



INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X

Impact Factor: 6.078

(Volume 10, Issue 2 - V10I2-1146)

Available online at: <https://www.ijariit.com>

Multi-scale Deep Learning for Histopathological Image Analysis: The Deep-Hipo Approach

Keerthana M.

km@gitam.in

GITAM (Deemed To Be University), Rudraram, GITAM (Deemed To Be University), Rudraram,
Telangana

Dr. Y. Md. Riyazuddin

rymd@gitam.edu

Telangana

ABSTRACT

The digitization of whole-slide imaging within digital pathology has propelled the advancement of computer assisted tissue examination utilizing machine learning methodologies, particularly convolutional neural networks (CNNs). Numerous CNN-based approaches have been proposed to effectively analyze histopathological images for tasks such as cancer detection, risk prediction, and cancer subtype classification. While many existing methods have relied on patch-based examination due to the immense size of histopathological images, such small window patches often lack sufficient information or patterns for the tasks at hand. Pathologists routinely inspect tissues at various magnification levels to scrutinize complex morphological patterns through microscopes. In response to these challenges, we propose a novel deep learning model for histopathology, named Deep-Hipo, which concurrently utilizes multi scale patches for precise histopathological image analysis. Deep-Hipo simultaneously extracts two patches of identical size from both high and low magnification levels, enabling capture of intricate morphological patterns within both large and small receptive fields of a whole-slide image. Deep-Hipo has demonstrated superior performance compared to current state-of-the-art deep learning methods. The efficacy of Deep-Hipo was evaluated across various types of whole-slide images of the stomach, well differentiated, moderately differentiated, and poorly differentiated adenocarcinoma, the poorly cohesive carcinoma with signet ring cell features, and normal gastric mucosa. Additionally, the optimally trained model was applied to histopathological images from cancer Genome Atlas (TCGA), including Stomach Adenocarcinoma (TCGA-STAD) and The Colon Adenocarcinoma (TCGA-COAD), which exhibit the similar pathological patterns to the gastric carcinoma.

Keywords - Deep-Hipo, Whole slide images (WSI), Convolutional neural network (CNN), N-Net, VGG16, Dense-Net, Efficient-Net, Multi-Resolution Deep learning network (MRD-Net), Histopathology, Google Brain's Inception V3 (GB-INCV3), CAT-Net.

1. INTRODUCTION

Histopathology stands as the primary method in cancer diagnosis and clinical decision-making. Nevertheless, traditional manual examinations and biopsy analyses are both time-consuming and financially burdensome. Cancer

diagnosis requires extensive scans across broad fields of varying morphology, encompassing benign regions, within a whole slide image [3]. Additionally, the escalating number of biopsies and cancer cases amplifies the workload of pathologists, leading to decreased efficiency and effectiveness in examining biopsies or surgical specimens.

The digitization of Whole Slide Images (WSI) of stained tissue specimens has facilitated cost-effective computational analysis of histopathological images using machine learning techniques. Consequently, an increasing number of studies have demonstrated the potential of machine learning techniques in various clinical applications over the past decade, including texture-based convolutional neural networks (CNN) for cancer detection, image segmentation for histopathological images, patient risk prediction, integrative deep learning for survival analysis, and cancer subtype classification [1, 5].

The evolution of machine learning techniques for digital pathology image analysis can be characterized by distinct periods before and after the emergence of deep learning, particularly convolutional neural networks (CNNs); it automatically learns complex spatial features [23]. Pre-deep learning methods predominantly relied on the handcrafted features predefined in advance, such as nuclei size and gland shape. For example, morphological features were extracted using an information maximum gain algorithm, and these handcrafted features were then utilized in Support Vector Machines (SVM) for cancer classification. Other techniques involved training regularized logistic regression models with features like size, shape, and intensity of nuclei to classify benign and malignant proliferative breast lesions. Additionally, feed-forward networks were fed handcrafted features of nuclei shape, distribution, and number to predict low and high-risk patients. Grey level co-occurrence matrices and run length matrices were computed in segmented breast biopsy WSIs, and SVMs were employed as classifiers [29]. A graph-based method was developed, incorporating cell nuclei morphology and neighborhood information for cancer classification. Although these non-deep learning techniques with handcrafted features offer readily interpretable results, they often struggle to define the complex morphological patterns of cancer, which are not always intuitive.

In recent years, deep learning-based methods, including CNNs, have become increasingly prevalent for histopathological image analysis. These methods, such as GoogLeNet, have demonstrated significantly improved performance by learning spatial features in a hierarchical structure without the need for feature extraction preprocessing. For instance, GoogLeNet identified metastatic breast cancer by examining patches randomly extracted from regions of interest, and heatmaps were generated to visualize tumor probabilities. Other approaches, like a spatially constrained CNN, have been developed to detect various nuclei shapes in cancer WSIs, with cancer regions identified based on nuclei classification. Additionally, fast CNN models and rotation equivariant CNNs have been proposed for efficient cancer region detection and pathology image analysis. These advancements also include the application of CNNs for image segmentation and the development of deep CNN models inspired by InceptionV3 to detect cancer metastases, where patches of tumor and normal tissue were integrated through InceptionV3 and SVM for localizing cancer regions. In the analysis of histopathological images, the ability to identify various sizes of complex morphological patterns is crucial. Multi-scale spatial features can be effectively learned through multi-scale receptive field deep learning techniques applied to intricate image data [17]. These methods often utilize asymmetric convolutional blocks to decompose symmetric kernels (such as 3×3 filters) into asymmetric convolutions (e.g., 1×3 filters). For example, InceptionV3 incorporates multi-scale kernels within an asymmetric CNN network by adjusting kernel sizes across convolutional and pooling layers for object classification. Conversely, Yolov3 and deeplabv3 are tailored for segmenting and detecting multiple objects in image data using multi-scale CNN networks, with Yolo v3 employing residual layers and deeplabv3 featuring pyramid pooling layers for multi-scale receptive fields [8,14].

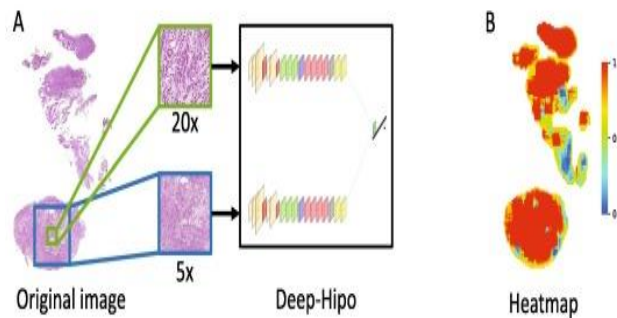


Figure 1: Overview of Deep-Hipo. (A) Deep-Hipo interprets patches of both large and small receptors. (B) Deep-Hipo produces a heatmap that visualizes cancer classification scores in a WSI.

While most deep learning models in digital pathology assume independence among patches within a histopathological image, pathologists typically interpret images by considering not only individual patches but also their neighboring regions. Some studies have begun to address these relationships between patches, with approaches such as ranking patches within tissues using a pre-trained CNN model and analyzing the patch relationships with a Recurrent Neural Network (RNN). Another method involves an expectation-maximization based CNN model that applies smoothing to patch predictive scores to reflect interdependencies. However, these approaches often struggle to capture large morphological patterns beyond the patch size.

Recent efforts in histopathological image analysis have explored the use of patches at multiple magnification levels. For instance, an adaptive weighting-based multi-field-of-view CNN integrates patches at different magnifications into separate CNNs, which are then combined to facilitate semantic segmentation in lung adenocarcinoma. Additionally, a CNN-based segmentation framework employs patches of varying sizes within the same magnification level to extract pixel and patch-based features concurrently for image segmentation. Furthermore, a multi-resolution deep learning framework (MRD-Net) combines patches of the same size but different magnifications into multi-channel images, which are trained using Resnet to identify lung adenocarcinoma growth patterns [29].

Our model makes several contributions, including learning multi-scale morphological patterns across different magnification levels simultaneously, interpreting these findings in line with histopathological domain knowledge, and achieving statistically significant performance improvements compared to current state-of-the-art methods. We present our proposed multi-scale receptive field deep learning method, describe our experimental setup, and present our results, comparing them with the benchmark method. Finally, we discuss interpretation of the morphological patterns in the connection with histopathological domain knowledge.

2. LITERATURE SURVEY

The combination of digital pathology with advanced machine learning techniques, such as deep learning, has been transforming the approach to analyzing whole-slide histopathological images (WSIs). Key applications in digital pathology utilizing machine learning encompass automatic cancer classification, survival analysis, and the subtyping based on pathological images. While most analyses of the pathological images involve processing them in patches due to the extensive size of histopathology images, there exist several applications that predict a single clinical outcome or conduct pathological diagnosis per slide [16]. However, current methods for slide-based analysis are often tailored to specific tasks, and there is a lack of comprehensive framework for slide-based analysis in WSI. We introduce a novel framework for the slide-based histopathology analysis that generates a WSI representation map, termed HipoMap, which can be applied to various slide-based challenges in conjunction with convolutional neural networks. HipoMap transforms WSIs of different shapes and sizes into a structured image-type representation

The rise of artificial intelligence (AI) and its implications in our daily lives has sparked significant interest in the field and its subcategories, particularly machine learning and deep learning. Researchers and developers have devised algorithms based on machine learning and deep learning to address various tasks associated with tumor pathologies, including tumor detection, classification, staging, diagnostic prediction, identification of pathological characteristics, understanding of disease progression, and recognition of genomic mutations. However, the integration of AI in healthcare faces several challenges, including the need to validate algorithms and computational technologies, as well as the necessity of training pathologists and physicians to utilize these systems effectively and their willingness to embrace the results [19]. This review examines how machine learning and deep learning methods could be integrated into routine tasks of healthcare providers, highlighting both the obstacles and opportunities associated with the application of artificial intelligence in tumor morphology analysis.

Historically, the diagnosis of pathology involved human pathologists examining stained tumor specimens on glass slides through a microscope to detect cancer. However, with the rapid advancement of deep learning in recent years, an increasing number of tissue slides are being digitized by scanners and stored as whole slide images (WSIs) [1, 6, 14]. Given the extensive digitization of WSIs, numerous endeavors have emerged to investigate the capabilities of deep learning in analyzing histopathological images. Histological images and associated tasks possess distinct features, often necessitating specialized processing techniques.

Neural networks serve as the core framework for most Deep Neural Network (DNN) algorithms, comprising interconnected units known as neurons organized into layers, including input, hidden, and output layers. DNNs

typically incorporate multiple hidden layers. The output of a neuron, referred to as its activation, results from a linear combination of its inputs and parameters (weights and bias), which is then transformed by an activation function. Commonly used activation functions in neural networks include sigmoid, hyperbolic tangent, and Rectified Linear Unit (ReLU) functions. At the final output layer, activations are mapped to a distribution over classes using the softmax function. Among the most prevalent and widely utilized supervised deep learning networks are Convolutional Neural Networks (CNNs), particularly renowned for their effectiveness in processing visual data such as images and video sequences. CNNs are composed of three primary types of layers: convolutional layers, pooling layers, and fully connected layers.

The convolutional layer stands as the cornerstone of the CNN architecture, consisting of multiple filters, also known as kernels, represented as a grid of discrete values. These values, termed kernel weights, are adjusted during the training process. The convolution operation involves the kernel sliding over the entire image both horizontally and vertically, with the dot product computed between the image and kernel by multiplying corresponding values and summing them to generate a scalar value at each position [28]. Specifically, each kernel convolves over the input matrix to produce a feature map. Following this, the feature maps obtained from the convolutional operation undergo sub-sampling in the pooling layer. Together, the convolution and pooling layers form a pipeline known as feature extraction. Finally, the fully connected layers integrate the features extracted by the preceding layers to execute the final classification task.

Various methods are utilized for cancer detection, including Magnetic Resonance Imaging (MRI) Scans, Computed Tomography (CT) Scans, Ultrasound, and Nuclear Imaging. However, none of these methods provide a completely accurate prediction of cancer. Tissue-based diagnosis typically involves staining elements of tissues, usually with hematoxylin and eosin (H&E), to highlight cell structures, types, and other foreign elements for examination under a microscope or via high-resolution images [11, 32]. Histopathology tests are essential for tumor detection, but they have limitations such as intra-observer variation and difficulty in identifying cancer cells and tissues due to their diverse appearances. Deep learning, a subset of machine learning inspired by the human brain's functioning, is adept at analyzing unstructured patterns. Deep learning models achieve high success rates by learning hierarchical representations and extracting diverse features without the need for prior domain knowledge. Unlike traditional methods that require extensive feature engineering, deep learning algorithms only necessitate properly formatted data and suitable network parameters.

Pre-designed networks like AlexNet, MobileNet, and Inception are commonly employed, while scholars have also proposed manual networks for breast cancer classification, such as Artificial Neural Networks based on Maximum Likelihood Estimation (MLE) and Radial Basis Function (RBF) Neural Networks. Several methodologies have been developed to improve classification accuracy with reduced computational complexity. Techniques like the AR β NN method, combining Neural Networks (NN) and multivariate adaptive regression splines (MARS), and systems integrating fuzzy-artificial immune systems with the K-NN algorithm have been proposed.

Scholars have also introduced descriptors such as CLBP, GLCM, LBP, LPQ, ORB, and PFTAS, achieving accuracies of up to 85.1% [15, 23]. While some researchers have utilized the BreakHis dataset, others have reported accuracy ranging from 80 to 85% using different parameters and the network setups. The image preprocessing, segmentation, feature extraction, and classification constitute the tasks implemented in deep learning algorithms for histopathological image analysis. These tasks involve converting data into a suitable format, separating regions of interest from the background, extracting visual content features, and classifying images using techniques like SVM or fully connected layers with activation functions such as Softmax.

3. PROPOSED METHODOLOGY

Deep-Hipo (deep learning for HIstoPathology) demonstrates a capacity for precise analysis of histopathological images through its ability to grasp multi-scale morphological patterns across diverse magnification levels of patches within a Whole Slide Image (WSI) concurrently. Employing a pixel-wise approach, Deep-Hipo calculates the likelihood of cancer occurrence for each pixel by traversing multi-scale windows across the histopathological image [1, 7]. Subsequently, it renders probability scores to identify regions potentially affected by cancer. The collection of probability scores by Deep-Hipo on a given histopathological image holds utility in tasks such as cancer classification, detection, and segmentation.

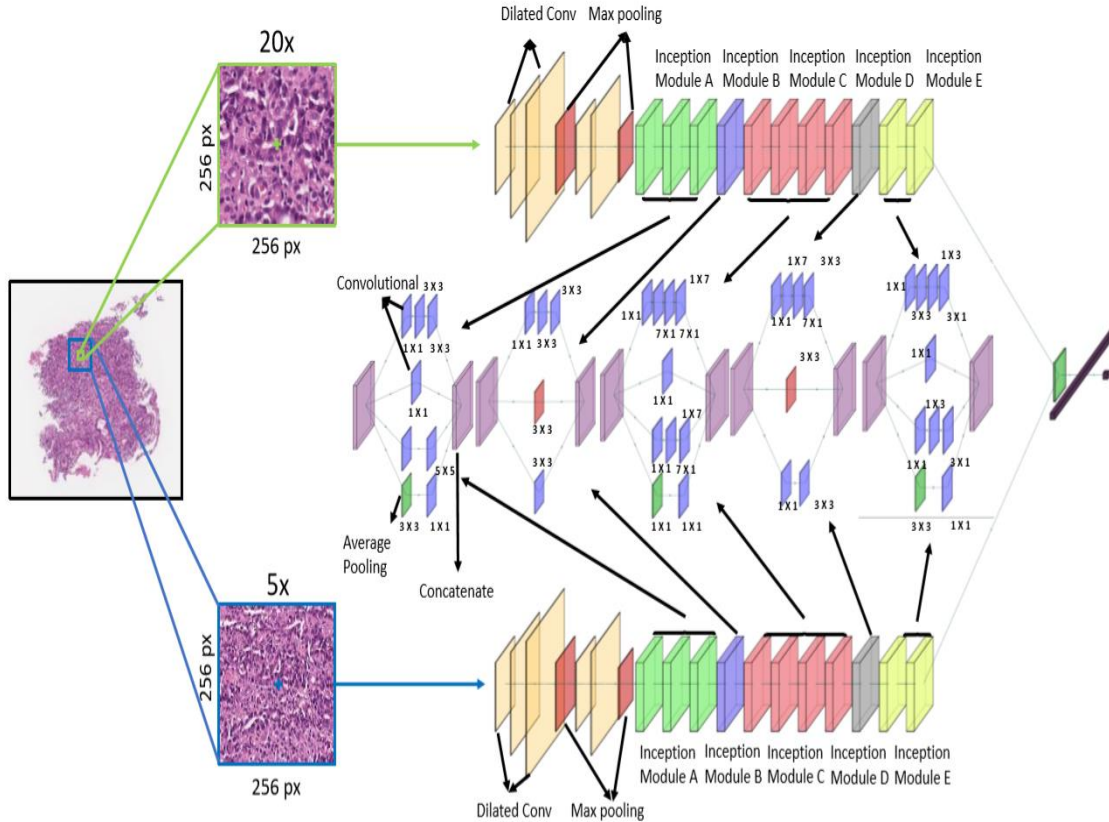


Figure 2: Architecture of Deep-Hipo. Two patches of the same pixel size are extracted in high and low magnification levels and the patches are introduced to the tracks of Deep-Hipo. Morphological patterns in the large and small receptive fields are learned simultaneously.

We introduce a deep learning model tailored for histopathology, capable of simultaneously processing multi-scale patches to enhance the precision of histopathological image analysis and identify cancerous regions. The key advancements offered by our model include:

1. The ability to assimilate multi-scale morphological patterns across both high and low magnification levels within a single model.
2. Interpretation of morphological patterns findings by establishing connections with domain-specific knowledge in histopathology.
3. Delivering the superior statistically significant performance compared to existing state-of-the-art methods in histopathological image analysis.

3.1 DATASET USED

The experiments were conducted using dataset of histopathological images derived from gastroscopic biopsy samples collected from 94 cases at the Gyeongsang National University Changwon Hospital (Changwon, Korea) to evaluate the effectiveness of our proposed method [5]. Hematoxylin and Eosin (H&E) staining was applied to the tissue specimens following standard protocols in routine clinical practice. Subsequently, all 94 biopsy samples were used to generate whole slide images (WSIs). Cancerous regions within the WSIs were manually annotated by a pathologist. The WSIs were categorized into four groups: well-differentiated (WD), moderately-differentiated (MD, 11 WSIs), poorly-differentiated adenocarcinoma (PD, 20 WSIs), and poorly cohesive carcinoma, which includes signet-ring cell features (PC, 20 WSIs), alongside samples of normal gastric mucosa (30 WSIs).

Adenocarcinoma represents the most prevalent histologic type of gastric carcinoma, typically characterized by irregularly distended, fused, or branching atypical glands of varying sizes. Traditionally, gastric adenocarcinoma is graded as well, moderately, and poorly differentiated based on gland formation. WD adenocarcinoma is typified by well-formed glands, while MD adenocarcinoma represents an intermediate stage between well and poorly differentiated cancer. PD adenocarcinoma is characterized by highly irregular glands, posing identification challenges.

PC carcinoma is a distinct subtype of gastric adenocarcinoma that encompasses signet-ring cell features and other cellular variations. Signet-ring cell features are observed in cancers predominantly or exclusively composed of the signet-ring cells, exhibiting a central, optically clear globoid droplet of cytoplasmic mucin with an eccentrically placed nucleus. Other cellular variants may include cancers composed of cells resembling histiocytes, lymphocytes, or plasma cells.

3.2 DATA PREPROCESSING

Preprocessing of the 94 WSIs involved eliminating background elements and noise, while also extracting regions of interest (ROIs). Gaussian blur smoothing was applied to address noise caused by tissue tears and the folds, as well as overstaining, whereas background removal was achieved through thresholding based on RGB values. Subsequently, annotated ROIs were identified within the WSIs. These preprocessing steps were carried out using the PyHistopathology3 Python package [8].

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




Figure 3: The architecture of each track in Deep-Hipo follows CAT-Net

From the 94 WSIs, a total of 179 and 164 tissue samples were obtained from 64 cancer WSIs and 30 control WSIs, respectively. The dataset was randomly partitioned into training (80%) and test (20%) sets, ensuring the preservation of tissue number ratios across the four categories. The training data was further divided into training (80%) and validation (20%) subsets. Each experiment utilized 219, 55, and 69 tissue samples for training, validation, and testing, respectively.

After the dataset was split, non-overlapping patches measuring 256×256 pixels were extracted at both 20x and 5x magnification levels, with each patch sharing a common center point. This process yielded approximately 80,000 cancer patches.

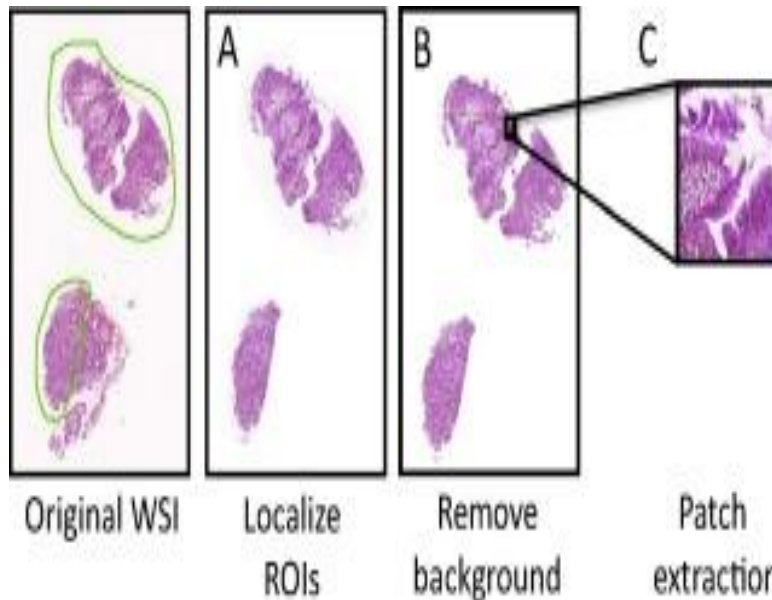


Figure 4: Data preprocessing: (A) cropping Regions Of Interest (ROIs); (B) cleaning a whole slide image (removing tissue tear/fold and overstaining and background); (C) extracting patches.

3.3 EXPERIMENTS

A comparative analysis was conducted between Deep-Hipo and several other benchmark methods [12], repeating the experiment twenty times to ensure the reproducibility of performance. Each benchmark method was implemented with specific experimental settings:

N-Net: Trained with patches of 256×256 pixels using Keras in Python with an Adaptive Moment Estimation (ADAM) optimizer.

VGG16: Trained with patches of 256×256 pixels using Keras in Python with a Stochastic Gradient Descent (SGD) optimizer.

Dense-Net: Modified DenseNet121 with one sigmoid activation node at the output layer, trained with patches of 256×256 pixels using Keras in Python with ADAM.

Efficient-Net: Replaced the output layer of EfficientNetB3 architecture with one sigmoid activation node, trained with patches of 256×256 pixels using Keras in Python with ADAM.

MRD-Net: Considered multiscale patches at both 20x and 10x magnification levels and combined them to form a new input of dimensions $224 \times 224 \times 6$, trained using ResNet with ADAM optimizer.

CAT-Net: Implemented in Keras with an SGD optimizer.

GB-INC3: Utilized a non-overlapping kernel window of 299×299 pixels, implemented in PyTorch with ADAM optimizer.

Deep-Hipo: Considered two tracks with patches of 5x and 20x magnification levels, sharing the same center points, implemented in Keras with ADAM optimizer.

4. EXPERIMENTAL RESULTS

The output layers of all benchmark models were configured with one node using sigmoid activation to compute the probability of cancer. Optimal hyperparameters were obtained through grid search with training and validation data, ensuring minimum validation loss for each experiment [13]. For CAT-Net and Deep-Hipo, batch normalization was used for faster training, and dropout was applied to prevent overfitting. In the case of Deep-Hipo, the most common optimal hyperparameters from the twenty experiments were a learning rate of $1e-5$, dropout of 0.3, and weight decay of 0.02.

The benchmark methods were evaluated across all categories of cancer tissues. Notably, Deep-Hipo exhibited the highest accuracy. Overall, Deep-Hipo achieved an accuracy of 0.937 ± 0.0032 , surpassing the second best model, CAT-Net, by 2.6.

Table 1: Patch-wise performance 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	0.951 (0.021)	0.949 (0.019)	0.947 (0.018)	0.953 (0.023)
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	0.949 (0.024)	0.945 (0.026)	0.939 (0.024)	0.950 (0.016)
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	0.918 (0.029)	0.915 (0.031)	0.908 (0.021)	0.921 (0.026)
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	0.931 (0.019)	0.929 (0.023)	0.922 (0.019)	0.934 (0.022)
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	0.937 (0.032)	0.934 (0.027)	0.929 (0.032)	0.939 (0.028)

Table 2: 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 ×	0.937 (0.032)	0.934 (0.027)	0.929 (0.032)	0.939 (0.028)
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)

Table 3: Performance comparison between dilated and conventional convolutional layers in Deep-Hipo

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

5. DISCUSSION

We assessed the optimal magnification levels for the large and small receptive fields in Deep-Hipo by conducting experiments and compared the accuracy, precision, recall, and F1-scores of Deep-Hipo across various combinations of magnification levels for the two tracks. The best performance, with values of 0.937 ± 0.0032 for accuracy, $0.934 \pm$

0.0027 for F1-score, 0.929 ± 0.0032 for precision, and 0.939 ± 0.0028 for recall, was obtained when using 20× and 5× magnification levels (refer to Table 2).

In Deep-Hipo, dilated convolutional layers were employed instead of conventional convolutional layers (refer to Table 3). Deep-Hipo with dilated convolutional layers exhibited an accuracy 4% higher than that with conventional convolutional layers.

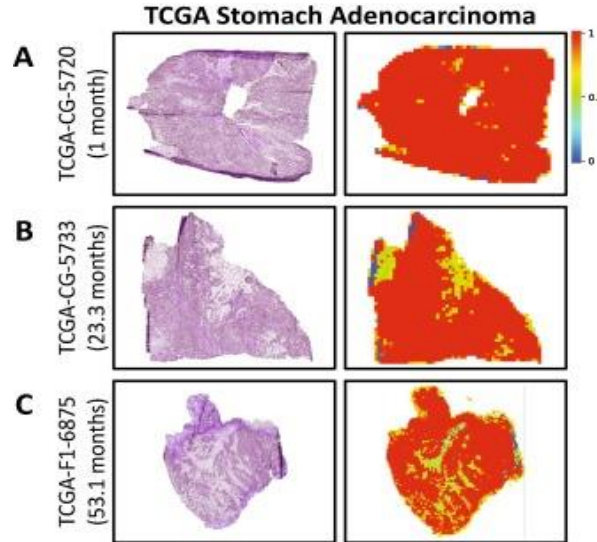


Figure 5: Heatmaps of TCGA Stomach Adeno- carcinoma (STAD). (A) Adenocarcinoma, poorly differentiated, stroma in green. (B) Mucinous adenocarcinoma, a few atypical cells in the mucin pools in green. (C) Adenocarcinoma, poorly differentiated, stroma in green.

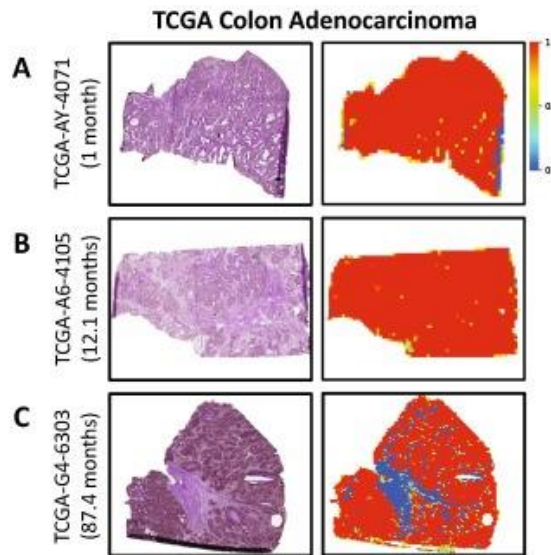


Figure 6: TCGA Colon Adenocarcinoma (COAD) heatmap. (A) Adenocarcinoma, well-differentiated, tissue folds in blue, (B) Adenocarcinoma, moderately differentiated, (C) Adenocarcinoma, moderately differentiated, stroma (no cancer cells) in blue and tissue folds in green.

We evaluated our proposed model using histo pathological images from TCGA Stomach and Colon Adenocarcinoma as an external cross validation (refer to Figure 4). In TCGA-STAD, we examined cancer tissues obtained from three patients with varying survival durations: short-term (TCGA-CG5720, 1 month), medium term (TCGA CG-5733, 23.3 months), long term (TCGA F1 6875, 53.1 months). In TCGA-CG-5720 and TCGA-F1-6875, stromal regions (without cancer cells) were identified by Deep-Hipo and represented in green, whereas in TCGA-CG-5733, a few atypical cells within the mucin pools were highlighted in green. Similarly, we examined TCGA-COAD tissues from three

patients(refer to Figure 6). In TCGA-G4-6303, stromal regions devoid of cancer cells were accurately identified and visualized in blue, while most cancerous regions were correctly depicted in red.

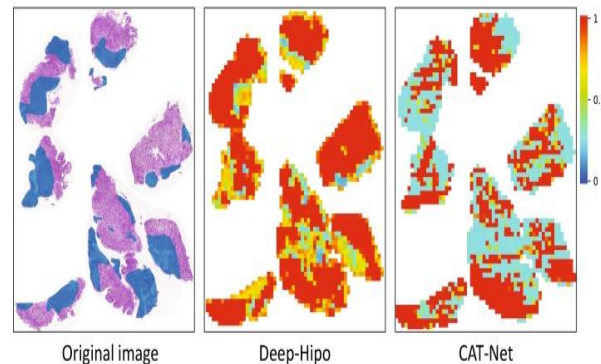


Figure 7: Performance comparison between Deep-Hipo and CAT-Net.

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. Deep-Hipo integrates the dual pathways of CAT-Net to encompass both large and small receptive fields within histopathological images. It's worth noting that while CAT-Net examines local patterns of individual patches independently, Deep-Hipo captures global patterns spanning across patches at 20× magnification, while also considering smaller receptive fields. Figure 7 illustrates the heat maps generated by Deep-Hipo and CAT-Net, alongside an original slide depicting a poorly cohesive carcinoma. In the original histopathological image (leftmost), areas annotated in blue predominantly represent normal gastric mucosa, devoid of cancer cell infiltration, while other areas indicate cancerous regions. Deep-Hipo exhibited a few regions erroneously classified as false positives (identified as cancerous but actually normal), but the majority of regions were accurately identified as true positives. Conversely, CAT-Net yielded significantly larger areas of false negatives (identified as normal but actually cancerous) and false positives compared to Deep-Hipo. Additionally, Deep-Hipo not only achieved higher accuracy but also produced smoother probability scores than CAT-Net, as it considers connected neighboring regions overlapped for analysis by a large receptive field.

6. CONCLUSION

In this study, we introduced Deep-Hipo, a novel deep learning architecture tailored for histopathology, which leverages multi-scale receptive fields extracted from diverse patches across multiple magnification levels in histopathological images. This approach mirrors the methodology employed by pathologists who examine tissues at varying magnification levels to scrutinize intricate morphological patterns under a microscope. The experimental findings from cross-validation underscored the robust performance of Deep-Hipo across metrics like accuracy, F1-score, precision, and recall. Moreover, the model's efficacy was validated through external assessment using WSIs of TCGA stomach adenocarcinoma and TCGA colon adenocarcinoma.

The heat maps generated in this study were produced by shifting a patch window by 16 pixels. Employing a smaller stride could potentially enhance the quality of the heatmap by generating more patches, albeit at the cost of increased computational complexity. To address this, parallel computation strategies could be explored to enable real-time analyses. While our focus in this study was on using Deep-Hipo for histopathological image analysis to predict cancer probabilities for individual patches, the multi-scale receptive field deep learning model holds promise for broader applications in histopathological analyses, including survival analysis, patient risk prediction, and cancer subtype classification, which could be achieved by reconfiguring the output layer accordingly.

7. REFERENCES

- [1] M.Z. Alom, T. Aspiras, T.M. Taha, V.K. Asari, T. Bowen, D. Billiter, S. Arkel, 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. Zwiggelaar (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 network for whole slide tissue image classification, 2016 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] Y. Liu, et al., 2017. Detecting cancer metastases on gigapixel pathology images. *Corr.* URL: <http://arxiv.org/abs/1703.02442>.
- [20] 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).
- [21] K. Nguyen, A. Sarkar, A.K. Jain, Prostate cancer grading: Use of graph cut and spatial arrangement of nuclei, *IEEE Trans. Med. Imag.* (2014).
- [22] K. Oikawa, et al., Pathological diagnosis of gastric cancers with a novel computerized analysis system, *J. Pathol. Inf.* (2017).
- [23] 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.
- [24] E. Radiya-Dixit, D. Zhu, A.H. Beck, Automated classification of benign and malignant proliferative breast lesions, *Scientific Rep.* (2017).
- [25] J. Redmon, A. Farhadi, Yolov3. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2017.
- [26] F. Sheikhzadeh, et al., Automatic labeling of molecular biomarkers of immunohistochemistry images using fully convolutional networks, *plos One* (2018).
- [27] 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>.
- [28] M. Tan, Q.V. Le, Efficient net: rethinking model scaling for convolutional neural networks, 2019.
- [29] D. Wang, A. Khosla, R. Gargeya, H. Irshad, A.H. Beck, Deep learning for identifying metastatic breast cancer, 2016. *Arxiv preprint arxiv:1606.05718*.

- [30] 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.
- [31] 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.
- [32] S. Wang, D.M. Yang, R. Rong, X. Zhan, G. Xiao, Pathology Image Analysis Using Segmentation Deep Learning Algorithms, 2019.