



# Deep-Hipo: Multi-scale receptive field deep learning for histopathological image analysis

Sai Chandra Kosaraju<sup>a,1</sup>, Jie Hao<sup>b,2</sup>, Hyun Min Koh<sup>c,\*,3</sup>, Mingon Kang<sup>a,\*,4</sup>

<sup>a</sup> Department of Computer Science, University of Nevada, Las Vegas, NV, USA

<sup>b</sup> Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, PA, USA

<sup>c</sup> Department of Pathology, Gyeongsang National University Changwon Hospital, Changwon, Republic of Korea

## ABSTRACT

Digitizing whole-slide imaging in digital pathology has led to the advancement of computer-aided tissue examination using machine learning techniques, especially convolutional neural networks. A number of convolutional neural network-based methodologies have been proposed to accurately analyze histopathological images for cancer detection, risk prediction, and cancer subtype classification. Most existing methods have conducted patch-based examinations, due to the extremely large size of histopathological images. However, patches of a small window often do not contain sufficient information or patterns for the tasks of interest. It corresponds that pathologists also examine tissues at various magnification levels, while checking complex morphological patterns in a microscope. We propose a novel multi-task based deep learning model for HistoPathology (named Deep-Hipo) that takes multi-scale patches simultaneously for accurate histopathological image analysis. Deep-Hipo extracts two patches of the same size in both high and low magnification levels, and captures complex morphological patterns in both large and small receptive fields of a whole-slide image. Deep-Hipo has outperformed the current state-of-the-art deep learning methods. We assessed the proposed method in various types of whole-slide images of the stomach: well-differentiated, moderately-differentiated, and poorly-differentiated adenocarcinoma; poorly cohesive carcinoma, including signet-ring cell features; and normal gastric mucosa. The optimally trained model was also applied to histopathological images of The Cancer Genome Atlas (TCGA), Stomach Adenocarcinoma (TCGA-STAD) and TCGA Colon Adenocarcinoma (TCGA-COAD), which show similar pathological patterns with gastric carcinoma, and the experimental results were clinically verified by a pathologist. The source code of Deep-Hipo is publicly available at <http://dataxlab.org/deep-hipo>.

## 1. Introduction

Histopathology is the gold-standard tool in cancer diagnosis and clinical decision making. However, conventional examinations and biopsy diagnoses done by hand are time-consuming and cost intensive [40]. Cancer diagnosis requires massive scans in large fields of various morphology, including benign regions, in a whole slide image. Moreover, the growing number of biopsies and cancer incidences increase the workload of pathologists, which deteriorates efficiency and effectiveness in the examination of biopsies or surgical specimens [19].

Digitizing Whole Slide Images (WSI) of stained tissue specimens has led to the cost-effective computational analysis of histopathological images using machine learning techniques. Accordingly, a growing number of studies have shown the potential of machine learning techniques in several clinical applications during the past decade [17,16], specifically: texture-based convolutional neural networks (CNN) for

cancer detection [32]; image segmentation for histopathological images [37]; patient risk prediction [41]; integrative deep learning for survival analysis [11]; and cancer subtype classification [38].

Machine learning techniques for digital pathology image analysis can be divided into the periods before and after the advent of deep learning (especially, convolutional neural networks), which automatically learns the filters of complex spatial features. Most non-deep learning machine learning techniques have utilized handcrafted features predefined in advance, such as nuclei size and gland shape. For instance, morphological features of interest were extracted by an information maximum gain algorithm, and the handcrafted features were introduced to a Support Vector Machine (SVM) to classify cancer [23]. Regularized logistic regression was trained with features including size, shape, and intensity of nuclei, to classify benign and malignant proliferative breast lesions [25]. Moreover, the handcrafted features of nuclei shape, distribution, and number were fed to a feed forward

\* Corresponding authors.

E-mail addresses: [sai.kosaraju@unlv.edu](mailto:sai.kosaraju@unlv.edu) (S.C. Kosaraju), [jie.hao@penndmedicine.upenn.edu](mailto:jie.hao@penndmedicine.upenn.edu) (J. Hao), [hyunminkoh@gnuch.co.kr](mailto:hyunminkoh@gnuch.co.kr) (H.M. Koh), [mingon.kang@unlv.edu](mailto:mingon.kang@unlv.edu) (M. Kang).

<sup>1</sup> Orcid: 0000-0002-4332-6217.

<sup>2</sup> Orcid: 0000-0001-8758-9265.

<sup>3</sup> Orcid: 0000-0002-7457-7174.

<sup>4</sup> Orcid: 0000-0002-9565-9523.

<https://doi.org/10.1016/j.ymeth.2020.05.012>

Received 16 March 2020; Received in revised form 1 May 2020; Accepted 14 May 2020

Available online 19 May 2020

1046-2023/ © 2020 Elsevier Inc. All rights reserved.

network to predict low and high risk patients [21]. Grey level co-occurrence matrices and grey level run length matrices were computed in segmented breast biopsy WSIs, and SVM was used as a classifier [3]. A graph-based method was developed, incorporating cell nuclei morphology as vertex attributes and neighborhood information as edge attributes, for cancer classification [22]. The results of those non-deep learning machine learning techniques that used handcrafted features are readily interpretable. However, the complex morphological patterns of cancer are often not intuitive, and are difficult to define.

Recently, deep learning-based methods have been widely used for histopathological image analysis [36,1,4]. Deep learning, such as CNNs, has shown tremendously improved performance, by learning spatial features in a hierarchical structure, without feature extraction processing in advance. GoogLeNet identified metastatic breast cancer by examining patches randomly extracted from regions of interest (regions annotated as cancer by a domain expert), and heatmaps were generated to visualize tumor probabilities [34]. A spatially constrained CNN was developed to detect various nuclei shapes in a cancer WSI, and cancer regions were identified from the nuclei classification [28]. A fast CNN model was proposed to identify cancer regions in breast cancer [9]. In the study, adaptive thresholding was applied for sampling patches in the training data, and a simplified version of InceptionV3 was adapted to efficiently predict cancer regions. Rotation equivariant CNN was suggested to analyze pathology images [33], in which the CNN network was trained with additional datasets generated by rotation transformation of patches extracted from WSIs. A fully CNN was applied to histopathological images for image segmentation [24]. A deep CNN model inspired from InceptionV3 detected cancer metastases [7], in which patches of tumor and normal tissue were introduced through InceptionV3 to SVM for localizing cancer regions.

In histopathological image analysis, it is essential to recognize various sizes of complex morphological patterns. Such multi-scale spatial features can be effectively learned by multi-scale receptive field deep learning in complex image data [39,8,12]. Multi-scale receptive field deep learning often deploys asymmetric convolutional blocks that factorize a symmetric kernel (e.g.,  $3 \times 3$  filter) into an asymmetric convolution (e.g.,  $1 \times 3$  filters). For instance, InceptionV3 implements multi-scale kernels in an asymmetric CNN network by varying kernel sizes in convolutional layers and pooling layers for classifying objects of interest [29]. In contrast, Yolov3 [26] and deeplabv3 [6] are optimized for the segmentation and detection of multiple objects in image data with multi-scale CNN networks. Yolov3 adopted residual layers and deeplabv3 suggested pyramid pooling layers for multi-scale receptive fields.

Most deep learning models in digital pathology assume that patches in a histopathological image are independent from each other and contain sufficient information to classify (or predict) clinical outcomes; however, most pathologists interpret histopathological images by examining not only a local region (one patch), but also its neighbor regions (surrounding patches). A few studies have considered the relationships between patches. A set of patches in a tissue were ranked by

a pre-trained CNN model, and the relationships of the ranked patches were analyzed by a Recurrent Neural Network (RNN) [5]. An expectation–maximization based CNN model proposed a post-processing strategy that applies smoothing on the independent predictive scores of patches to reflect dependency between the patches [13]. However, those approaches still rely on the predictive performance of each patch, and often fail to learn large morphological patterns bigger than the patch size.

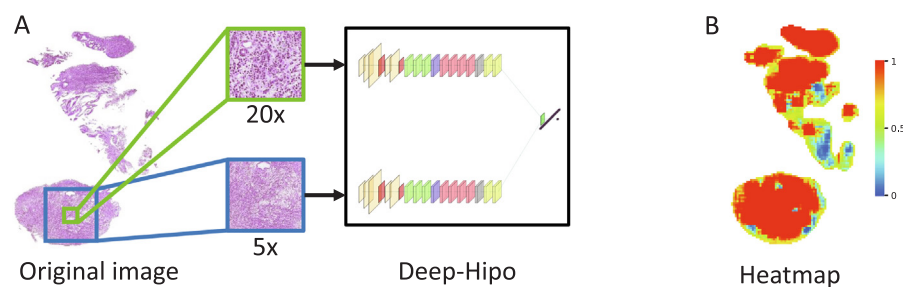
Histopathological image analysis with patches of multiple magnification levels has been recently conducted. An adaptive weighting based multi-field-of-view CNN feeds patches at multiple magnifications into multiple CNNs (one CNN on each magnification level), and the outputs of the CNNs are weighted for semantic segmentation in lung adenocarcinoma given pixel-based annotation [31]. A CNN-based segmentation framework was proposed, where a patch at different sizes in the same magnification level was used to extract pixel/patch-based features simultaneously for histopathological image segmentation [10]. A multi-resolution deep learning framework (MRD-Net) was explored to identify lung adenocarcinoma growth patterns, where two patches of the same size, but different magnifications, were combined together as a pathological image of multiple channel, and the multi-channel images were trained by Resnet [2]. In this paper, we propose a deep learning model for histo- pathology that takes multi-scale patches simultaneously to improve accuracy for histopathological image analysis, as well as identifying cancer regions in a WSI. The main contributions of our model are:

1. learning multi-scale morphological patterns of histo-pathology in both high and low magnification levels simultaneously in a model;
2. interpreting the findings of morphological patterns by connecting to histopathological domain knowledge; and
3. producing better statistically significant performance than the current state-of-the-art methods in histopathological image analysis.

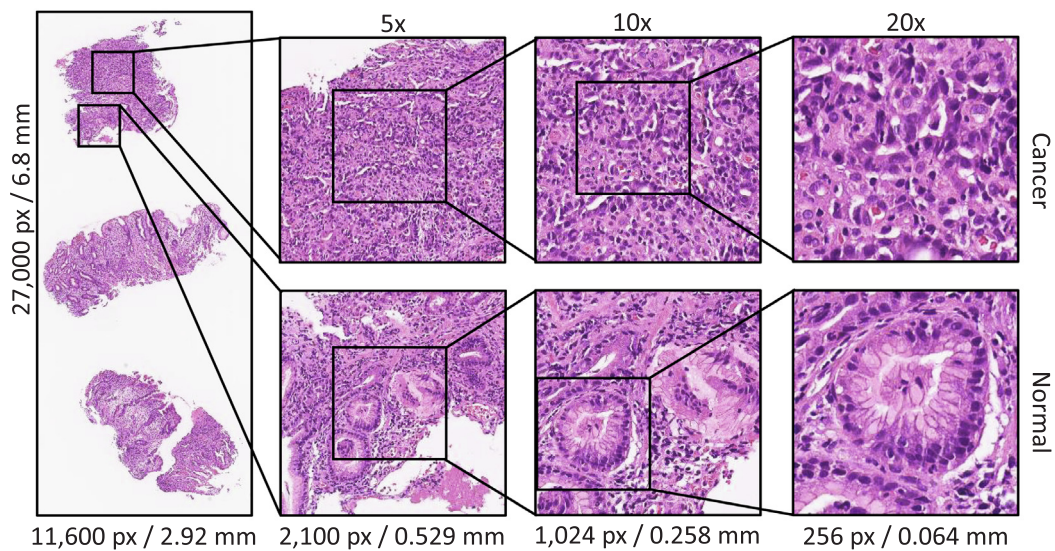
The remainder of this paper is structured as follows. In Section 2, we present our proposed method of multi-scale receptive field deep learning. Then we demonstrate our experimental settings and the results of our method, comparing it with benchmark methods, in Section 3. Finally, we interpret the findings of morphological patterns by connecting to histopathological domain knowledge in Section 4.

## 2. Methods

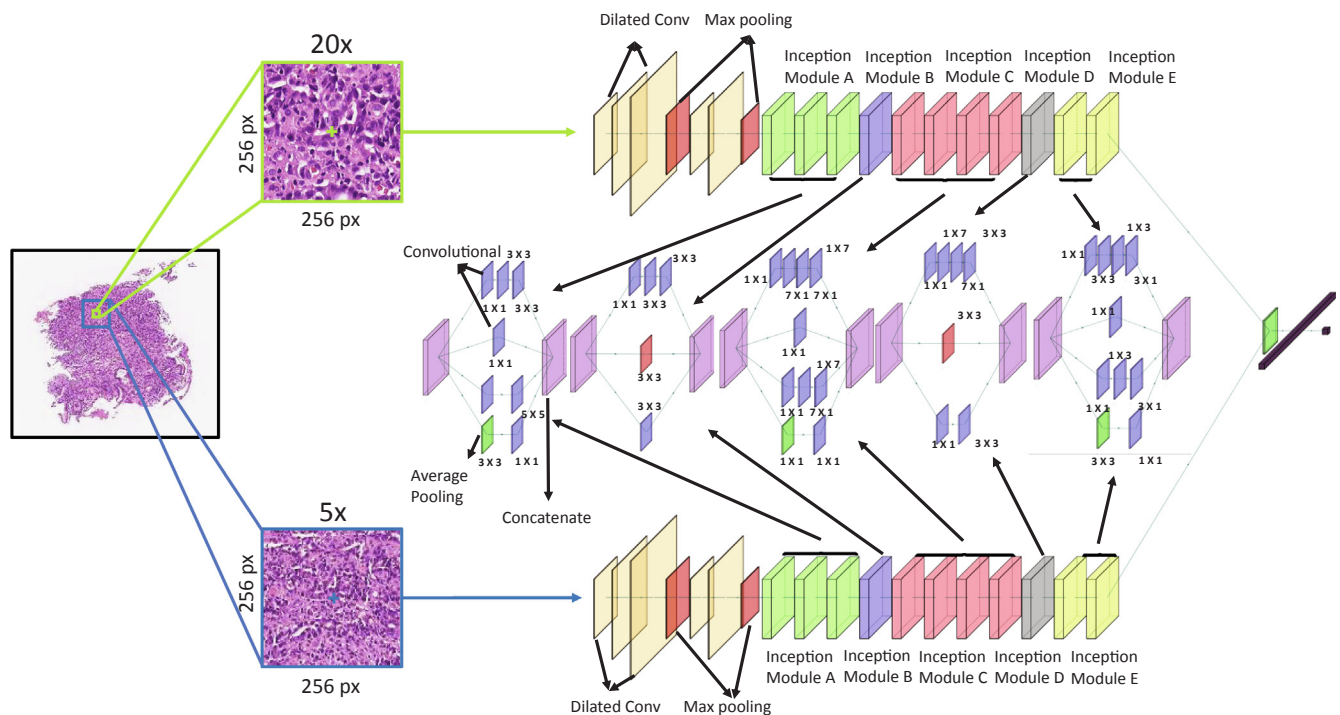
In this section, we describe our proposed method, a multi-scale receptive field deep neural network (a.k.a. Deep-Hipo, which stands for deep learning for HIstoPatholOgy) that accurately analyzes histopathological images by learning multi-scale morphological patterns from various magnification levels of patches in a WSI simultaneously. Deep-Hipo computes a cancer probability on each pixel, sliding multi-scale windows over a histopathological image, and then visualizes the probability scores to locate cancerous regions (see Fig. 1). The set of



**Fig. 1.** Overview of Deep-Hipo. (A) Deep-Hipo interprets patches of both large and small receptive fields by sliding windows over a WSI. (B) Deep-Hipo produces a heatmap that visualizes cancer classification scores in a WSI.



**Fig. 2.** Original slide and patches of cancer (top) and normal (bottom) in the magnification levels of 5 $\times$ , 10 $\times$ , and 20 $\times$ . The patches at the 5 $\times$  magnification level show glands, whereas nuclei features are shown at the 20 $\times$  magnification level.



**Fig. 3.** Architecture of Deep-Hipo. Two patches of the same pixel size are extracted in high and low magnification levels on each pixel, and the patches are introduced to the tracks of Deep-Hipo. Morphological patterns in the large and small receptive fields are learned simultaneously.

probability scores generated by Deep-Hipo on a histopathological image can be used for cancer classification, detection, and segmentation.

Cancer regions in a histopathological image are shown as a mixture of multiple, complex morphological patterns. Among the various morphological patterns, small morphological patterns (e.g., atypical nuclei) can be recognized in a small receptive field with a higher magnification level in a WSI, whereas large morphological patterns (e.g., atypical glands) that lie across large receptive fields can be observed with a lower magnification level. Fig. 2 shows an original histopathological image of a carcinoma and its patches in 5 $\times$ , 10 $\times$ , and 20 $\times$  magnification levels, where the patches on top are cancerous and those on the bottom are normal. The patches in the 5 $\times$  magnification level show

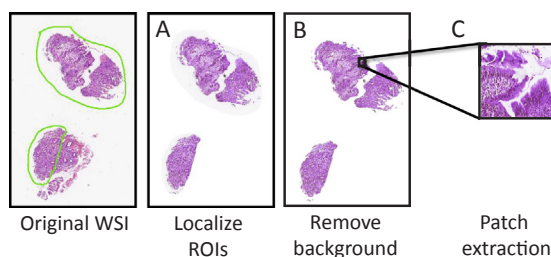
patterns of glands, whereas nuclei features are shown at the 20 $\times$  magnification level.

Deep-Hipo implements multi-scale receptive fields by taking two patch images (at high and low magnification levels) simultaneously to two tracks of convolutional neural networks, and then concatenates the last hidden layers to compute the probability of cancer on a region (see Fig. 3). One track takes a patch image at a high magnification level (e.g. 20 $\times$ ), while another track takes one at a low magnification level (e.g. 5 $\times$ ), while both patch images of the same pixel size are extracted from the same center position in a WSI. For instance, in Fig. 2, a patch of 256  $\times$  256 pixels in 5 $\times$  magnification is equivalent to one of 2,100  $\times$  2,100 pixels in 20 $\times$  magnification. However, snapshots of the same size patches at different magnification levels can capture multi-



**Table 1**  
The architecture of each track in Deep-Hipo follows CAT-Net [32].

Layer (Number)	Filter Size	Stride	Output (W × H × N)	
Dilated Conv (1)	3 × 3	2	256 × 256 × 3	
Dilated Conv (2)	3 × 3	1	149 × 149 × 32	
Dilated Conv (3)	3 × 3	1	147 × 147 × 32	
Max Pooling (4)	3 × 3	2	147 × 147 × 64	
Dilated Conv (5)	1 × 1	1	73 × 73 × 64	
Dilated Conv (6)	3 × 3	1	73 × 73 × 80	
Max Pooling (7)	3 × 3	2	71 × 71 × 192	
Inception Module A (8)	(1 × 1) (3 × 3) (5 × 5)	1	35 × 35 × 192	
Inception Module A (9)	(1 × 1) (3 × 3) (5 × 5)	1	35 × 35 × 256	
Inception Module A (10)	(1 × 1) (3 × 3) (5 × 5)	1	35 × 35 × 288	
Inception Module B (11)	(1 × 1) (3 × 3)	2	35 × 35 × 288	
Inception Module C (12, 13, 14, 15)	(1 × 1) (1 × 7)*2 (7 × 1)*2	1	17 × 17 × 768	
Inception Module D (16)	(1 × 1) (1 × 7) (7 × 1) (3 × 3)		17 × 17 × 768	
Inception Module E (17)	(1 × 1) (1 × 3) (3 × 1) (3 × 3)		8 × 8 × 1280	
Inception Module E (18)	(1 × 1) (1 × 3) (3 × 1) (3 × 3)	1	8 × 8 × 2048	
Average Pooling (19)	(1 × 1)		8 × 8 × 2048	



**Fig. 4.** The pipeline of preprocessing: (A) cropping Regions Of Interest (ROIs) if annotation is available; (B) cleaning a whole slide image (removing tissue tear/fold and overstaining and background); and (C) extracting patches.

scale morphological patterns. Thus, multi-scale morphological patterns can be extracted from the small receptive fields in the track with the high magnification level, and from the large receptive fields in the track with the low magnification level in Deep-Hipo.

Each track consists of 21 layers of dilated convolutional layers, max pooling, five Inception modules types, fully connected layers, and an output layer, following the architecture of our previous study [32] (see Fig. 3). Dilated convolutional layers are used instead of conventional convolutional layers in Deep-Hipo to capture complex morphological patterns, enlarging the field of view without a loss of spatial information [35]. The number of parameters do not increase when enlarging a

kernel size in dilated convolutional layers, which makes model training computationally efficient. Then Inception modules are adapted to recognize various sizes of morphological patterns with kernel factorization, into asymmetric convolutions. The Inception modules share a similar structure, but have different kernel sizes and numbers of convolution layers. The various-sized convolutional layers in the Inception modules are concatenated into a single output vector, which forms an input to the next layer. The last Inception modules of the two tracks are concatenated, and the flattened layer is subsequently introduced to the fully connected layer. The output layer includes only one node for binary classification, which estimates posterior probability, determining how likely a given patch is cancerous. The overall network architecture is depicted in Fig. 3, and the details of the Inception modules are elaborated in Table 1.

### 3. Experimental results

#### 3.1. Dataset

We conducted experiments with a histopathological image dataset obtained from gastroscopic biopsy specimens of 94 cases at Gyeongsang National University Changwon Hospital (Changwon, Korea) between February 2016 and July 2017 to assess the performance of the proposed method. The tissue specimens were stained with Hematoxylin and Eosin (H&E) using standard protocols in routine clinical care. Then all 94 whole slide images (WSIs) were generated.<sup>5</sup> Cancer regions were annotated on the cancer WSIs by a pathologist at Gyeongsang National University Changwon Hospital. The WSIs were manually categorized into four groups: well-differentiated (WD, 13 WSIs); moderately-differentiated (MD, 11 WSIs); poorly-differentiated adenocarcinoma (PD, 20 WSIs); and poorly cohesive carcinoma, including signet-ring cell features (PC, 20 WSIs); as well as normal gastric mucosa (30 WSIs).<sup>6</sup>

Adenocarcinoma is the most common histologic type of gastric carcinoma. Adenocarcinoma is usually composed of irregularly distended, fused or branching atypical glands of various sizes [14]. Gastric adenocarcinoma is traditionally graded as well, moderately, and poorly on the basis of gland formation. WD adenocarcinoma is composed of well-formed glands; MD adenocarcinoma comprises cancer that is intermediate, between well and poorly-differentiated; PD adenocarcinoma is composed of highly irregular glands that are identified with difficulty [18]; PC carcinoma is a unique subtype of gastric adenocarcinoma, which encompasses signet-ring cell features and other cellular variants. Signet-ring cell features are defined in a cancer composed predominantly or exclusively of signet-ring cells, characterized by a central, optically clear globoid droplet of cytoplasmic mucin, with an eccentrically placed nucleus. Other cellular variants include cancers composed of cells resembling histiocytes, lymphocytes or plasma cells [14,18].

#### 3.2. Data preprocessing

Data preprocessing was performed on the 94 WSIs to remove background and noises as well as to extract annotated regions of interests (ROIs) (see Fig. 4). Noises due to tissue tears and folds, and overstaining were removed by Gaussian blur smoothing, whereas background was filtered out by thresholding on RGB values. Then annotated ROIs were localized from the WSIs. The pre-processing was done by an open source python package, PyHistopathology.<sup>7</sup>

<sup>5</sup> This study was approved by the Institutional Review Board of Gyeongsang National University Changwon Hospital with a waiver for informed consent (2018-08-005-001).

<sup>6</sup> The histologic types and differentiation grades of the carcinomas were determined according to the classification system of the World Health Organization, fourth edition.

**Table 2**

Patch-wise performance of the benchmark methods on the testing data of the four categories (WD, MD, PD, and PC).

	Methods	Accuracy	F1 score	Precision	Recall
Well-differentiated (WD)	N-Net	0.724 (0.062)	0.709 (0.054)	0.719 (0.058)	0.702 (0.061)
	VGG-16	0.832 (0.069)	0.814 (0.065)	0.826 (0.079)	0.802 (0.063)
	Dense-Net	0.889 (0.056)	0.872 (0.048)	0.881 (0.057)	0.861 (0.044)
	Efficient-Net	0.865 (0.036)	0.852 (0.031)	0.856 (0.038)	0.848 (0.037)
	MRD-Net (20 × 10 ×)	0.838 (0.046)	0.827 (0.041)	0.832 (0.039)	0.822 (0.047)
	MRD-Net (20 × 5 ×)	0.878 (0.036)	0.864 (0.038)	0.861 (0.031)	0.822 (0.035)
	GB-INCV3	0.908 (0.035)	0.904 (0.038)	0.901 (0.035)	0.907 (0.041)
	CAT-Net	0.926 (0.024)	0.924 (0.027)	0.921 (0.029)	0.929 (0.027)
	Deep-Hipo	<b>0.951 (0.021)</b>	<b>0.949 (0.019)</b>	<b>0.947 (0.018)</b>	<b>0.953 (0.023)</b>
Moderately-differentiated (MD)	N-Net	0.716 (0.049)	0.704 (0.067)	0.714 (0.056)	0.695 (0.062)
	VGG-16	0.821 (0.076)	0.805 (0.065)	0.813 (0.067)	0.797 (0.073)
	Dense-Net	0.878 (0.052)	0.862 (0.055)	0.868 (0.049)	0.853 (0.046)
	Efficient-Net	0.852 (0.034)	0.843 (0.037)	0.846 (0.029)	0.839 (0.034)
	MRD-Net (20 × 10 ×)	0.829 (0.051)	0.821 (0.044)	0.824 (0.043)	0.817 (0.046)
	MRD-Net (20 × 5 ×)	0.878 (0.036)	0.864 (0.038)	0.861 (0.031)	0.822 (0.035)
	GB-INCV3	0.898 (0.043)	0.896 (0.041)	0.894 (0.049)	0.901 (0.043)
	CAT-Net	0.919 (0.023)	0.915 (0.025)	0.912 (0.026)	0.918 (0.027)
	Deep-Hipo	<b>0.949 (0.024)</b>	<b>0.945 (0.026)</b>	<b>0.939 (0.024)</b>	<b>0.950 (0.016)</b>
Poorly-differentiated (PD)	N-Net	0.689 (0.055)	0.672 (0.058)	0.684 (0.066)	0.665 (0.057)
	VGG-16	0.784 (0.078)	0.759 (0.068)	0.765 (0.072)	0.751 (0.073)
	Dense-Net	0.859 (0.052)	0.845 (0.045)	0.851 (0.052)	0.836 (0.047)
	Efficient-Net	0.827 (0.033)	0.821 (0.037)	0.824 (0.029)	0.811 (0.026)
	MRD-Net (20 × 10 ×)	0.806 (0.042)	0.789 (0.051)	0.795 (0.047)	0.784 (0.044)
	MRD-Net (20 × 5 ×)	0.839 (0.043)	0.831 (0.037)	0.826 (0.041)	0.836 (0.035)
	GB-INCV3	0.874 (0.038)	0.868 (0.044)	0.865 (0.034)	0.872 (0.039)
	CAT-Net	0.898 (0.021)	0.893 (0.023)	0.891 (0.025)	0.901 (0.021)
	Deep-Hipo	<b>0.918 (0.029)</b>	<b>0.915 (0.031)</b>	<b>0.908 (0.021)</b>	<b>0.921 (0.026)</b>
Poorly-Cohesive (PC)	N-Net	0.698 (0.049)	0.681 (0.044)	0.694 (0.043)	0.673 (0.055)
	VGG-16	0.798 (0.075)	0.772 (0.072)	0.786 (0.067)	0.763 (0.069)
	Dense-Net	0.868 (0.049)	0.851 (0.045)	0.855 (0.043)	0.843 (0.048)
	Efficient-Net	0.831 (0.039)	0.825 (0.035)	0.828 (0.033)	0.818 (0.038)
	MRD-Net (20 × 10 ×)	0.812 (0.052)	0.795 (0.055)	0.804 (0.053)	0.787 (0.047)
	MRD-Net (20 × 5 ×)	0.847 (0.046)	0.839 (0.044)	0.833 (0.045)	0.839 (0.037)
	GB-INCV3	0.881 (0.046)	0.871 (0.038)	0.869 (0.042)	0.882 (0.044)
	CAT-Net	0.901 (0.025)	0.899 (0.022)	0.895 (0.033)	0.904 (0.029)
	Deep-Hipo	<b>0.931 (0.019)</b>	<b>0.929 (0.023)</b>	<b>0.922 (0.019)</b>	<b>0.934 (0.022)</b>
Overall	N-Net	0.707 (0.053)	0.691 (0.062)	0.703 (0.061)	0.684 (0.057)
	VGG-16	0.809 (0.074)	0.787 (0.061)	0.797 (0.064)	0.778 (0.068)
	Dense-Net	0.874 (0.043)	0.857 (0.047)	0.864 (0.048)	0.848 (0.041)
	Efficient-Net	0.844 (0.035)	0.835 (0.039)	0.839 (0.032)	0.829 (0.033)
	MRD-Net (20 × 10 ×)	0.821 (0.046)	0.808 (0.052)	0.814 (0.057)	0.802 (0.043)
	MRD-Net (20 × 5 ×)	0.859 (0.034)	0.848 (0.034)	0.843 (0.035)	0.853 (0.038)
	GB-INCV3	0.890 (0.041)	0.885 (0.045)	0.882 (0.043)	0.891 (0.048)
	CAT-Net	0.911 (0.031)	0.908 (0.031)	0.905 (0.027)	0.913 (0.025)
	Deep-Hipo	<b>0.937 (0.032)</b>	<b>0.934 (0.027)</b>	<b>0.929 (0.032)</b>	<b>0.939 (0.028)</b>

Note that the average and standard deviations (in parenthesis) of the twenty experiments are shown, where a bold-face indicates the highest average on the experiments. Wilcoxon rank-sum tests between Deep-Hipo and CAT-Net of the second highest performance were performed. Statistically significant p-values ( $p < 0.0001$ ) were shown in all the experiments.

The total numbers of 179 and 164 tissue samples were obtained from 64 cancer WSIs and 30 control WSIs, respectively. We randomly split the samples into training (80%) and test data (20%), preserving the tissue number ratios in the four categories. Then the training data was split again into training (80%) and validation (20%). On each experiment, 219, 55, and 69 tissue samples were available for training, validation, and test data, respectively. After splitting the dataset, non-overlapping patches of  $256 \times 256$  pixels were extracted in both  $20 \times$  and  $5 \times$  magnification levels, where each patch shared a common center point. Finally, we obtained approximately 80 k cancer patches, and another 80 k normal patches at each magnification level.

### 3.3. Experiments

We compared the performance of Deep-Hipo with current state-of-the-art benchmark methods, repeating the experiment twenty times for reproducibility of performance. The benchmark methods included N-Net [27], VGG16 [34], Dense-Net [15], Efficient-Net [30], Multi-Resolution Deep learning network (MRD-Net) [2], Google Brain's Inception V3 (GB-INCV3) [20], and CAT-Net [32]. Note that we did not consider image segmentation methods that require pixel-based annotation for the comparison. The experimental settings for the benchmark methods were as follows:

- **N-Net:** N-Net was trained with the same size patches of  $256 \times 256$  pixels, and the output layer was replaced by one node with sigmoid activation to compute a probability of cancer. N-Net was implemented using Keras in Python with an Adaptive Moment Estimation (ADAM) optimizer.

<sup>7</sup> <http://dataxlab.org/pyhistopathology>.

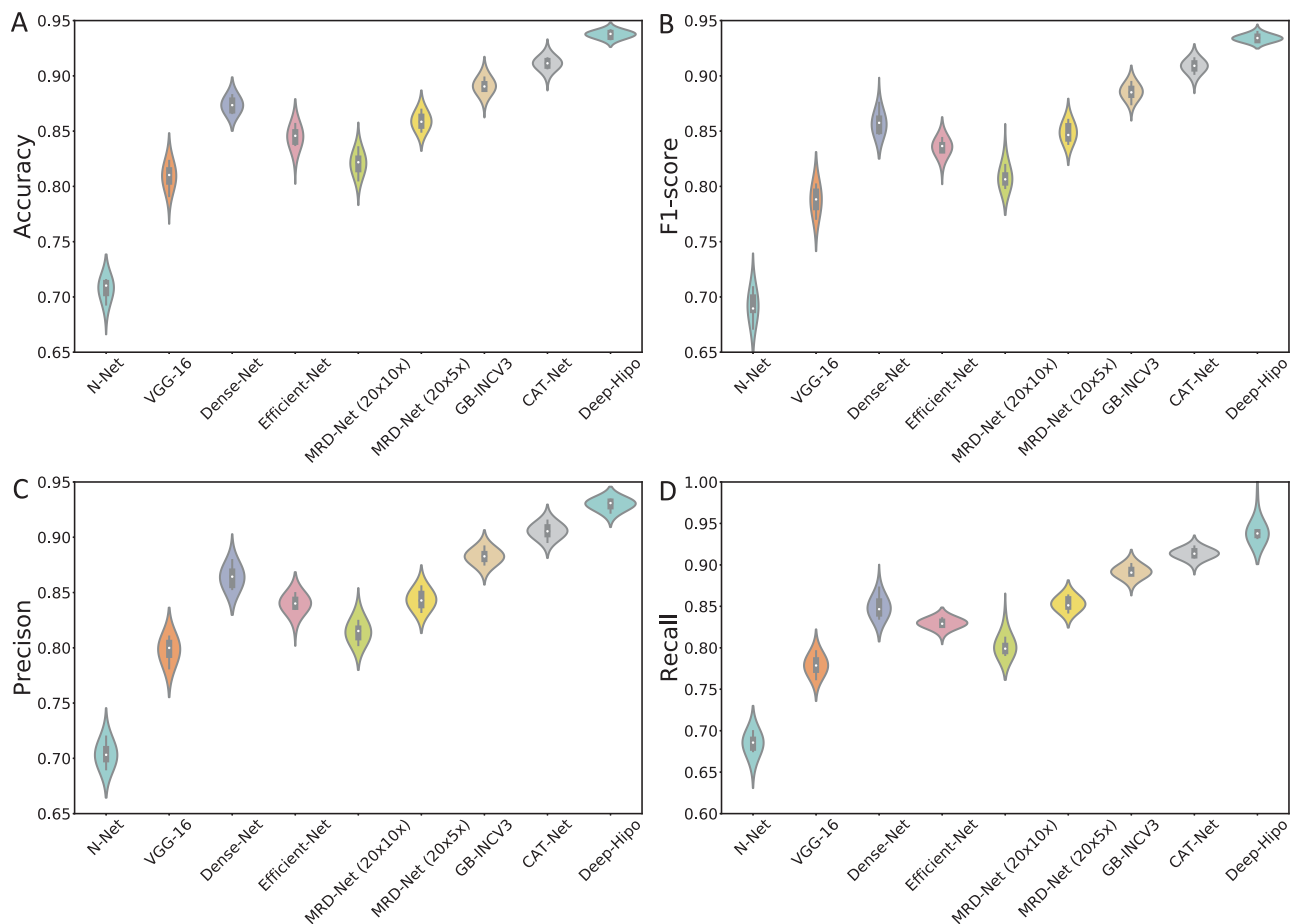


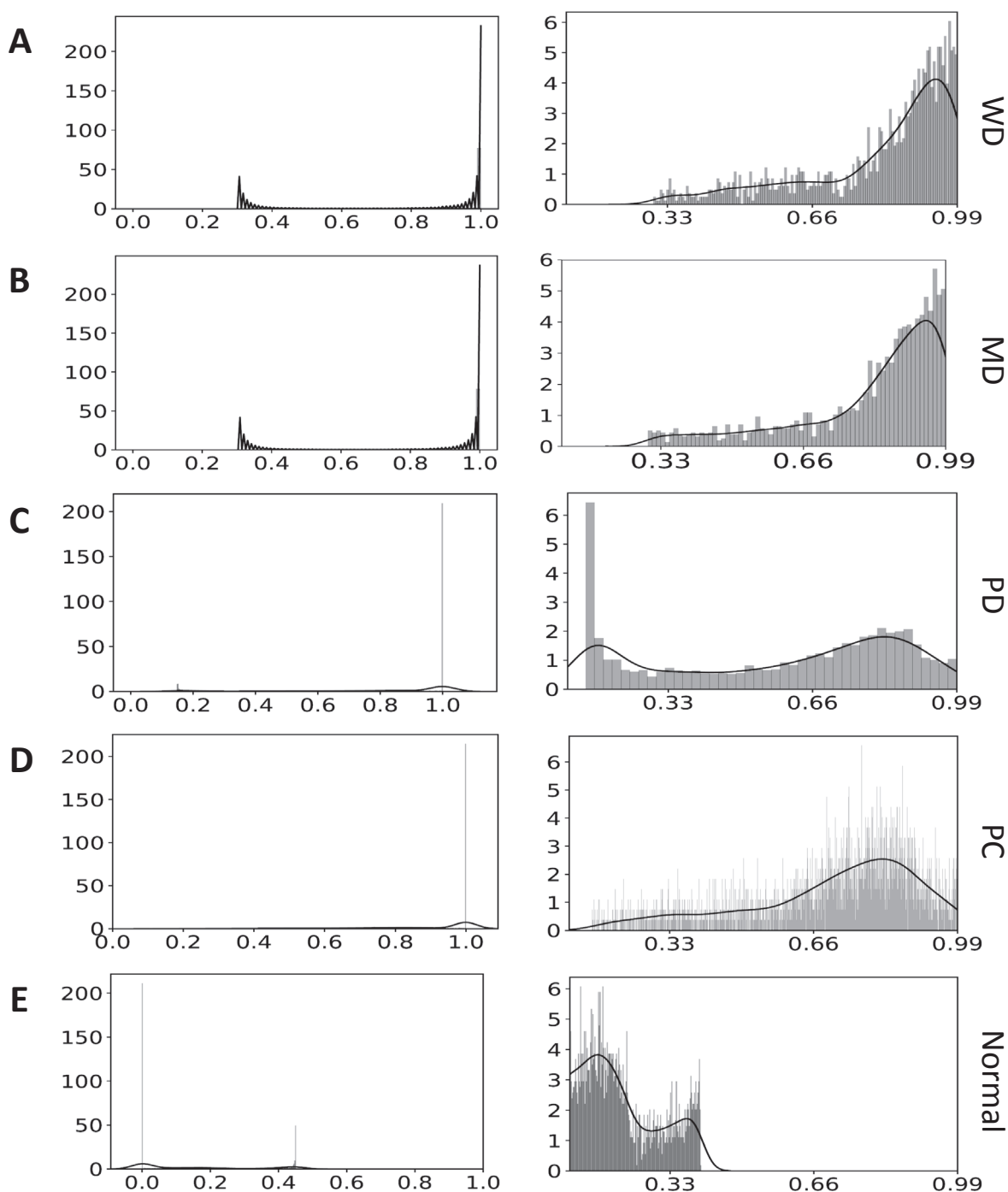
Fig. 5. Overall patch-wise performance of the benchmark methods: (A) Accuracy, (B) F1-score, (C) Precision, and (D) Recall.

- **VGG16:** VGG16 was trained with patches of  $256 \times 256$  pixels using Keras in Python with a Stochastic Gradient Descent (SGD) optimizer.
- **Dense-Net:** DenseNet121 was modified by replacing the output layer with one sigmoid activation node, and the model was trained using patches of  $256 \times 256$  pixels. DenseNet was implemented using Keras in Python with ADAM.
- **Efficient-Net:** The output layer of EfficientNetB3 architecture was replaced with one sigmoid activation node, and patches of  $256 \times 256$  pixels were used to train and validate the model. Efficient-Net was implemented using Keras in Python with ADAM.
- **MRD-Net ( $20 \times 10 \times$ ):** MRD-Net ( $20 \times 10 \times$ ) took multi-scale patches of  $224 \times 224 \times 3$  pixels at both  $20 \times$  and  $10 \times$  magnification levels, and then combined them to form a new input of dimensions  $224 \times 224 \times 6$ . The multi-channel images were used to train with ResNet. MRD-Net was implemented in Keras in Python with ADAM, where the output layer was replaced with sigmoid activation node.
- **MRD-Net ( $20 \times 5 \times$ ):** MRD-Net ( $20 \times 5 \times$ ) considered patch images at  $20 \times$  and  $5 \times$  magnification levels.
- **GB-INCv3:** A non-overlapping kernel window of  $299 \times 299$  pixels, instead of  $256 \times 256$  pixels, was slid on the WSIs to generate patches as proposed in [20], so the number of patches in the test data was fewer than the other benchmarks, since patch size was relatively larger. GB-INCv3 was implemented by its open source code in PyTorch, where an Adaptive Moment Estimation (ADAM) optimizer was used.
- **CAT-Net:** CAT-Net was implemented in Keras. In CAT-Net, an SGD optimizer was used that produced better performance than ADAM.
- **Deep-Hipo:** Deep-Hipo considered two tracks, with patches of  $5 \times$  and  $20 \times$  magnification levels, that share the same center points. Justification of optimal magnification levels of the two tracks is

discussed in Section 4. Deep-Hipo was implemented in Keras, where ADAM was used.

The output layers of all benchmark models were replaced by one node with the activation function of sigmoid to compute the probability of cancer. The optimal hyper-parameters of learning rate, dropout, weight decay, and optimizer (ADAM or sophisticated gradient decent (SGD)) were obtained by a grid search with training and validation data, with minimum validation loss on each experiment for all of the models. For CAT-Net and Deep-Hipo, we used batch normalization to achieve faster training, and drop-out to avoid overfitting. For Deep-Hipo, the most common optimal hyper-parameters in the twenty experiments of learning rate, dropout and weight decay, from the grid search were  $1e-5$ ,  $0.3$  and  $0.02$  respectively.

We measured patch-based accuracy, precision, recall, and F1-score on the test data of the twenty experiments. Note that the models were trained with the training data evenly distributed to the four categories, and the performance was measured with the test data on each category separately. As shown in Table 2 and Fig. 5, Deep-Hipo outperformed the other eight benchmark methods on all categories of cancer tissues. Specifically, Deep-Hipo produced the highest accuracies of  $0.951 \pm 0.0021$  (mean  $\pm$  standard deviation),  $0.949 \pm 0.0024$ ,  $0.918 \pm 0.0029$ , and  $0.931 \pm 0.0019$  for all of the categories of WD, MD, PD, and PC, respectively. Deep-Hipo achieved an overall accuracy of  $0.937 \pm 0.0032$ , which is 2.6% higher than the second best model, CAT-Net. Furthermore, Deep-Hipo also achieved highest overall F1-score, precision, and recall of  $0.934 \pm 0.0027$ ,  $0.929 \pm 0.0032$ , and  $0.939 \pm 0.0028$ , respectively. The significant performance was statistically assessed by a Wilcoxon rank-sum test, compared with a CAT-Net of the second highest performance. The p-values for overall accuracy



**Fig. 6.** The histogram of the probability scores, generated by Deep-Hipo, on the four categories of (A) WD, (B) MD, (C) PD, (D) PC, and (E) normal tissues.

**Table 3**

Performance comparison with various magnification levels in the two tracks in Deep-Hipo.

Small receptive field	Large receptive field	Accuracy	F1-score	Precision	Recall
20 ×	5 ×	<b>0.937 (0.032)</b>	<b>0.934 (0.027)</b>	<b>0.929 (0.032)</b>	<b>0.939 (0.028)</b>
20 ×	10 ×	0.919 (0.043)	0.910 (0.033)	0.903(0.0039)	0.917 (0.0028)
10 ×	5 ×	0.858 (0.053)	0.853 (0.057)	0.845 (0.052)	0.859 (0.048)

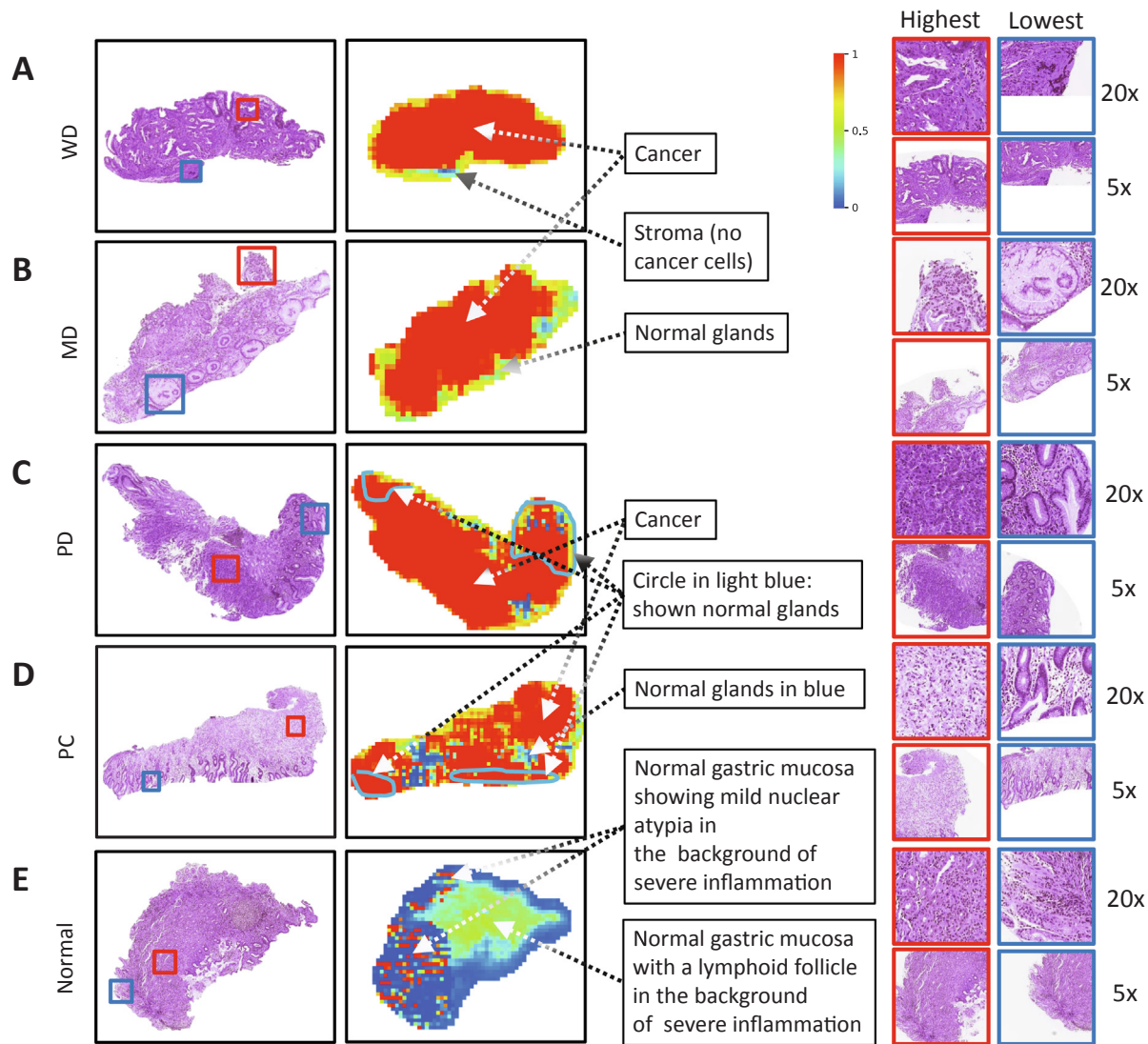
( $p = 6.54e-5$ ), F1-score ( $p = 8.05e-5$ ), precision ( $p = 1.62e-6$ ), and recall ( $p = 7.41e-5$ ) were lower than 0.0001.

The performance of Deep-Hipo in the four categorical data groups is well aligned to the difficulty levels of manual interpretation in histopathology, where the accuracies of WD and MD were 1.8–3.3% higher

than PD and PC. Well- and moderately-differentiated adenocarcinomas consist of well-formed glands in more than 50% of cancer. However, poorly-differentiated adenocarcinomas have poorly formed glands and may form solid areas or single cells with barely recognizable glands, and poorly cohesive carcinomas, including signet-ring cell features,

**Table 4**  
Performance comparison between dilated and conventional convolutional layers in Deep-Hipo.

Architecture	Accuracy	F1-score	Precision	Recall
Dilated convolutional layers	<b>0.937 (0.032)</b>	<b>0.934 (0.027)</b>	<b>0.929 (0.032)</b>	<b>0.939 (0.028)</b>
Conventional convolutional layers	0.894 (0.038)	0.886 (0.037)	0.875(0.0043)	0.897 (0.036)



**Fig. 7.** Examples of microscopy images (leftmost) from the test data in the four categories of (A) WD, (B) MD, (C) PD, and (D) PC as well as (E) normal and their heatmap results (second left) by Deep-Hipo. The patches in the right side show the ones with the highest (border in red) and lowest (border in blue) probability of cancer in the tissues. The position of the patches of the highest and lowest probability are pointed out by two boxes in red and blue, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

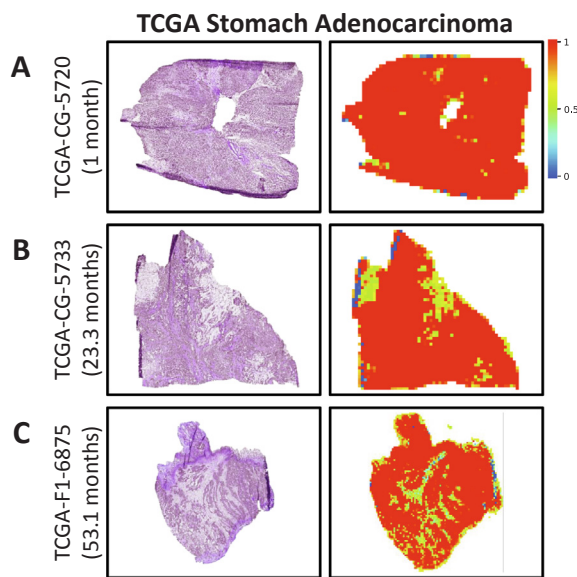
show scattered single cells or small aggregated cells without gland formation. Thus, Deep-Hipo is more accurate in well and moderately-differentiated adenocarcinomas than in poorly-differentiated adenocarcinomas and poorly cohesive carcinomas [18].

We examined the cancer probability distributions produced by Deep-Hipo on the test data in the four categories of (A) WD, (B) MD, (C) PD, (D) PC, and (E) normal tissues (see Fig. 6). The optimally trained Deep-Hipo produced a set of cancer probabilities on the patches over the test data, and the histograms were generated with those probabilities. The figures on the right side are the histograms, excluding the patches of zero and one cancer probabilities, while the patches dominated the original histograms are on the left side. The probability

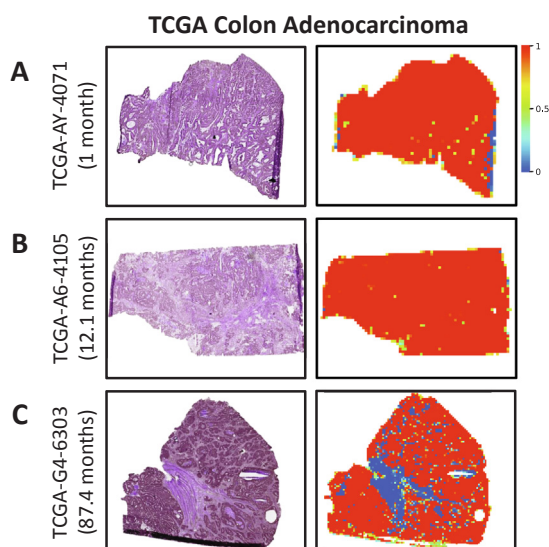
distributions of WD and MD showed similarity, but WD had less gland fusion and cribriform formation than moderately-differentiated adenocarcinomas. Poorly-differentiated adenocarcinomas and poorly-cohesive carcinomas, including signet-ring cell features, may show probabilities seen as either inflammation or cancer. This may be because they lack gland formation and infiltrate normal tissue in the form of individual cells.

In this study, we considered 5 × and 20 × magnification levels for the large and small receptive fields in Deep-Hipo, respectively. The optimal magnification levels of the two tracks were chosen by the following experiments. We compared accuracy, precision, recall, and F1-scores of Deep-Hipo with all the available combinations of





**Fig. 8.** Heatmaps of TCGA Stomach Adenocarcinoma (STAD). (A) Adenocarcinoma, poorly-differentiated, stroma (no cancer cells) in green. (B) Mucinous adenocarcinoma, a few atypical cells in the mucin pools in green. (C) Adenocarcinoma, poorly-differentiated, stroma (no cancer cells) in green.



**Fig. 9.** Heatmaps of TCGA Colon Adenocarcinoma (COAD). (A) Adenocarcinoma, well-differentiated, tissue folds in blue, (B) Adenocarcinoma, moderately-differentiated, (C) Adenocarcinoma, moderately-differentiated, stroma (no cancer cells) in blue and tissue folds in green.

magnification levels for the two tracks. The highest performance,  $0.937 \pm 0.0032$ ,  $0.934 \pm 0.0027$ ,  $0.929 \pm 0.0032$ , and  $0.939 \pm 0.0028$  on accuracy, F1-score, precision, and recall, respectively, was achieved with  $20\times$  and  $5\times$  magnification levels (see Table 3). The lower performance, with low =  $10\times$  and high =  $5\times$ , and low =  $20\times$  and high =  $10\times$ , may show that small morphological patterns are not recognized well in the  $10\times$  magnification level as the small receptive field, whereas large morphological patterns are also not captured in the  $10\times$  magnification level as the large receptive field, respectively.

Deep-Hipo adapted dilated convolutional layers, instead of conventional convolutional layers, for effectively capturing morphological patterns. We compared the performance of Deep-Hipo using dilated or conventional convolutional layers (see Table 4). Deep-Hipo with dilated convolutional layers showed an accuracy of 4% higher than

conventional convolutional layers.

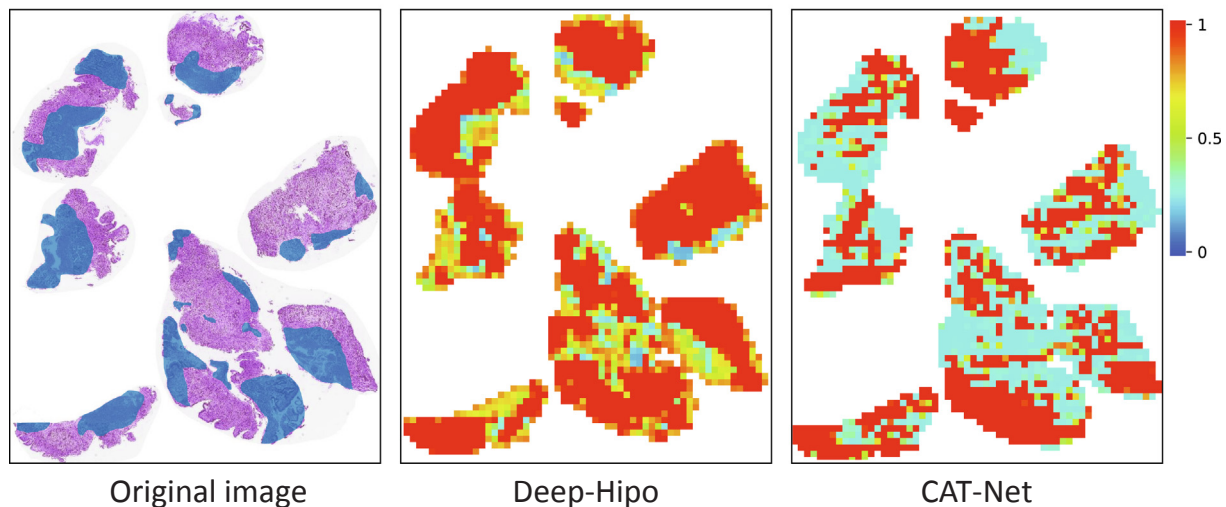
#### 4. Discussion

In order to interpret the morphological pattern findings by connecting them to histopathological domain knowledge, we trained the proposed model again using all available datasets with the most common hyper-parameters (i.e., learning rate:  $1e-5$ ; dropout: 0.3; weight decay: 0.02). The optimally trained model was applied to the cancer tissues of the four categories and the normal tissues. Note that the dataset was weakly supervised, where cancer regions were annotated as broad regions of cancer instead of pixel-based annotation (see Fig. 4). Thus, some regions of cancer tissues may contain normal glands, whereas some normal tissues may show some inflammation.

Fig. 7 illustrates representative examples of the microscopy images (leftmost) in the four categories, as well as normal tissues, and their heatmap results (second left) from Deep-Hipo. The patches on the right side show the those with the highest (border in red) and lowest (border in blue) probabilities of cancer in the tissues. The patch positions of the highest and lowest probabilities are shown by two boxes (based on  $20\times$ ) in red and blue, respectively. The various-sized boxes of the patches depend on the tissue size. Most cancer regions in (A)–(D) of Fig. 7 are recognized as cancer (in red) by Deep-Hipo. Stroma, which are not cancer cells as well as normal glands, were correctly recognized by Deep-Hipo in the four types of gastric carcinoma. In PD and PC, some normal glands, circled in light blue, were misclassified as cancer. In the normal tissue of Fig. 7(E), Deep-Hipo correctly identified regions of inflammation, colored in blue or light green. The regions in blue (left side of the tissue) with scattered red contain normal glands showing mild atypia in the background of severe inflammation, whereas the regions in light green (centered in the tissue) show normal gastric mucosa with a lymphoid follicle in the background of severe inflammation.

We, furthermore, assessed our proposed model with histo-pathological images of The Cancer Genome Atlas (TCGA) Stomach Adenocarcinoma (TCGA-STAD) and TCGA Colon Adenocarcinoma (TCGA-COAD), which show similar pathological patterns with gastric carcinoma, as the external cross-validation (see Fig. 8). In TCGA-STAD, the cancer tissues obtained from three patients of short- (TCGA-CG-5720, 1 month), medium- (TCGA-CG-5733, 23.3 months), and long-term (TCGA-F1-6875, 53.1 months) survival were examined. TCGA-CG-5720 and TCGA-F1-6875 were classified as poorly-differentiated adenocarcinoma, whereas TCGA-CG-5733 was classified as mucinous adenocarcinoma. In TCGA-CG-5720 and TCGA-F1-6875, stroma (no cancer cells) were recognized by Deep-Hipo in green, while a few atypical cells in the mucin pools were captured in green in TCGA-CG-5733. Similarly, TCGA-COAD tissues from the three patients of TCGA-AY-4071 (survival month = 1), TCGA-A6-4105 (survival = 12.1), and TCGA-G4-6303 (survival = 87.4) were investigated (see Fig. 9). In TCGA-G4-6303, stroma (no cancer cells) were correctly observed in blue, while most cancer regions were also correctly captured in red.

Deep-Hipo incorporates the two tracks of CAT-Net [32] for large and small receptive fields of histopathological images. Note that CAT-Net analyzes only local patterns of a patch independently, whereas Deep-Hipo captures global patterns across patches in  $20\times$ , as well as taking in small receptive fields. Fig. 10 depicts the heatmaps of Deep-Hipo and CAT-Net, as well as an original slide of a poorly cohesive carcinoma. In the original histopathological image (leftmost), the annotated regions in blue mainly consist of normal gastric mucosa, without cancer cell infiltration, whereas the others are cancerous. Deep-Hipo showed a few regions classified as false positive (classified as cancer but actually normal), but most regions are correctly classified as true positive. In contrast, CAT-Net produced much larger regions of false negatives (classified as normal but actually cancer) and false positives than Deep-Hipo. Moreover, Deep-Hipo produces not only higher accuracy, but also smoother probability scores than CAT-Net, for



**Fig. 10.** Performance comparison between Deep-Hipo and CAT-Net. In the original histopathological image of poorly cohesive carcinoma (leftmost), the annotated regions in blue mainly consist of normal gastric mucosa without cancer cells infiltration, whereas the others are cancerous. Deep-Hipo showed a few regions classified as false positive, but most regions are correctly classified as true positive. In contrast, CAT-Net produced much larger regions of false negatives and false positives than Deep-Hipo.

connected neighbors are overlapped for the analysis by a large receptive field in Deep-Hipo.

## 5. Conclusion

In this paper, we developed a novel deep learning model for histopathology, named Deep-Hipo, that implements multi-scale receptive fields from multiple patches in multi-scale magnification levels from a histopathological image. The exploration of multi-scale patches at various magnification levels is consistent with pathologists examining tissues at various magnification levels, while checking complex morphological patterns in a microscope.

The experimental results of cross-validation showed statistically significant performance of Deep-Hipo in accuracy, F1-score, precision, and recall, and the external assessment with TCGA stomach adenocarcinoma and TCGA colon adenocarcinoma WSIs verified its performance.

In this study, we have not considered  $40\times$  magnification levels, due to the extreme computational cost and the lack of a  $40\times$  magnification level in many TCGA WSIs. However, the performance comparison of Deep-Hipo architectures: (1) two tracks with  $5\times$  and  $40\times$  magnification levels; or (2) multiple tracks, e.g. with  $5\times$ ,  $20\times$ , and  $40\times$  magnification levels, would be desired for the future work.

The heatmaps in the study were generated by shifting a patch window by 16 pixels. If smaller stride (the amount by which a patch window shifts) is considered, more patches will be generated, and the quality of the heatmap will be improved. However, generating a heatmap is computationally expensive due to huge numbers of patch extractions. For real-time analyses, parallel computation can be a potential solution.

In this study, we developed the proposed method, Deep-Hipo, for histopathological image analysis that assigns a cancer probability on each patch. However, the multi-scale receptive field deep learning model can be potentially extended for the other histopathological analyses, such as survival analysis, patient risk prediction, and cancer subtype classification, by reconstructing the output layer.

## CRediT authorship contribution statement

MK and HK designed and supervised the overall research project. SK mainly developed and implemented the proposed algorithms and

performed the experiments. SK mainly wrote the manuscript under the supervision of MK. JH and SK developed preprocessing software of the histopathological image datasets, and JH mainly conducted quality control of the project. HK generated the histopathological image dataset and discussed about the pathological interpretation. All authors have read and approved the final version of the manuscript.

## Acknowledgments

This research was supported in part by the Ministry of Science, ICT, Korea, under the High-Potential Individuals Global Training Program (2019001601), supervised by the Institute for Information & Communications Technology Planning & Evaluation (IITP).

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.ymeth.2020.05.012>.

## References

- [1] M.Z. Alom, T. Aspiras, T.M. Taha, V.K. Asari, T. Bowen, D. Billiter, S. Arkell, Advanced deep convolutional neural network approaches for digital pathology image analysis: a comprehensive evaluation with different use cases, , 2019.
- [2] N. Alsubaie, M. Shaban, D. Snead, A. Khurram, N. Rajpoot, A multi-resolution deep learning framework for lung adenocarcinoma growth pattern classification, in: M. Nixon, S. Mahmoodi, R. Zwigglar (Eds.), *Medical Image Understanding and Analysis*, Springer International Publishing, Cham, 2018, pp. 3–11.
- [3] A.D. Belsare, M.M. Mushrif, M.A. Pangarkar, N. Meshram, Classification of breast cancer histopathology images using texture feature analysis, in: *TENCON 2015–2015 IEEE Region 10 Conference*, 2015, pp. 1–5.
- [4] K. Bera, K.A. Schalper, D.L. Rimm, V. Velcheti, A. Madabhushi, Artificial intelligence in digital pathology – new tools for diagnosis and precision oncology, *Nat. Rev. Clin. Oncol.* (2019).
- [5] G. Campanella, M.G. Hanna, L. Geneslaw, et al., Clinical-grade computational pathology using weakly supervised deep learning on whole slide images, 2019.
- [6] L. Chen, G. Papandreou, F. Schroff, H. Adam, Rethinking atrous convolution for semantic image segmentation, *CoRR* (2017) URL: <http://arxiv.org/abs/1706.05587>.
- [7] K. Fan, S. Wen, Z. Deng, Deep learning for detecting breast cancer metastases on WSI, in: *Smart Innovation, Systems and Technologies*, 2019.
- [8] Y. Gao, L. Zhang, Y. Wang, J. Jiang, Multi-scale receptive field neural networks for object tracking, *Int. Congress Ser.* 1291 (2006) 229–232.
- [9] Z. Guo, L.H.N.H., et al., A fast and refined cancer regions segmentation framework in whole-slide breast pathological images, 2019.
- [10] Z. Guo, H. Liu, H. Ni, X. Wang, M. Su, W. Guo, K. Wang, T. Jiang, Y. Qian, A fast and refined cancer regions segmentation framework in whole-slide breast pathological

- images, *Scientific Rep.* 9 (1) (Dec 2019).
- [11] J. Hao, S.C. Kosaraju, N.Z. Tsaku, D.H. Song, M. Kang, PAGE-Net: Interpretable and Integrative Deep Learning for Survival Analysis Using Histopathological Images and Genomic Data. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 2020.
  - [12] D. He, K. Xu, D. Wang, Design of multi-scale receptive field convolutional neural network for surface inspection of hot rolled steels, *Image Vis. Comput.* 89 (2019) 12–20 URL: <http://www.sciencedirect.com/science/article/pii/S0262885619300915>.
  - [13] L. Hou, D. Samaras, T.M. Kurc, Y. Gao, J.E. Davis, J.H. Saltz, Patch-based convolutional neural networks, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2424–2433.
  - [14] B. Hu, N.E. Hajj, S. Sittler, N. Lammert, R. Barnes, A. Meloni-Ehrig, Gastric cancer: classification, histology and application of molecular pathology, 2012.
  - [15] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2261–2269.
  - [16] A. Janowczyk, A. Madabhushi, Deep learning for digital pathology image analysis: a comprehensive tutorial with selected use cases, *J. Pathol. Inf.* 7 (1) (2016) 29.
  - [17] K.J. Kaplan, L.K. Rao, Digital pathology: historical perspectives, current concepts and future applications, 2016.
  - [18] G. Lauwers, F. Carneiro, D.Y. Graham, WHO Classification of Tumours of the Digestive System – 3rd Chapter. WHO Classification of Tumours of the Digestive System, fourth ed., IARC, Lyon, 2010.
  - [19] G. Litjens, C.I. Sánchez, N. Timofeeva, M. Hermsen, I. Nagtegaal, I. Kovacs, C. Hulsbergen-Van De Kaa, P. Bult, B. Van Ginneken, J. Van Der Laak, Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis, *Scientific Rep.* (2016).
  - [20] Y. Liu, et al., 2017. Detecting cancer metastases on gigapixel pathology images. CoRR. URL: <http://arxiv.org/abs/1703.02442>.
  - [21] X. Luo, X. Zang, L. Yang, J. Huang, F. Liang, J. Rodriguez-Canales, I.I. Wistuba, A. Gazdar, Y. Xie, G. Xiao, Comprehensive computational pathological image analysis predicts lung cancer prognosis, *J. Thoracic Oncol.* (2017).
  - [22] K. Nguyen, A. Sarkar, A.K. Jain, Prostate cancer grading: Use of graph cut and spatial arrangement of nuclei, *IEEE Trans. Med. Imag.* (2014).
  - [23] K. Oikawa, et al., Pathological diagnosis of gastric cancers with a novel computerized analysis system, *J. Pathol. Inf.* (2017).
  - [24] B. Peng, L. Chen, M. Shang, J. Xu, Fully convolutional neural networks for tissue histopathology image classification and segmentation, in: 2018 25th IEEE International Conference on Image Processing (ICIP), 2018, pp. 1403–1407.
  - [25] E. Radiya-Dixit, D. Zhu, A.H. Beck, Automated classification of benign and malignant proliferative breast lesions, *Scientific Rep.* (2017).
  - [26] J. Redmon, A. Farhadi, Yolov3. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2017.
  - [27] F. Sheikhzadeh, et al., Automatic labeling of molecular biomarkers of immunohistochemistry images using fully convolutional networks, *PLoS One* (2018).
  - [28] K. Sirinukunwattana, S.E.A. Raza, Y. Tsang, D.R.J. Snead, I.A. Cree, N.M. Rajpoot, Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images, *IEEE Trans. Med. Imag.* 35 (5) (May 2016) 1196–1206.
  - [29] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, CoRR (2015) URL: <http://arxiv.org/abs/1512.00567>.
  - [30] M. Tan, Q.V. Le, Efficient net: rethinking model scaling for convolutional neural networks, 2019.
  - [31] H. Tokunaga, Y. Teramoto, A. Yoshizawa, R. Bise, Adaptive weighting multi-field-of-view cnn for semantic segmentation in pathology, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 12589–12598.
  - [32] N.Z. Tsaku, S.C. Kosaraju, T. Aqila, M. Masum, D.H. Song, A.M. Mondal, H.M. Koh, M. Kang, Texture-based deep learning for effective histopathological cancer image classification, in: 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2019, pp. 973–977.
  - [33] B.S. Veeling, J. Linmans, J. Winkens, T. Cohen, M. Welling, Rotation equivariant cnns for digital pathology, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, Springer International Publishing, Cham, 2018, pp. 210–218.
  - [34] D. Wang, A. Khosla, R. Gargya, H. Irshad, A.H. Beck, Deep learning for identifying metastatic breast cancer, 2016. arXiv preprint arXiv:1606.05718.
  - [35] P. Wang, P. Chen, Y. Yuan, D. Liu, Z. Huang, X. Hou, G. Cottrell, Understanding convolution for semantic segmentation, 2018 IEEE Winter Conference on Applications of Computer Vision (WACV) (2018) 1451–1460.
  - [36] S. Wang, D.M. Yang, R. Rong, X. Zhan, J. Fujimoto, H. Liu, J. Minna, I.I. Wistuba, Y. Xie, G. Xiao, Artificial intelligence in lung cancer pathology image analysis, 2019.
  - [37] S. Wang, D.M. Yang, R. Rong, X. Zhan, G. Xiao, Pathology Image Analysis Using Segmentation Deep Learning Algorithms, 2019.
  - [38] J.W. Wei, L.J. Tafe, Y.A. Linnik, et al., Pathologist-level classification of histologic patterns on resected lung adenocarcinoma slides with deep neural networks, *Sci. Rep.* 9 (2019) 3358.
  - [39] Z. Yan, Y. Zhan, S. Zhang, D. Metaxas, X.S. Zhou, Multi-instance multi-stage deep learning for medical image recognition, in: *Deep Learning for Medical Image Analysis*, 2017.
  - [40] S. You, Y. Sun, L. Yang, J. Park, H. Tu, M. Marjanovic, S. Sinha, S.A. Boppart, Real-time intraoperative diagnosis by deep neural network driven multiphoton virtual histology, *NPJ Precis. Oncol.* (2019).
  - [41] X. Zhu, J. Yao, F. Zhu, J. Huang, Wsisa: making survival prediction from whole slide histopathological images, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 6855–6863.