



OPEN

Deep learning model to predict Epstein–Barr virus associated gastric cancer in histology

Yejin Jeong^{1,13}, Cristina Eunbee Cho^{2,13}, Ji-Eon Kim^{3,13}, Jonghyun Lee⁴, Namkug Kim², Woon Yong Jung⁵, Joohon Sung¹, Ju Han Kim⁶, Yoo Jin Lee⁷, Jiyeon Jung⁸, Juyeon Pyo⁹, Jisun Song¹⁰, Jihwan Park¹¹, Kyoung Min Moon¹²✉ & Sangjeong Ahn^{6,7}✉

The detection of Epstein–Barr virus (EBV) in gastric cancer patients is crucial for clinical decision making, as it is related with specific treatment responses and prognoses. Despite its importance, the limited medical resources preclude universal EBV testing. Herein, we propose a deep learning-based EBV prediction method from H&E-stained whole-slide images (WSI). Our model was developed using 319 H&E stained WSI (26 EBV positive; TCGA dataset) from the Cancer Genome Atlas, and 108 WSI (8 EBV positive; ISH dataset) from an independent institution. Our deep learning model, EBVNet consists of two sequential components: a tumor classifier and an EBV classifier. We visualized the learned representation by the classifiers using UMAP. We externally validated the model using 60 additional WSI (7 being EBV positive; HGH dataset). We compared the model's performance with those of four pathologists. EBVNet achieved an AUPRC of 0.65, whereas the four pathologists yielded a mean AUPRC of 0.41. Moreover, EBVNet achieved a negative predictive value, sensitivity, specificity, precision, and F1-score of 0.98, 0.86, 0.92, 0.60, and 0.71, respectively. Our proposed model is expected to contribute to prescreen patients for confirmatory testing, potentially to save test-related cost and labor.

The Epstein–Barr virus (EBV) is present in approximately 10% of gastric cancer patients worldwide¹, with the most prevalent among EBV-attributed malignancies^{2,3}. The presence of EBV has been recognized as a potential biomarker for precision oncology in gastric cancers^{4–8}. EBV-associated gastric cancers exhibit characteristic genetic and epigenetic alteration, which has multipronged effects on their unique phenotypes^{9,10}. EBV-positive gastric cancer has high immunogenicity¹¹ with overexpression of *PD-L1* and *PD-L2*^{12,13}, clinically approved biomarkers for immune checkpoint inhibitors^{14,15}. Deregulated immune response genes in this subtype of tumor affect the tumor immune microenvironment¹⁶, leading to a unique spatial arrangement of tumor cells within exuberant lymphoid stroma: so-called “lymphoepithelioma-like carcinoma”¹⁷. Prominent tumor-infiltrating lymphocytes, the characteristic morphologic feature in EBV-positive gastric cancer, can be a surrogate indicator of tumor behavior and prognosis^{17–19}, associated with a lower frequency of lymph node metastasis^{5–8,20,21}.

¹Genome & Health Data Lab, School of Public Health, Seoul National University, Seoul, Republic of Korea. ²Department of Convergence Medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea. ³Wonkwang University Medical Research Convergence Center, Wonkwang University Hospital, Iksan, Republic of Korea. ⁴Department of Medical and Digital Engineering, Hanyang University College of Engineering, Seoul, Republic of Korea. ⁵Department of Pathology, Hanyang University Guri Hospital, Hanyang University College of Medicine, Guri, Republic of Korea. ⁶Division of Biomedical Informatics, Seoul National University College of Medicine, Seoul National University Biomedical Informatics (SNUBI), Seoul, Republic of Korea. ⁷Department of Pathology, Korea University Anam Hospital, Korea University College of Medicine, 73 Goryeodae-ro, Seongbuk-gu, Seoul 02841, Republic of Korea. ⁸Department of Pathology, Kangnam Sacred Heart Hospital, College of Medicine, Hallym University, Seoul, Republic of Korea. ⁹Department of Pathology, International St. Mary's Hospital, Catholic Kwandong University College of Medicine, Incheon, Republic of Korea. ¹⁰Department of Pathology, Ewha Womans University Seoul Hospital, Ewha Womans University College of Medicine, Seoul, Republic of Korea. ¹¹School of Software Convergence, College of Software Convergence, Dankook University, Yongin-si, Republic of Korea. ¹²Department of Pulmonary, Allergy, and Critical Care Medicine, Gangneung Asan Hospital, College of Medicine, University of Ulsan, 38, Bangdong-gil, Sacheon-myeon, Gangneung-si 25440, Gangwon-do, Republic of Korea. ¹³These authors contributed equally: Yejin Jeong, Cristina Eunbee Cho and Ji-Eon Kim. ✉email: pulmogicu@ulsan.ac.kr; vanitasahn@gmail.com

With a low risk of lymph node metastasis in this subtype, early gastric cancers associated with EBV have been proposed as candidates for local excision regardless of the depth of the submucosal invasion^{5–8,20,21}. Therefore, a revised criteria was proposed for endoscopic resection in patients with EBV-positive early gastric cancer^{6,8,21}. Taken together, the identification of EBV status in gastric cancer, which links to a treatment-relevant phenotype, is important to provide effective therapeutic options and strategies²².

Manual microscopic inspection of H&E histology may be used to predict EBV status in gastric cancer, and may therefore have a role in screening EBV status for further confirmatory test. However, morphologic assessment has a disadvantage of low inter- and intra-rater agreement. A paramount diagnostic method for identifying a tumor as “EBV-positive” is the presence of EBV-encoded RNA in situ hybridization (EBER-ISH)^{11,23}. However, the limitations on medical resources preclude universal testing of EBV status. Therefore, the development of a widely accessible and cost-effective tool for the EBV testing is essential.

As an unprecedented breakthrough in artificial intelligence technology, histology-based deep learning approaches are expected to identify novel biomarkers for oncology practice with a precision beyond human performance²⁴. Prior studies have proven to facilitate learning of morphologic feature representation, correlating to molecular alterations, from digitized whole-slide images (WSI). They have inferred genetic traits including EBV status, actionable driving mutation, microsatellite instability, gene signature, and molecular tumor subtypes, in various malignant images (Supplementary Table S1 and S2)^{25–51}. These studies shed light on the morphology-molecular association or “histo-genomics”^{52,53}, which may contribute toward the discovery of cost-effective biomarkers and improved therapeutic options⁵⁴.

In this study, we present a deep learning-based system for automated EBV prediction directly from H&E stained WSI. Our model was trained on WSI obtained from the TCGA dataset, where the EBV results were derived from RNA sequencing (RNA-seq). However, as EBER-ISH is a more common method to detect EBV status in practice^{11,23}, a dataset labeled with EBER-ISH was used for fine-tuning to mitigate estimation bias and ensure robustness. To assess its generalizability, we externally validated our network on a dataset originating from a different institution.

Results

Patch-wise performance on TCGA dataset (internal validation). The baseline framework has two sequential binary classifiers: a tumor classifier and an EBV classifier (Fig. 1). The performance of each classifier, trained using a different patch size and model, are shown in Table 1. To select the models for baseline framework, we compared the results of the internal validation using patch-wise performance on the hold-out TCGA dataset (Supplementary Table S3). As shown in Table 1, the models yielded better performance with a patch size of 512 × 512 pixels than with 256 × 256 pixels, while the computation time on larger images was longer than on smaller images (Supplementary Table S4).

We embedded the tumor classifier to select representative tumor patches. For the tumor classifier, we used ResNet50 because of its high specificity (Table 1). For the EBV classifier, we implemented InceptionV3, which yields higher sensitivity, since our aim was to enable our network to be used as a screening tool for EBV identification.

The confusion matrix, where the performance of the two classifiers was evaluated, is shown in Supplementary Figure S1. The baseline frameworks with different combinations of sequential binary classifiers were internally validated on the hold-out TCGA dataset, which was assessed with the 3-class classifier as well (Supplementary Fig. S2 and Supplementary Table S5). The results showed that when using the sequential binary classifiers, the false positive rate drastically decreased from 0.056 to 0.005 (Supplementary Figure S2 and Supplement table S5) and the overall performance was better in the sequential binary classifiers than in the 3-class classifier. The sequential binary classifiers with the best performance yielded a macro-negative predictive value (NPV) of 0.99 and a macro-sensitivity of 0.96.

Representation visualization. To have an insight into the model’s representation, we visualized the learned feature space in two dimensions using the hold-out TCGA dataset. The discriminative patches with similar features according to the classifiers were clustered closely in distinct regions, which proves the disentangled representation of each model (Fig. 2). Interestingly, several misclassified patches according to the binary classifiers are found at the edges of the other side cluster (Fig. 2a,b, arrows), whereas the misclassified patches predicted by the 3-class classifier are located far from their own cluster (Fig. 2c, arrows). Based on this result, it can be said that the binary classifiers learned the representation more efficiently than the 3-class classifier.

Comparison of the baseline framework before and after fine-tuning (external validation based on slide-level performance). In comparison with the slide-level performances using WSIs from Hanyang University Guri Hospital (Guri, South Korea; denoted by HGH), the performance after fine-tuning using WSIs from the International St. Mary’s Hospital (Incheon, South Korea; denoted by ISH) improved as follows: the accuracy increased from 0.77 to 0.92, the area under the receiver operating characteristic (AUROC) increased from 0.77 to 0.88, and the area under the precision-recall curve (AUPRC) increased from 0.38 to 0.65 (Table 2, Fig. 3).

We implemented EBV probability heatmaps to visualize the effect of fine-tuning (Fig. 4a,b) and to identify the potential reasons for misclassification (Fig. 4c,d). After fine-tuning, regions with high probabilities of EBV-positive tumor tissue increased significantly (Fig. 4a,b). Tumors with a low probability for EBV (Fig. 4c) show differentiated histology, whereas the characteristic features of lymphoepithelial carcinoma were found with a high probability (Fig. 4d).

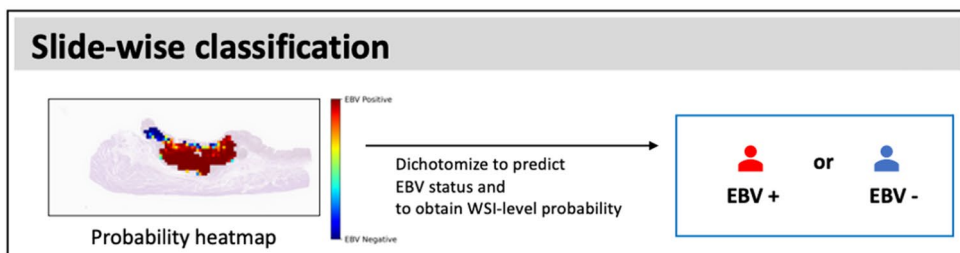
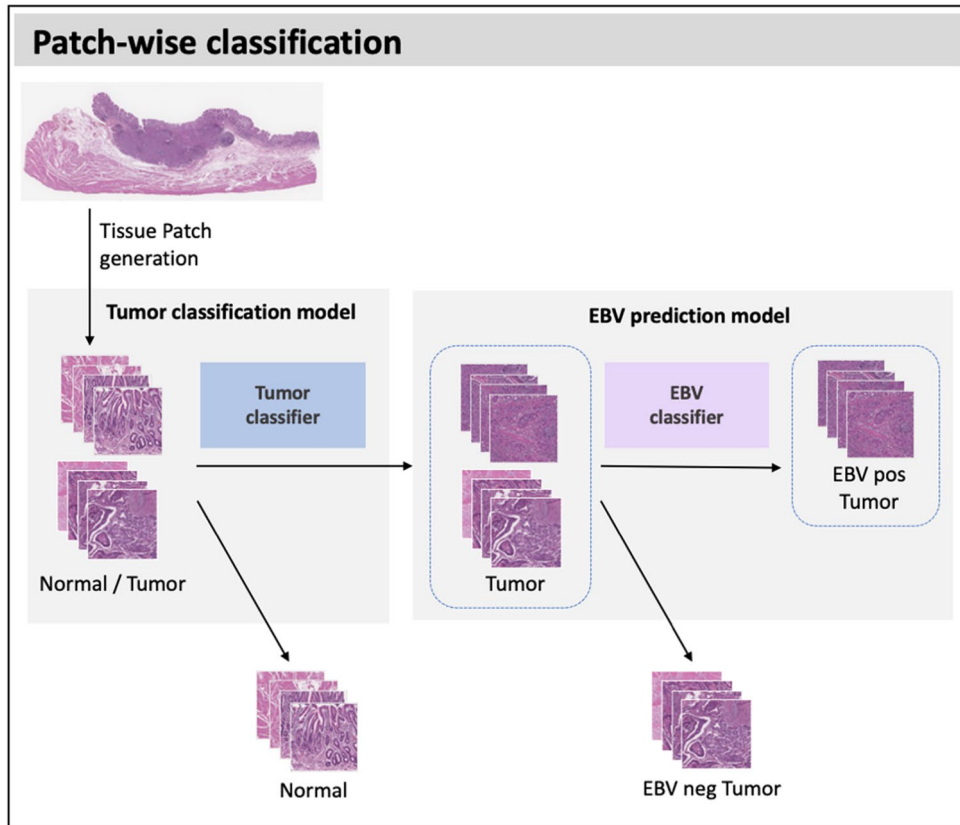
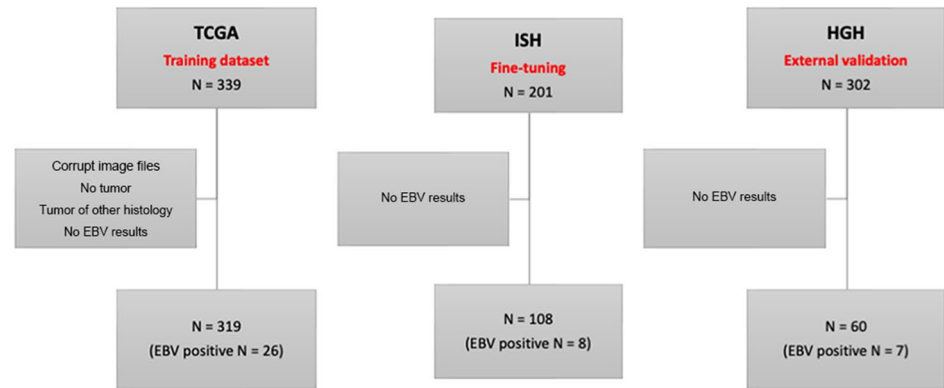


Figure 1. Overview of EBVNet. EBVNet was trained on the TCGA dataset, and fine-tuned on the ISH dataset. EBVNet before and after fine-tuning were externally validated on the HGH dataset. Each H&E histology slide was preprocessed by discarding non tissue-containing background using Otsu’s thresholding and generating non-overlapping patches. All the patches were stain normalized (not shown in the figure). First, the tumor classification model (blue box) receives all the tissue-containing patches and returns the probability of each patch being a tumor patch. The patches with probabilities of higher than 0.5 are assigned as tumor patches, otherwise, they are assigned as normal patches. The tumor-predicted patches are then fed into the second classifier (EBV prediction model, purple box), with the resulting probability of being an EBV positive tumor patch. The tumor-predicted patches with probabilities higher than 0.1 are assigned as EBV positive tumor patches, otherwise, they are assigned as EBV negative patches. Based on the patch-wise classification result, the EBV probability score (EPS), which is the ratio of the number of EBV-positive tumor-predicted patches to the number of tumor-predicted patches, is calculated for each slide. The slide with calculated EPS higher than 0.2 is assigned as an EBV positive slide, otherwise, it is assigned as an EBV negative slide.

	Model	Patch size (pixels)	Accuracy	NPV ^a	Specificity	Sensitivity	Precision	F1-score
Tumor classifier	InceptionV3	256 × 256	0.98	0.97	0.96	0.98	0.98	0.98
		512 × 512	0.98	0.99	0.95	0.99	0.98	0.99
	ResNet50	256 × 256	0.98	0.97	0.96	0.99	0.98	0.98
		512 × 512	0.98	0.98	0.97	0.99	0.99	0.99
EBV classifier	InceptionV3	256 × 256	0.93	0.99	0.93	0.90	0.56	0.69
		512 × 512	0.99	0.99	0.99	0.92	0.92	0.92
	ResNet50	256 × 256	0.92	0.97	0.94	0.71	0.54	0.61
		512 × 512	0.98	0.99	0.99	0.91	0.86	0.89

Table 1. Patch-wise model performances in the tumor classifier (upper) and the EBV classifier (lower). Bold font: maximum value of specific metrics across different model and patch size. ^aNegative predictive value.

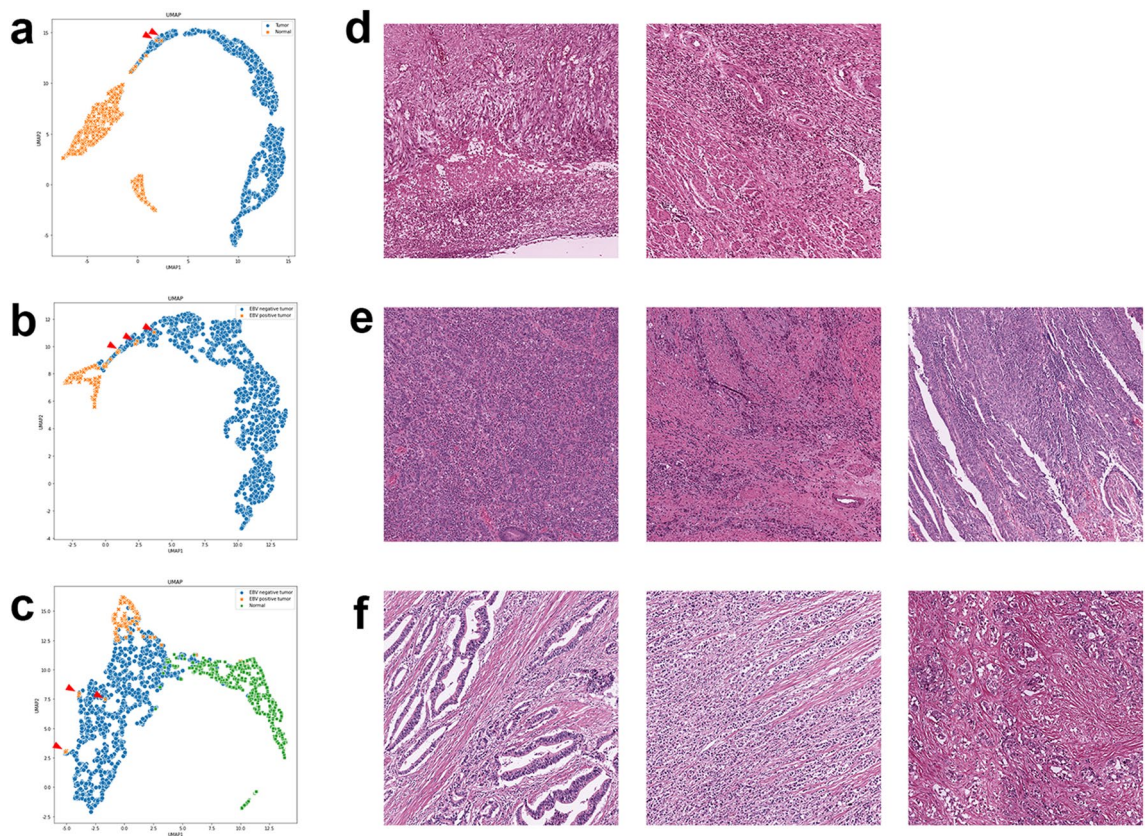


Figure 2. Uniform Manifold Approximation and Projection (UMAP) visualization of neural network latent space and examples of false results. All input patches were fed into classification models and the corresponding post-convolution dense vectors (1024-length) were generated. All the vectors were mapped into 2-dimensional space using UMAP. Each point corresponds to a patch in (a) the tumor classifier, (b) the EBV classifier, and (c) the 3-Class classifier, colored by tissue type (tumor vs. normal) and tumor type (EBV positive vs. EBV negative). Histology images (d–f) of false predictions (red triangles) of each classifier (a–c), which are located apart from their corresponding cluster on UMAP. These points were examined to investigate potential reasons for misclassification.

Analysis of false results. We reviewed falsely predicted image patches on the hold-out TCGA test set (internal validation) (Fig. 2d–f). False-positive image patches by the tumor classifier include inflammatory necrotic tissue, florid granulation tissue, and dense infiltration of inflammatory cells, which could be mistaken as poorly-differentiated adenocarcinoma (Fig. 2d). False-negative tumor patches that were EBV positive but predicted as EBV negative by the EBV classifier and 3-class classifiers show a lack of lymphoepithelial features (Fig. 2e,f).

EBVNet identified one false negative and four false positives on the slide-wise HGH external validation set (Supplementary Figs. S3 and S4). The false negative exhibits differentiated histology with well-formed tubules

	Accuracy	NPV ^a	Specificity	Sensitivity	Precision	F1-score	AUROC ^b	AUPRC ^c
Before transfer learning	0.77	0.95	0.77	0.71	0.29	0.29	0.77	0.38
After transfer learning	0.92	0.98	0.92	0.86	0.60	0.71	0.88	0.65
Performance gain (value, %)	+0.15 (19.48)	+0.03 (3.16)	+0.15 (19.48)	+0.15 (21.13)	+0.31 (106.90)	+0.42 (144.83)	+0.11 (14.29)	+0.27 (71.05)

Table 2. Slide-wise performance on the external dataset of the HGH cohort. ^aNegative predictive value. ^bThe area under the receiver-operating characteristics curve. ^cThe area under the precision-recall curve.

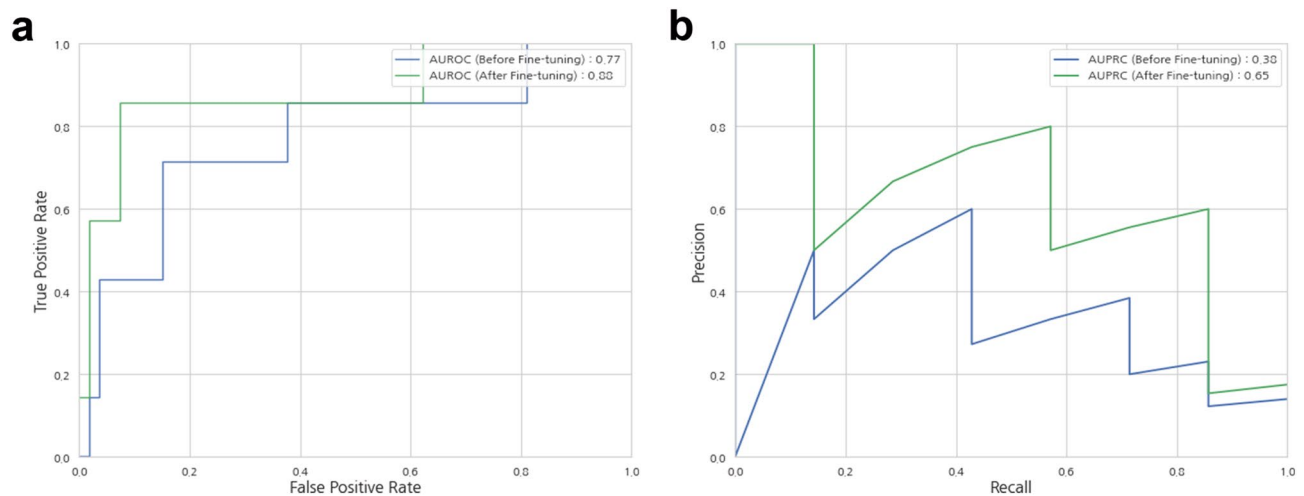


Figure 3. Performance of the models. The area under the receiver operating characteristic (AUROC) curve (a) and area under the precision-recall curve (AUPRC) (b) of slide-level EBV status inference using the original classifiers solely trained with TCGA dataset (blue) and re-weighted classifiers fine-tuned with ISH dataset (green).

(Supplementary Fig. S3). Of the four false positives, the algorithm falsely identified prominent lymphoid stroma and lymphoepithelial carcinoma (Supplementary Fig. S4A–C), and poorly cohesive carcinoma (Supplementary Fig. S4D), as EBV-positive tumors.

Comparison of deep-learning model to pathologists based on slide-level performance. The performance results of all four pathologists were below the models' AUROC and AUPRC, with a mean AUROC of 0.75 and a mean AUPRC of 0.41 (Supplementary Table S7). While the NPV and specificity were similar, EBVNet exhibited higher sensitivity than the pathologists.

Discussion

In this study, we present EBVNet, a deep learning model to predict EBV status directly from histological images. Our model achieved higher performance than the network without fine-tuning, and outperformed experienced pathologists on a reader study. Our study also validated the feasibility of fine-tuning for domain adaptation and model generalizability.

Until now, seven studies on EBV status prediction via a deep learning approach using digitalized WSIs have been published (Supplementary Table S1)^{25–31}. Of these, the most similar to our pipeline are those proposed by Zhang et al. and Zheng et al.^{28,29}; in these two previous studies, a two-step approach utilizing a tumor classifier and an EBV classifier was implemented. Most studies have trained EBV classifiers using tumor patches generated from tumor annotation, with^{28,29} or without^{25–27} training tumor classifiers—a tumor annotation-based approach. Other studies have trained the model without using tumor or normal patches—an annotation-free training approach^{30,31}. The latter is called weakly supervised learning where only a slide-level label (weak label) per image was available for model development^{30,31}. Of the studies with weakly supervised learning, Muti et al. reported a comparable performance with an AUROC of 0.859 on the external dataset. However, the results of these previous studies elucidate that supervised learning approaches, where the model was trained on tumor patches, outperformed the weakly supervised approach with an AUROC of 0.941²⁹.

Most histology-based deep learning studies implemented patch-based workflow due to hardware memory constraints. Patch selection techniques that take meaningful representative images for network training are paramount for higher network performance⁵⁵. In the current study, we assumed that normal patches cannot provide any information about the genetic trait of the tumor. Therefore, we only used tumor patches based on the tumor

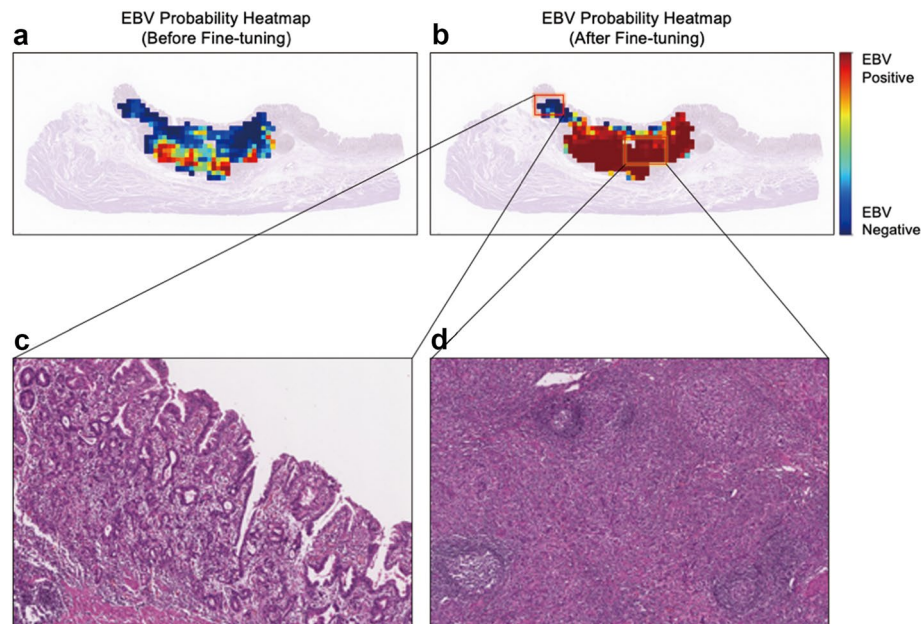


Figure 4. Representative heatmaps of an EBV positive case. The colors of the overlaid heatmaps represent the output of the EBV classification model (second classifier in the sequential classification framework), which means predicted probabilities of being an EBV-positive tumor, as defined in the color bar. In the top row, each heatmap corresponds to the EBV positive tumor probability values from classifiers before (a) and after (b) fine-tuning with the ISH dataset. A zoomed-in view of a region with low probability (c) and high probability (d) of being an EBV positive tumor, according to the fine-tuned EBV classifier (EBVNet), is visualized in the bottom row.

annotation dataset when training the EBV classifier, similar to previous studies^{25–29}. The tumor classifier in our proposed framework may have been useful in selecting representative patches (tumor patches) and in removing noise images of normal patches. We elucidated that the implementation of this patch selection network in sequential binary classifiers led to better performance than with a 3-class classifier.

However, training a discriminative model with supervision using tumor annotation has two drawbacks. First, annotation is a laborious, time-consuming, and expensive task, requiring domain knowledge and expertise. Second, detection of biomarkers using a model trained on annotated data is restricted to the detection of tumor regions that are already known. In contrast, unsupervised or weakly supervised learning allows the identification of novel biomarkers that are biologically relevant to the target of the model^{56,57}. In the recent study by Brockmoeller et al., the model was trained on slide-level labels using tumor and non-tumor regions, and the image biomarker in the normal area, not in the tumor area, was first identified to predict lymph node metastasis⁵⁶.

The increasing growth of deep learning owes plenty to “open data mentality”, with researchers sharing their datasets and code. Nevertheless, there is a lack of annotation data in histology-based deep learning, as the annotation of WSI is expensive and time-consuming. We would like to support further open-source development by sharing the annotations of TCGA-STAD employed in this study (<https://github.com/EBVNET/EBVNET>).

Our study has certain limitations that require further analysis. First, our proposed pipeline employed a fixed pooling operator for a slide-level aggregation. This simple decision fusion method of aggregating patch-level labels to a slide-level label is inconsistent with the decision process in pathology. In addition, this approach disregards spatial relationships between patches, resulting in a loss of global contextual information present in WSI. However, since EBV-positive tumors have morphologically homogeneous characteristics, simple thresholding or major voting, which are the most widely used approaches for the post-processing strategies of classification tasks, would have been effective³⁵. Second, the use of EBV identification as a screening tool is more helpful on biopsy specimens, where lesions are much smaller and likely to be ignored. However, we validated EBVNet on the external dataset of HGH, which consists of WSI from gastrectomy specimens (Supplementary Table S6). The performance of our pipeline should be further fine-tuned and investigated on the hard cases of these biopsy slides. Finally, our network should be ameliorated for clinical applications achieving 100% sensitivity with an acceptable false rate. A large dataset for training or fine-tuning will be helpful to achieve this.

In conclusion, our study has illuminated the potential of deep learning systems to identify histology-based biomarkers. Our EBVNet is expected to serve as an alternative to effectively screen EBV status using ubiquitously available H&E slides.

Methods

Experimental design. In this study, three different datasets were used for training, fine-tuning, and external validation, respectively (Fig. 1). EBVNet, which consisted of the tumor classifier and EBV classifier (Fig. 1), was trained on a public dataset (TCGA-STAD). For fine-tuning and external validation, we used two datasets

from different institutes, the ISH and HGH (Supplementary Table S6). The ISH dataset was employed for fine-tuning, and the HGH dataset for external validation. We also compared the performance of EBVNet before and after fine-tuning using slide-wise performance on the HGH dataset.

Data acquisition and annotation. We retrieved anonymized histology images (diagnostic slides, FFPE tissue) from the TCGA-STAD project through the Genomic Data Commons Portal⁵⁸. A total 319 WSI were used to train the deep learning model after removal of WSI with corrupt image files, lack of visible tumor tissue, missing values in EBER-RNA results, or histology for tumors other than adenocarcinoma (Fig. 1).

For fine-tuning, 108 WSI of gastric cancer were collected from ISH. To externally validate the developed model, we used an independent cohort that encompasses 60 WSI of stomach cancer with surgical resection from HGH.

All slides from the ISH and HGH cohorts were scanned at 40× objective magnification (~0.25 μm/pixel) using Leica Aperio ScanScope AT2 (Leica Biosystems, Wetzlar, Germany). This study was approved by the Institutional Review Board at ISH (IRB no., IS21SIME0031) and HGH (IRB no., 2020-09-002), respectively, and conducted in accordance with the Declaration of Helsinki. Informed consent from patients was waived with IRB approval.

Digitalized H&E histology slides were re-surveyed, and adenocarcinoma regions in the TCGA and the ISH dataset were manually annotated by an expert gastrointestinal pathologist (S. Ahn) using the ASAP software, v.1.9.0 (Geert Litjens, Nijmegen, Netherlands). Every region with adenocarcinoma was assigned as “tumor.” As all WSIs used in this study contained tumor areas, which also have areas of normal tissue, the remaining unannotated areas were considered as normal tissue. We generated tumor patches based on tumor annotation, and normal patches derived from unannotated areas (Supplementary Table S3). Randomly selected patches generated from the TCGA datasets and not used in training classifiers, were used for internal validation (the hold-out TCGA dataset).

Full details for the strategies of image preprocessing, including patch generation, are provided in the Supplementary method.

EBV labels. For TCGA samples, the EBV status was retrieved directly from the molecular result, as described by Liu et al. and Tathiane et al.^{9,59} without re-analysis for EBV. For samples in the ISH and HGH cohorts, EBV status was determined via EBER-ISH results.

Model development. Deep neural networks, denoted by EBVNet, were trained on image patches with the aim of predicting EBV status in gastric cancers. The baseline framework consists of two sequential binary classifiers (Fig. 1). The first binary classifier is a tumor classification model (Normal vs. Tumor), and it is followed by the second binary classifier, which is the EBV prediction model (EBV negative tumor vs. EBV positive tumor). Among all the patches fed to the first classifier, only the patches predicted as tumors enter the second classifier. For training binary classifiers, two networks were used: ResNet50 and Inception V3. The ResNet network is well known for its residual connections and therefore used to mitigate the vanishing gradient and stabilize the learning process. The inception network consists of inception blocks with different parallel convolution filters for the extraction of representative features with fewer parameters. As an alternative approach to the sequential binary classifiers, we conducted an experiment to train a simple multi-classification network for EBV positive tumor, EBV negative tumor, and normal tissues, denoted by a 3-class classifier. For the 3-class classifier, Inception V3 was employed.

Limited generalizability of a deep learning model may hinder its clinical applicability resulting in poor performance on real-world data. One way to mitigate this problem is to re-weight the model on target datasets with domain-specific features. Following this approach, fine-tuning can be used to ensure domain adaptation. Herein, the weight values of each binary classifier trained on the source TCGA dataset were employed to fine-tune the ISH dataset, as the ISH and HGH cohorts exhibited similar clinical characteristics and color spaces (Supplementary Table S6). All the layers were unfrozen and re-trained using the ISH dataset.

Full details on neural network training, model selection, hyperparameter optimization and data augmentation, are provided in the Supplementary method.

Inference of EBV status. For patch-wise EBV status inference in a sequential binary classifier framework, first, patch images were given as input to the tumor classification model (Fig. 1), which then provided the probability of each patch being a tumor patch as output. The patches with predicted probabilities higher than 0.5 were assigned as tumor-predicted patches (N_{tumor}). Second, tumor-predicted patches were given as input to the EBV prediction model, which then provided the probability of each patch being an EBV positive tumor patch as output. The patches with predicted probabilities higher than 0.1 were assigned as EBV positive-predicted patches (N_{EBVpos}). Unlike the 0.5 threshold value of the tumor classifier, the cut-off for the EBV classifier was set to the low value of 0.1 to ensure that the sensitivity is high when detecting an EBV positive tumor patch. For inference in a 3-class classifier framework, patch images were given as input to the classification model (Normal vs. EBV negative tumor vs. EBV positive tumor), which then returned the probabilities of each class as output. The class with the highest probability was assigned as the prediction for each patch.

For slide-wise classification (EBV positive slide vs. EBV negative slide) after the patch-wise EBV status inference, we defined the EBV probability score (EPS) as the ratio of EBV positive-predicted patches to tumor-predicted patches ($N_{\text{EBVpos}} / N_{\text{tumor}}$). The cut-off value for EPS was calculated using Youden's J statistic⁶⁰, which is the cut-off value with the highest difference between true positive and false positive rate. The cut-off value was calculated for the HGH dataset as 0.2.

Internal and external validation. The binary classifiers (the tumor classifier and the EBV classifier) and the 3-class classifier trained on the TCGA dataset were internally validated on the hold-out TCGA dataset using patch-wise performance.

We externally validated the performance of EBVNet on the HGH dataset. Slide-wise performance was assessed through the dichotomized outcomes of WSI in EBV status. The performance of EBVNet before fine-tuning was also compared to that after fine-tuning.

Feature visualization. To visualize and interpret the deep learning predictions, we employed three approaches. First, we identified the falsely predicted patches for each class of the TCGA dataset, allowing observers to identify potential reasons for misclassification. Second, in order to appreciate the variability of learned representation, we used the three classifiers to generate and visualize post-convolution layer activation for the hold-out TCGA dataset. Activation values in the post-convolution layer were calculated on each image patch. Once calculated, activation vectors across all patches were mapped with the dimensionality reduction technique known as Uniform Manifold Approximation and Projection (UMAP)⁶¹. Finally, we rendered patch-level predictions for EBV status as activation maps, visualizing prediction scores as heatmap overlays on the original WSI in the HGH cohort.

Reader study. To compare the performance of EBVNet with that of pathologists, we conducted a reader study in which four board-certified pathologists with 4–20 years of experience (JJ, JP, JS, and S Ahn) reviewed 168 WSI from both cohorts. The pathologists were blind to all clinical information, as well as EBVNet's performance. For each WSI, they recorded whether they thought the cancer could be classified as EBV positive or EBV negative.

Evaluation metrics. We compared the performance of each algorithm trained on the TCGA dataset, using patch-wise accuracy, sensitivity, specificity, NPV, precision, and F1-score. We measured the slide-wise performance of our network on HGH dataset using the AUPRC, AUROC, accuracy, sensitivity, specificity, NPV, precision, and F1-score. These metrics were defined as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN},$$

$$\text{Sensitivity(Recall)} = \frac{TP}{TP + FN},$$

$$\text{Specificity} = \frac{TN}{FP + TN},$$

$$\text{Precision} = \frac{TP}{TP + FP},$$

$$\text{Negative predictive values(NPV)} = \frac{TN}{FN + TN},$$

$$\text{F1-score} = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}},$$

where TP, TN, FP, FN represented the total number of true positive, true negative, false positive, and false negative, respectively.

Data availability

The data that support the findings of this study were derived in part from TCGA database (<https://www.cancer.gov/tcga>) and their annotation are provided via an open-source license at <https://github.com/EBVNET/EBVNET>.

Code availability

All source codes are available under an open-source license at <https://github.com/EBVNET/EBVNET>.

Received: 30 March 2022; Accepted: 18 October 2022

Published online: 02 November 2022

References

1. Wakiguchi, H. Overview of Epstein–Barr virus-associated diseases in Japan. *Crit. Rev. Oncol. Hematol.* **44**, 193–202 (2002).
2. Lee, J. H. *et al.* Clinicopathological and molecular characteristics of Epstein–Barr virus-associated gastric carcinoma: A meta-analysis. *J. Gastroenterol. Hepatol.* **24**, 354–365 (2009).
3. Murphy, G., Pfeiffer, R., Camargo, M. C. & Rabkin, C. S. Meta-analysis shows that prevalence of Epstein–Barr virus-positive gastric cancer differs based on sex and anatomic location. *Gastroenterology* **137**, 824–833 (2009).
4. Panda, A. *et al.* Immune activation and benefit from avelumab in EBV-positive gastric cancer. *J. Natl. Cancer Inst.* **110**, 316–320 (2018).

5. van Beek, J. *et al.* EBV-positive gastric adenocarcinomas: A distinct clinicopathologic entity with a low frequency of lymph node involvement. *J. Clin. Oncol.* **22**, 664–670 (2004).
6. Cheng, Y., Zhou, X., Xu, K., Huang, J. & Huang, Q. Very low risk of lymph node metastasis in Epstein–Barr virus-associated early gastric carcinoma with lymphoid stroma. *BMC Gastroenterol.* **20**, 273 (2020).
7. van Beek, J. *et al.* Morphological evidence of an activated cytotoxic T-cell infiltrate in EBV-positive gastric carcinoma preventing lymph node metastases. *Am. J. Surg. Pathol.* **30**, 59–65 (2006).
8. Osumi, H. *et al.* Epstein–Barr virus status is a promising biomarker for endoscopic resection in early gastric cancer: Proposal of a novel therapeutic strategy. *J. Gastroenterol.* **54**, 774–783 (2019).
9. Liu, Y. *et al.* Comparative molecular analysis of gastrointestinal adenocarcinomas. *Cancer Cell* **33**, 721–735e728 (2018).
10. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513**, 202–209 (2014).
11. Ignatova, E. *et al.* Epstein–Barr virus-associated gastric cancer: Disease that requires special approach. *Gastric Cancer* **23**, 951–960 (2020).
12. Saito, R. *et al.* Overexpression and gene amplification of PD-L1 in cancer cells and PD-L1(+) immune cells in Epstein–Barr virus-associated gastric cancer: The prognostic implications. *Mod. Pathol.* **30**, 427–439 (2017).
13. Sasaki, S. *et al.* EBV-associated gastric cancer evades T-cell immunity by PD-1/PD-L1 interactions. *Gastric Cancer* **22**, 486–496 (2019).
14. Shitara, K. *et al.* Efficacy and safety of pembrolizumab or pembrolizumab plus chemotherapy vs chemotherapy alone for patients with first-line, advanced gastric cancer: The KEYNOTE-062 phase 3 randomized clinical trial. *JAMA Oncol.* **6**, 1571–1580 (2020).
15. Kang, Y. K. *et al.* Nivolumab in patients with advanced gastric or gastro-oesophageal junction cancer refractory to, or intolerant of, at least two previous chemotherapy regimens (ONO-4538-12, ATTRACTION-2): A randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet* **390**, 2461–2471 (2017).
16. Kim, S. Y. *et al.* Deregulation of immune response genes in patients with Epstein–Barr virus-associated gastric cancer and outcomes. *Gastroenterology* **148**, 137–147.e139 (2015).
17. Song, H. J. *et al.* Host inflammatory response predicts survival of patients with Epstein–Barr virus-associated gastric carcinoma. *Gastroenterology* **139**, 84–92.e82 (2010).
18. Song, H. J. & Kim, K. M. Pathology of Epstein–Barr virus-associated gastric carcinoma and its relationship to prognosis. *Gut Liver* **5**, 143–148 (2011).
19. Camargo, M. C. *et al.* Improved survival of gastric cancer with tumour Epstein–Barr virus positivity: An international pooled analysis. *Gut* **63**, 236–243 (2014).
20. Tokunaga, M. & Land, C. E. Epstein–Barr virus involvement in gastric cancer: Biomarker for lymph node metastasis. *Cancer Epidemiol. Biomark. Prevent.* **7**, 449–450 (1998).
21. Park, J. H. *et al.* Epstein–Barr virus positivity, not mismatch repair-deficiency, is a favorable risk factor for lymph node metastasis in submucosa-invasive early gastric cancer. *Gastric Cancer* **19**, 1041–1051 (2016).
22. Sohn, B. H. *et al.* Clinical significance of four molecular subtypes of gastric cancer identified by the Cancer Genome Atlas Project. *Clin. Cancer Res.* **23**, 4441–4449 (2017).
23. Gulley, M. L. & Tang, W. Laboratory assays for Epstein–Barr virus-related disease. *J. Mol. Diagn.* **10**, 279–292 (2008).
24. Bera, K., Schalper, K. A., Rimm, D. L., Velcheti, V. & Madabhushi, A. Artificial intelligence in digital pathology—New tools for diagnosis and precision oncology. *Nat. Rev. Clin. Oncol.* **16**, 703–715 (2019).
25. Flinner, N. *et al.* Deep learning based on hematoxylin-eosin staining outperforms immunohistochemistry in predicting molecular subtypes of gastric adenocarcinoma. *J. Pathol.* **257**, 218–226 (2022).
26. Hinata, M. & Ushiku, T. Detecting immunotherapy-sensitive subtype in gastric cancer using histologic image-based deep learning. *Sci. Rep.* **11**, 22636 (2021).
27. Kather, J.N., *et al.* Deep learning detects virus presence in cancer histology. *BioRxiv* 690206 (2019).
28. Zhang, B., Yao, K., Xu, M., Wu, J. & Cheng, C. Deep learning predicts EBV status in gastric cancer based on spatial patterns of lymphocyte infiltration. *Cancers (Basel)* **13**, 6002 (2021).
29. Zheng, X. *et al.* A deep learning model and human-machine fusion for prediction of EBV-associated gastric cancer from histopathology. *Nat. Commun.* **13**, 2790 (2022).
30. Muti, H. S. *et al.* Development and validation of deep learning classifiers to detect Epstein–Barr virus and microsatellite instability status in gastric cancer: a retrospective multicentre cohort study. *Lancet Digit. Health* **3**, e654–e664 (2021).
31. Ghaffari Laleh, N. *et al.* Benchmarking weakly-supervised deep learning pipelines for whole slide classification in computational pathology. *Med. Image Anal.* **79**, 102474 (2022).
32. Coudray, N. *et al.* Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat. Med.* **24**, 1559–1567 (2018).
33. Kather, J. N. *et al.* Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat. Med.* **25**, 1054–1056 (2019).
34. Liu, S. *et al.* Isocitrate dehydrogenase (IDH) status prediction in histopathology images of gliomas using deep learning. *Sci. Rep.* **10**, 7733 (2020).
35. Schmauch, B. *et al.* A deep learning model to predict RNA-Seq expression of tumours from whole slide images. *Nat. Commun.* **11**, 3877 (2020).
36. Kather, J. N. *et al.* Pan-cancer image-based detection of clinically actionable genetic alterations. *Nat. Cancer* **1**, 789–799 (2020).
37. Sirinukunwattana, K. *et al.* Image-based consensus molecular subtype (imCMS) classification of colorectal cancer using deep learning. *Gut* **70**, 544–554 (2021).
38. Jang, H. J., Lee, A., Kang, J., Song, I. H. & Lee, S. H. Prediction of clinically actionable genetic alterations from colorectal cancer histopathology images using deep learning. *World J. Gastroenterol.* **26**, 6207–6223 (2020).
39. Sun, M. *et al.* Prediction of BAP1 expression in uveal melanoma using densely-connected deep classification networks. *Cancers (Basel)* **11**, 1579 (2019).
40. Sha, L. *et al.* Multi-field-of-view deep learning model predicts non small cell lung cancer programmed death-ligand 1 status from whole-slide hematoxylin and eosin images. *J. Pathol. Inform.* **10**, 24 (2019).
41. Yamashita, R. *et al.* Deep learning model for the prediction of microsatellite instability in colorectal cancer: A diagnostic study. *Lancet Oncol.* **22**, 132–141 (2021).
42. Xu, Z. *et al.* Deep learning predicts chromosomal instability from histopathology images. *iScience* **24**, 102394 (2021).
43. Couture, H. D. *et al.* Image analysis with deep learning to predict breast cancer grade, ER status, histologic subtype, and intrinsic subtype. *NPJ Breast Cancer* **4**, 30 (2018).
44. Schrammen, P. L. *et al.* Weakly supervised annotation-free cancer detection and prediction of genotype in routine histopathology. *J. Pathol.* **256**, 50–60 (2022).
45. Schaumberg, A.J., Rubin, M.A. & Fuchs, T.J. *H&E-Stained Whole Slide Image Deep Learning Predicts SPOP Mutation State in Prostate Cancer.* (bioRxiv, 2018).
46. Kather, J.N., *et al.* *Deep Learning Detects Virus Presence in Cancer Histology.* (bioRxiv, 2019).
47. Xu, H., Park, S., Lee, S.H. & Hwang, T. *Using Transfer Learning on Whole Slide Images to Predict Tumor Mutational Burden in Bladder Cancer Patients.* (bioRxiv, 2019).

48. Fu, Y. *et al.* Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. *Nat. Cancer* **1**, 800–810 (2020).
49. Kim, R.H., *et al.* *A Deep Learning Approach for Rapid Mutational Screening in Melanoma.* (bioRxiv, 2019).
50. Zhang, H., *et al.* Predicting tumor mutational burden from liver cancer pathological images using convolutional neural network. in *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (eds. Yoo, I., Bi, J. & Hu, X.). 920–925 (IEEE, 2019).
51. Bilal, M., *et al.* *Novel Deep Learning Algorithm Predicts the Status of Molecular Pathways and Key Mutations in Colorectal Cancer from Routine Histology Images.* (medRxiv, 2021).
52. Shia, J. *et al.* Morphological characterization of colorectal cancers in The Cancer Genome Atlas reveals distinct morphology-molecular associations: Clinical and biological implications. *Mod. Pathol.* **30**, 599–609 (2017).
53. Greenson, J. K. *et al.* Pathologic predictors of microsatellite instability in colorectal cancer. *Am. J. Surg. Pathol.* **33**, 126–133 (2009).
54. Cooper, L. A. *et al.* Novel genotype-phenotype associations in human cancers enabled by advanced molecular platforms and computational analysis of whole slide images. *Lab. Invest.* **95**, 366–376 (2015).
55. Salvi, M., Acharya, U. R., Molinari, F. & Meiburger, K. M. The impact of pre- and post-image processing techniques on deep learning frameworks: A comprehensive review for digital pathology image analysis. *Comput. Biol. Med.* **128**, 104129 (2021).
56. Brockmoeller, S. *et al.* Deep learning identifies inflamed fat as a risk factor for lymph node metastasis in early colorectal cancer. *J. Pathol.* **256**, 269–281 (2022).
57. Schlegl, T., Seeböck, P., Waldstein, S. M., Langs, G. & Schmidt-Erfurth, U. f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. *Med. Image Anal.* **54**, 30–44 (2019).
58. National Cancer Institute. *Genomic Data Commons (GDC) Data Portal.* Vol. 2022 (National Cancer Institute, 2021).
59. Malta, T. M. *et al.* Machine learning identifies stemness features associated with oncogenic dedifferentiation. *Cell* **173**, 338–354e315 (2018).
60. Youden, W. J. Index for rating diagnostic tests. *Cancer* **3**, 32–35 (1950).
61. McInnes, L., Healy, J. & Melville, J. *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction.* (arXiv, 2020).

Acknowledgements

This work was supported by research grants as follows; 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: HI21C0940); from the Medical Research Promotion Program through the Gangneung Asan Hospital funded by the Asan Foundation (grant number: 2021 B002); and from a Korea University Grant (grant number: K2225541).

Author contributions

S.A., K.M.M., and N.K. performed study concept and design; S.A., K.M.M., Y.J., C.E.C., J.E.K., J.P., and J.L. performed data analysis, and experiment for model development; S.A. performed writing; K.M.M., Y.J., C.E.C., J.E.L., J.H.K., N.K., and J.S. performed review and revision of the paper; W.Y.J., and Y.J.L. provided material support; S.A., J.J., J.P., and J.S. performed histologic review; J.E.K. and J.L. provided technical support for web page creation. All authors read, edited, and approved the final paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-22731-x>.

Correspondence and requests for materials should be addressed to K.M.M. or S.A.

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

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



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

© The Author(s) 2022