.3%)	5 (2.2%)	
Absent	250 (95.8%)	63 (95.5%)	387 (98.7%)	260 (97.7%)	219 (97.8%)	
Missing	28	0	0	0	0	
Lymphovascular invasion						0.0004
Present	62 (23.8%)	12 (18.2%)	53 (13.5%)	28 (10.5%)	42 (18.8%)	
Absent	199 (76.2%)	54 (81.8%)	339 (96.5%)	238 (89.5%)	182 (81.2%)	
Missing	28	0	0	0	0	
Lymph node metastasis						0.36964
Present	36 (12.5%)	5 (7.6%)	41 (10.5%)	20 (7.5%)	22 (9.8%)	
Absent	253 (87.5%)	61 (92.4%)	351 (89.5%)	246 (92.5%)	202 (90.2%)	
Missing	0	0	0	0	0	

Supplementary Table 4. Basic characteristics according to cohort (ISH vs others)

Characteristics	ISH (66)	Data sets other than ISH (948)	*p-value*
Size (mm)	23	30.4	0.00282
Submucosal invasion			0.00018
Present	15 (22.7%)	531 (46.5%)	
Absent	51 (77.3%)	612 (53.5%)	
Missing	0	28	
Perineural invasion			0.22225
Present	3 (4.5%)	27 (2.4%)	
Absent	63 (95.5%)	1116 (97.6%)	
Missing	0	28	
Lymphovascular invasion			0.61096
Present	12 (18.2%)	185 (16.2%)	
Absent	54 (81.8%)	958 (83.8%)	
Missing	0	28	
Lymph node metastasis			0.67305
Present	5 (7.6%)	119 (10.2%)	
Absent	61 (92.4%)	1052 (89.8%)	
Missing	0	0	



Supplementary Figure 3. Feature importance plot in XGBoost.



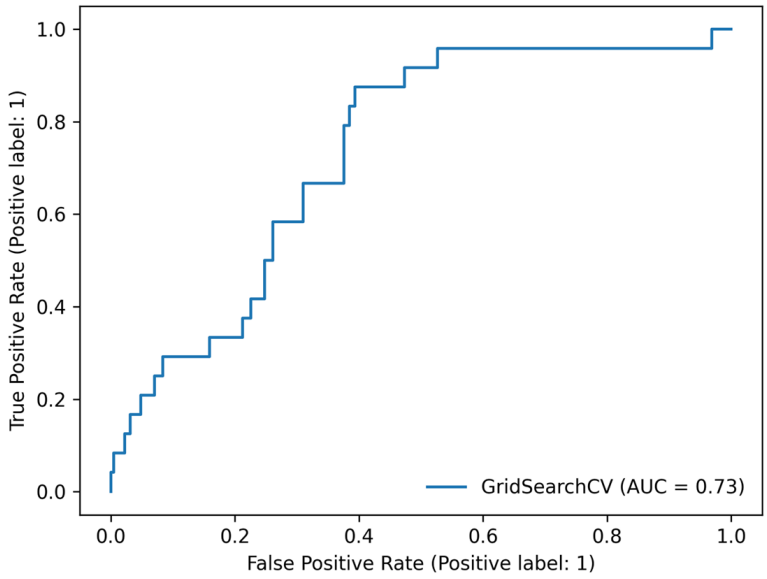
Supplementary Figure 4. Confusion matrix and ROC curve for predicting LNM using XGBoost.

Supplementary Table 5. Performance of SGD for predicting lymph node metastasis

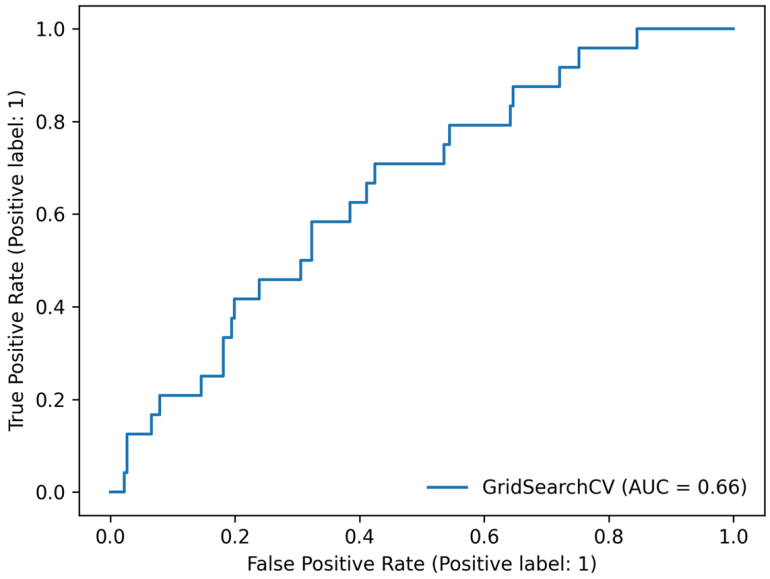
	HGH	KBSMC	SS	ISH	KUMC	Total
AUC (± CI)	0.7731 (±0.0083)	0.58 (±0.0283)	0.82 (±0.0059)	0.8461 (±nan)	0.7042 (±0.0050)	0.7345 (±0.0023)
Accuracy	0.6	0.78	0.64	0.79	0.58	0.65
Weighted Precision	0.85	0.87	0.93	0.95	0.85	0.87
Weighted Recall	0.6	0.78	0.64	0.79	0.58	0.65
Weighted F1-Score	0.67	0.82	0.72	0.84	0.66	0.72

Supplementary Table 6. Performance of LGB for predicting lymph node metastasis

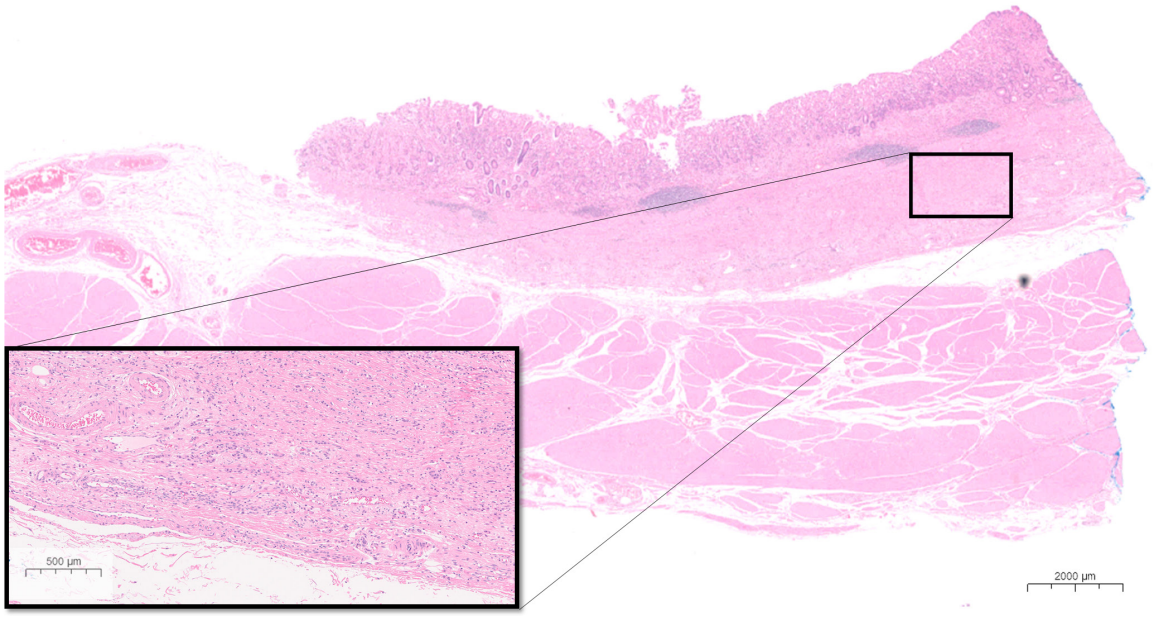
	HGH	KBSMC	SS	ISH	KUMC	Total
AUC (± CI)	0.6639 (± 0.0111)	0.69 (± 0.0122)	0.7317 (± 0.0175)	0.8462 (± nan)	0.6056 (±0.0114)	0.6574 (± 0.0030)
Accuracy	0.53	0.69	0.51	0.71	0.59	0.59
Weighted Precision	0.87	0.86	0.88	0.94	0.85	0.87
Weighted Recall	0.53	0.69	0.51	0.71	0.59	0.59
Weighted F1-Score	0.61	0.76	0.61	0.78	0.67	0.67



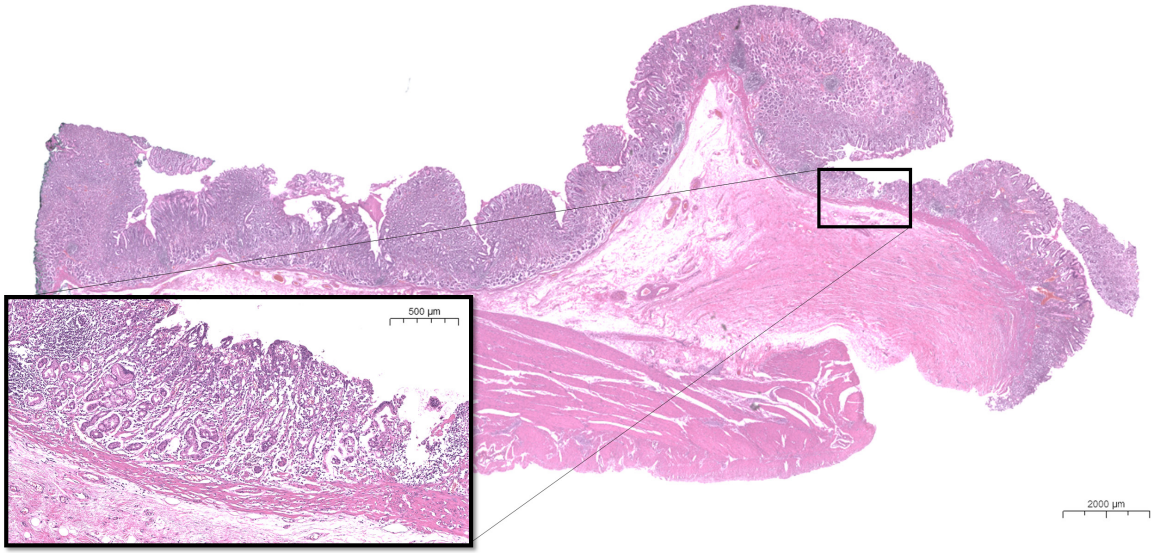
Supplementary Figure 5. The ROC curve for predicting LNM using SGD.



Supplementary Figure 6. The ROC curve for predicting LNM using LGB.



Supplementary Figure 7. Representative image of a slide that the poorly differentiated tumor component is scattered diffusely as individual cells in the stroma (H&E, ×10, ×100, inlet).



Supplementary Figure 8. Representative image of a slide that inadequetely reflect the original diagnosis of SM3 invasion, as the tumor in the scanned slide is confined to the mucosa (H&E, ×10, ×100, inlet).