Importance Of Stain Normalization In Enhancing Semantic Segmentation Of Oral Squamous Cell Carcinoma Histopathological Images

Rupesh Mandal[1], Prithvijit Das[2], Hrishiraj Modi[3], Nupur Choudhury[4]*, Anuran Patgiri[5], Uzzal Sharma[6], Mrinmoy Mayur Choudhury[7], Muktanjalee Deka[8], Jyoti Barman[9]

[1]School of Technology, Assam Don Bosco University, Guwahati, Assam, India; rupesh.mandal@dbuniversity.ac.in
[2]School of Technology, Assam Don Bosco University, Guwahati, Assam, India; prithvijitdas488@gmail.com
[3]School of Technology, Assam Don Bosco University, Guwahati, Assam, India; hrishiraj.modi.12b.399@gmail.com
[4*]School of Technology, Assam Don Bosco University, Guwahati, Assam, India; nupur.choudhury@dbuniversity.ac.in
[5]School of Technology, Assam Don Bosco University, Guwahati, Assam, India; anuranpatgiriworks@gmail.com
[6]Department of Computer Science, Birangana Sati Sadhani Rajyik Vishwavidyalaya, Golaghat, Assam, India; druzzalsharma@gmail.com
[7]State Cancer Institute, Guwahati Medical College, Guwahati, Assam, India; mrinmoygmc@gmail.com
[8]School of Technology, Assam Don Bosco University, Guwahati, Assam, India; muktanjalee@rediffmail.com
[9]School of Technology, Assam Don Bosco University, Guwahati, Assam, India; jyoti.barman@dbuniversity.ac.in

***Correspondence:** Nupur Choudhury
*Email: nupur.choudhury@dbuniversity.ac.in

Abstract: In the field of histological image analysis for Oral Squamous Cell Carcinoma (OSCC), few studies have explored the integration of stain normalization methods as a preprocessing step in deep learning workflows. Standard Hematoxylin and Eosin (H&E) staining techniques are commonly used for tissue visualization, where Hematoxylin highlights cell nuclei in blue and Eosin colors the cytoplasm, muscle fibers, and other components in pink. Our research investigates the impact of applying Reinhard's color transfer technique for stain normalization on the performance of deep learning models for semantic segmentation. We propose that stain normalization not only enhances model robustness but also serves as an effective data augmentation strategy, which is especially valuable given the limited availability of OSCC histological data. To accommodate hardware constraints while maintaining high performance, we employ a lightweight UNet architecture with an EfficientNet backbone. This model strikes a balance between computational efficiency and accuracy, achieving results comparable to those of the widely adopted ResNet models, which are often considered the gold standard in Convolutional Neural Network (CNN) architectures. Our findings support that stain normalization can serve as a valuable data augmentation method, enhancing the effectiveness of deep learning models in histological segmentation tasks for OSCC.

Keywords: OSCC, stain normalization, semantic segmentation, data augmentation, UNET.

1. Introduction

Oral Squamous Cell Carcinoma (OSCC) constitutes approximately 90% of all oral malignancies, serves as the leading cause of human mortality among all oral cancers [28,29]. Originating from the stratified squamous epithelium which is a component of the oral mucosa, this tumour has been accounted for 377,713 OSCC cases globally in 2020 according to the Global Cancer Observatory (GCO) [30,31]. Males are mostly affected by Oral Squamous Cell Carcinoma OSCC in comparison with females, where men in their mid-age are the most vulnerable [32]. Pathogenic bacteria, particularly Porphyromonas gingivalis, Fusobacterium nucleatum and Prevotella intermedia, can be closely linked to Oral Squamous Cell Carcinoma (OSCC) [33]. The hematoxylin and eosin (H and E) stain is a universally adopted and robust stain which serves as a primary contrast method in diagnosis of biopsy specimen [34]. The H and E stain is essential for pathologists in daily diagnostic work in providing detailed views of tissue structures, including cell nuclei, cytoplasm, and extracellular components. For over 150 years, this staining procedure remains unchanged. [9].

A robust segmentation model necessitates a potent backbone network serving as an encoder, proficient in capturing multi-scale information, adept at preserving spatial details, and characterized by low computational complexity[13] . The domain of oral cancer segmentation via deep learning methodologies is notably constrained, with current efforts predominantly relying on the ORCA[1] and OCDC[21] datasets, which are publicly accessible online.

In recent advancements within medical image processing, transformer-based models have emerged as dominant in semantic segmentation, exhibiting exceptional performance in extracting spatial information from images[15], [16], [17]. The self-attention mechanism, introduced by Vaswani et al.[22] , excels in capturing global interactions among input vectors. This innovative approach was directly applied in Vision Transformers (ViT) by Alexey et al[10]. on image patches to encapsulate global interactions among individual pixels. Subsequent methodologies have been proposed to compute self-attention with reduced computational demands[12], [13], [17]. Nevertheless, our research emphasizes the impact of image preprocessing techniques on model performance.

The exploration of image preprocessing methods within the context of histological images remains sparse. Despite the proposition of various preprocessing techniques such as histogram equalization algorithms[23] , intensity correction[24] , gamma correction[25] , and convolution-based methods like blurring, edge enhancement, and sharpening[23] , no substantial improvements in accuracy have been documented. Techniques like Retinex and Multiscale Retinex perform optimally in low illumination and fuzzy enviroments[18] . However, our literature review indicates a lack of re-search on stain normalization as a preprocessing strategy for histological images in OSCC (Oral Squamous Cell Carcinoma).

Stain normalization is crucial due to the color variations observed in histological images, which arise from different scanners, slide preparation techniques, staining procedures, stain concentrations, temperatures, and other factors [19]. Our proposed methodology utilizes Reinhard's color transfer algorithm for stain normalization [20]. This technique, though simple, proves highly effective when an appropriate reference image is selected. We have further enhanced this algorithm to emphasize the cell nuclei regions within the images by iteratively selecting pixel regions where the blue channel exhibits dominant intensity values. This enhancement contributes to more stable training and improved generalization capabilities of the model. Given our hardware resource constraints, we have opted for the lightweight EfficientNet[26] as the encoder within the UNet[27] architecture. EfficientNet is an optimal choice due to its balance between model complexity and performance, achieved through a principled compound scaling method that uniformly scales the network's depth, width, and resolution.

2. Related Works

Deep learning is poised to revolutionize the diagnosis and prognosis of Oral Squamous Cell Carcinoma(OSCC) by extracting intricate visual patterns from histological images. Ahmed et al. [4] proposed a novel attention block called as CSAF that could emphasize important channel and spatial areas. Their observations indicated that using attention mechanisms along with residual connections and ASPP gives significant improvement over the baseline UNet model. Jelena et al. [3] proposed a 2-stage multiclass grading system where the first stage included data augmentation, training classification model, preprocessing data with SWT functions and selected model was used as a backbone for the DeepLabv3+ semantic segmentation model. Ac-cording to the observations Xception model as backbone with Haar SWT function gave the best results. Andrea et al.[2] worked with an ensemble learning approach where multiple classification networks each taking different tiles of a split image as input were used as backbone for the Unet architecture with a single decoder. The Oral Cancer Annotated(ORCA) dataset introduced by Martino et al.[1] trained the dataset with four deep learning methods, the observations indicated that combining R channel of RGB with V channel of HSV colour space gave optimal results.

Since the proposal of Vision Transformers(ViT)[10] used for image classification, the self attention mechanism has paved its way to semantic segmentation due to its ability to capture long range interdependencies between pixels. The UNetR model for volumetric image segmentation proposed by Ali et al.[16] implemented the transformer in a U shaped encoder decoder architecture similar to UNet. The output of the transformer is reshaped and downsampled to different dimensions to feed to the de-coder which comprises of convolutional layers. An advancement of the UNetR architecture proposed by Abdelrahman et al.[17] the UNetR++ uses an efficient EPA block to compute channel wise and spatial wise self attention parallelly. To bridge the semantic gap between encoder and decoder the DCA attention block proposed by Gorkem et al.[35] takes input from all the encoders and computes channel and spatial cross attention to capture long range dependencies. Hao et al.[15] also worked with multi scale features by computing the cross attention between two parallel Axial attentions[12] to capture global information.

Histological tissues are naturally transparent, requiring staining for proper visualization. Hematoxylin and Eosin (H&E) is a commonly used staining combination, where hematoxylin stains cell nuclei blue, and eosin stains cytoplasmic components and extracellular matter pink. However, variations in staining can significantly impact the accuracy of automatic image analysis, making stain normalization essential. This area of research has seen significant advancements, with various algorithms being developed to address the challenge. Hoque et al. [19] introduced a method that decomposes images into individual stain components, followed by normalization based on illumination conditions using the Multiscale Retinex algorithm. Babak et al. [36] proposed the Whole Slide Images Color Standardizer (WSICS) algorithm, which utilizes spatial information and the HSD color model for precise stain differentiation. The WSICS algorithm also applies weighted smoothing to ensure artifact-free standardization, making it highly effective in mitigating staining variations. A study closely related to our work is the research by Francesco et al. [37], which focused on stain normalization in colorectal cancer tissues. Their findings demonstrated that using the WSICS algorithm significantly enhanced the performance of deep learning models, underscoring its importance in this domain.

3. Methodology

In this section we focus on the dataset details and the preprocessing methods used, details about the architecture of the model to be used and the loss function.

3.1. Datasets

We used two publicly available datasets for our experiment with stain normalization techniques with deep learning: OCDC(Oral Cavity Derived Cancer)[21] dataset and ORCA(Oral Cancer Annotated)[1] dataset.

The OCDC dataset consists of 1,020 histological images, meticulously labeled and validated by expert pathologists for binary classification, distinguishing between tumor and non-tumor regions. This dataset is structured with 840 images designated for training and 180 images reserved for testing, specifically addressing the binary class segmentation problem of identifying tumor and non-tumor areas within histological samples. Each image in the dataset is sized at 640x640 pixels and has been digitized at a 20x magnification level. The original images were captured as Whole Slide Images (WSI) using the Slide Scanner Aperio AT2 at 20x magnification. These tissue specimens were sourced from the archives of the Department of Oral and Maxillofacial Pathology at the Federal University of Uberlandia. The images depict tissue sections that encompass a variety of cellular and structural elements, including blood vessels, keratin, lymphocytes, glands, muscle, and tumor cells. The diversity of these elements within the tissue sections presents a challenging and realistic scenario for the binary classification of tumor and non-tumor regions, making the OCDC dataset a valuable resource for advancing research in histopathological image analysis.

Our second dataset, ORCA, comprises 100 samples used for both training and testing, derived from The Cancer Genome Atlas (TCGA) dataset. The Cancer Genome Atlas, provided by the National Cancer Institute (NCI), includes clinicopathological data and unannotated Whole Slide Images (WSIs) of over 20,000 primary cancer cases, spanning 33 different cancer types. Each ORCA image has a high resolution of 4500x4500 pixels and is categorized into three distinct classes: carcinoma pixels, non-carcinoma tissue pixels, and non-tissue pixels. The dataset includes a variety of color stains, reflecting the diverse staining techniques used in histopathology. These variations enhance the dataset's complexity, making it an invaluable resource for advancing the analysis of different cancer types and improving the robustness of diagnostic algorithms.

3.2. Stain Normalization

In our work, we utilized Reinhard’s color transfer algorithm [20]. The first step involves converting the image to LAB color space to separate luminance and chromaticity components. The algorithm aims to standardize each image by transforming it so that the mean and standard deviation of its color channels match those of a reference image selected by the user. This process ensures consistency across all images in the dataset. Equation 1 outlines the process where i and j are the row and column positions in the image matrix x. For an input image matrix x, we first compute the z-score for every pixel value to preserve the relative positions of the data. This involves normalizing the pixel values by subtracting the mean of the input image and dividing by its standard deviation. Next, we multiply the result by the standard deviation σ_R of the reference image, aligning the input image's standard deviation with that of the reference. Finally, we shift the mean of the input image from 0 to the mean of the reference image ¯x_R by adding ¯x_R. This ensures that the input image has the same mean and standard deviation as the reference image, effectively standardizing the color distribution.

$$x_{ij} = z_{ij} \cdot \sigma_R + \bar{x}_R \qquad (1)$$

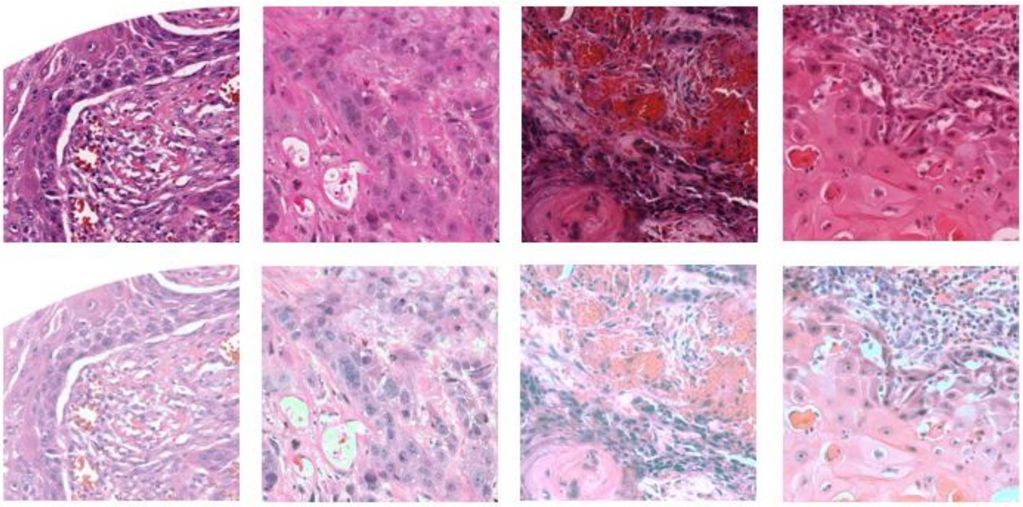


Figure 1. Samples of ORCA dataset(first row) along with its stain normalized transformation(second row).

3.3. Unet Efficientnet

We employed the UNet architecture [27] with EfficientNetB0 [26] as the backbone. EfficientNet is renowned for its high accuracy, while also being lightweight and computationally efficient. The main building block for the encoder backbone is the Mobile Inverted Bottleneck (MBConv) block. As illustrated in Figure. 2, the MBConv block begins by expanding the number of channels using 1x1 convolutions denoted by Ex(.) in Equation 2. This is followed by a Depthwise Convolution, which is then enhanced by Squeeze-and-Excitation attention denoted by D (.) and Se (.) respectively in Equation 2. Finally, the block contracts the number of channels back using another set of 1x1 convolutions denoted by Cn (.) in Equation 2, thus earning the name "inverted bottle-neck." Additionally, a residual connection is incorporated to facilitate gradient flow.

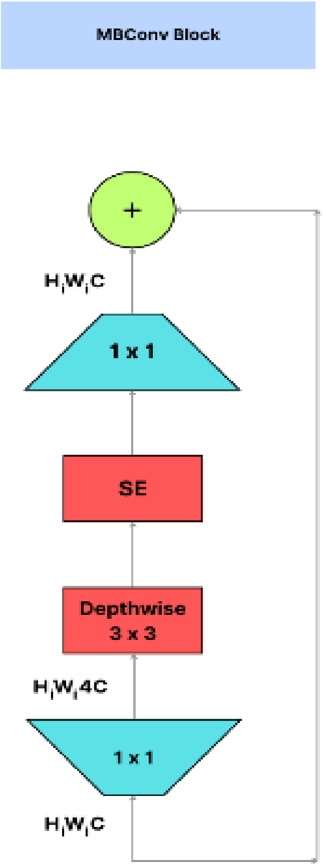


Figure 2. Overview of the MBConv Block used as basic building blocks in models belonging to EfficientNet family.

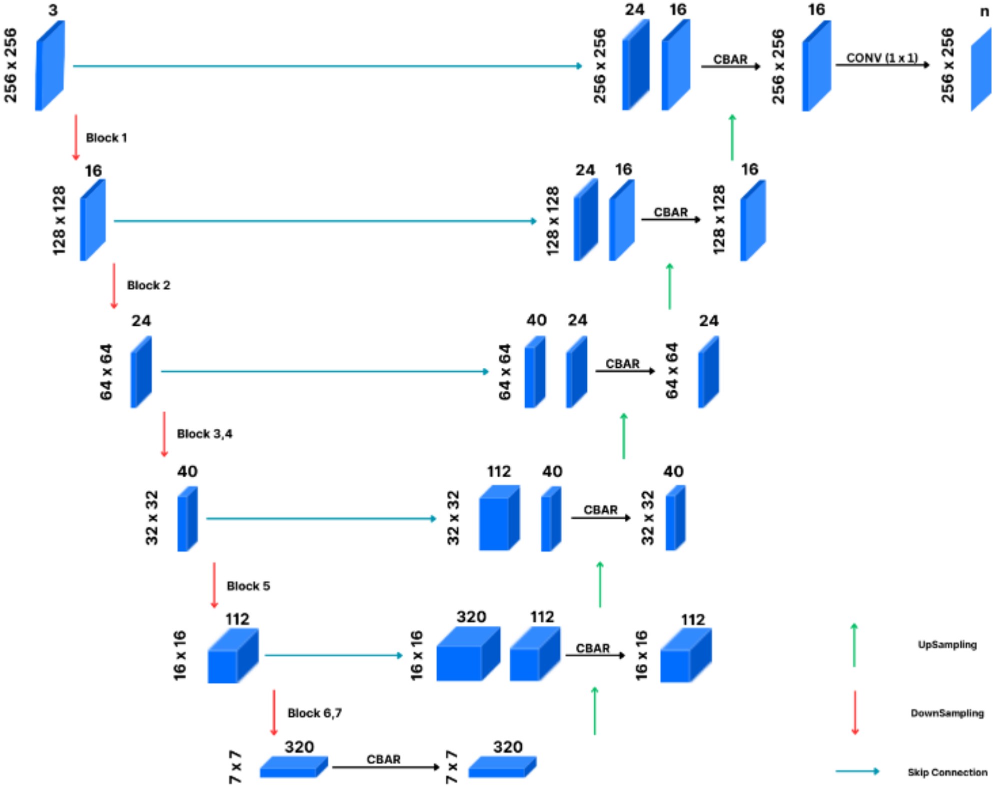


Figure 3. Architecture of the UNet-EfficientNet model with input as images with dimensions of 256x256x3, and the output is of size 256x256xn, where n represents the number of target classes.

Equations 2 and 3 define the operations for the Encoder and Decoder within the U-Net architecture, respectively. In this context, $\mathbf{X}_{\mathbf{En}}^{\mathbf{i}}$ represents the feature map output from the i-th MBConv block, while $\mathbf{X}_{\mathbf{De}}^{\mathbf{j}}$ and $\mathbf{X}_{\mathbf{En}}^{\mathbf{j}}$ denotes the j-th encoder and decoder blocks. The $CBAR\,(.)$ operation involves concatenating the feature map from the encoder's skip connection with the upsampled feature map from the preceding decoder block. This concatenated feature map undergoes a series of transformations: a 3x3 convolution, followed by batch normalization, activation, and finally, a residual connection. The symbols for ReLu, Batch normalization and 3x3 Convolution are denoted by $\delta\,(.)$, $BN\,(.)$ and $C_{3\times3}\,(.)$ respectively where X is the input for the CBAR function in Equation 4.

$$\mathbf{X}_{\mathbf{En}}^{\mathbf{i}} = \boldsymbol{Cn}\left(\boldsymbol{Se}\left(\boldsymbol{D}\left(\boldsymbol{Ex}\left(\mathbf{X}_{\mathbf{En}}^{\mathbf{i-1}}\right)\right)\right)\right) + \mathbf{X}_{\mathbf{En}}^{\mathbf{i-1}} \tag{2}$$

$$\mathbf{X}_{\mathbf{De}}^{\mathbf{j}} = CBAR\left(concat\left(\mathbf{X}_{\mathbf{En}}^{\mathbf{j}}\,,\,U\left(\mathbf{X}_{\mathbf{De}}^{\mathbf{j-1}}\right)\right)\right) \tag{3}$$

$$\mathbf{CBAR}\,(\mathbf{X}) = \boldsymbol{\delta}\,(\boldsymbol{BN}\,(\boldsymbol{C}_{3\times3}\,(\boldsymbol{X}))\, + \boldsymbol{X} \tag{4}$$

3.4. Loss Function

Our approach combines two loss functions, each targeting a critical aspect of effective semantic segmentation. The first is the Dice loss function[39], which promotes greater overlap between the predicted segmentation and the ground truth. By minimizing the Dice loss, we guide the model to converge towards an optimal alignment between prediction and reality. The Dice loss function, as defined in Equation 5, inversely correlates with the Dice coefficient—where a higher Dice coefficient indicates better overlap. Therefore, by subtracting the Dice coefficient from 1, we transform it into a loss function, where a lower Dice loss signifies more substantial overlap, and a higher Dice loss indicates less overlap.

The second loss function employed is the Focal Loss[38], specifically designed to address class imbalance issues. As illustrated in Equation 6, Focal Loss modifies the standard cross-entropy loss to focus on hard-to-classify examples, rather than being dominated by easy, well-classified instances. It introduces a modulating factor that downweights the influence of easy predictions while upweighting those that are difficult to classify.

This adjustment ensures that the model pays more attention to challenging cases, leading to a more balanced and robust segmentation performance. In equation 6 the expression for dice loss function is expressed where P_{true} is the ground truth probability for a pixel (usually 1 for foreground and 0 for background in binary segmentation).
P_{pred} is the predicted probability of the pixel belonging to the foreground (output from the model, typically between 0 and 1). ξ is a small smoothing constant added to avoid division by zero, typically a very small value. The expression for focal loss is defined in equation 8 where p_t is the transformed probability based on whether the pixel is from the foreground or background and γ is the focusing parameter that controls the weight given to hard examples. Equation 9 expresses the summation of dice loss and focal loss.

$$\mathbf{DL}\left(\mathbf{P_{true}}, \mathbf{P_{pred}}\right) = 1 - \frac{2 * P_{true} * P_{pred} + \xi}{P_{true} + P_{pred} + \xi} \tag{5}$$

$$\mathbf{p_t} = \begin{cases} \boldsymbol{y_{pred}} & \textbf{if } \boldsymbol{y_{true}} = \mathbf{1} \\ \mathbf{1} - \boldsymbol{y_{pred}} & \textbf{if } \boldsymbol{y_{true}} = \mathbf{0} \end{cases} \tag{6}$$

$$\mathbf{FL}\left(\mathbf{P_t}\right) = -\left(1 - P_t\right)^{\gamma} \log\left(P_t\right) \tag{7}$$

$$\mathbf{TL}\left(\mathbf{P_{true}}, \mathbf{P_{pred}}\right) = FL\left(P_t\right) + DL\left(P_{true}, P_{pred}\right) \tag{8}$$

4. Experiments and Results

4.1. Evaluation Metrics

In this work, we will evaluate the performance of our model on both the normal and stain-normalized datasets using several key metrics: Jaccard's coefficient(IoU), mean Intersection over Union (mIoU), Dice coefficient, accuracy, precision, and error rate. The mathematical expressions for these metrics are provided below.

Table 1. Architecture of EfficientNet B0 model excluding the Fully Connected layers.

CONV BLOCK	STAGE NUMBER
Conv, 3 x 3	
MBConv1, 3 x 3	Block 1
MBConv6, 3 x 3 MBConv6, 3 x 3	Block 2
MBConv6, 5 x 5 MBConv6, 5 x 5	Block 3
MBConv6, 3 x 3 MBConv6, 3 x 3 MBConv6, 3 x 3	Block 4
MBConv6, 5 x 5 MBConv6, 5 x 5 MBConv6, 5 x 5	Block 5
MBConv6, 5 x 5 MBConv6, 5 x 5 MBConv6, 5 x 5 MBConv6, 5 x 5	Block 6

$$\textbf{IoU} = \frac{TP}{TP + FP + FN} \tag{9}$$

$$\textbf{mean IoU} = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i + FP_i + FN_i} \tag{10}$$

$$\textbf{Dice} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \tag{11}$$

$$\textbf{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

$$\textbf{Error Rate} = \frac{FP + FN}{TP + TN + FP + FN} \tag{13}$$

The terms TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative, respectively, and are fundamental in understanding the performance of classification models.

4.2. Experimental Settings

In our experiments, we utilized the TensorFlow 2.10 deep learning framework to train our model, leveraging the computational power of an Nvidia RTX 3080 GPU. This GPU, with its 10,496 CUDA cores and 10GB of GDDR6X memory, allowed for efficient parallel processing and accelerated the training and inference phases of our work. The model was trained using the Nadam optimizer, which combines the benefits of both the Nesterov-accelerated gradient and Adam optimization methods. We set the learning rate to 0.001 and trained the models for 30 epochs on both the ORCA and OCDC datasets. A batch size of 16 was used throughout the training process to balance memory usage and convergence speed.

To enhance the robustness of our models, we implemented data augmentation techniques using Keras's ImageDataGenerator. Specifically, we applied horizontal and vertical flips to the training images, which helped increase the diversity of the training data and improve the model's generalization capabilities. Additionally, we propose that stain normalization could serve as an excellent data augmentation technique for histological images, potentially enhancing the model's ability to handle variations in staining across different samples.

4.3. Experiments on ORCA Dataset

The ORCA dataset consists of 100 validation samples and 100 test samples. Since the training samples used by Martino et al. [1] are not publicly available, we utilized samples from both the validation and test sets to create our training set. The original image resolution is quite high at 4500x4500 pixels. Given our computational resource limitations, we first resized the images to 1024x1024 pixels, followed by dividing them into patches of 256x256 pixels each. Reducing the size of the resized images increases the coverage of global regions by the patches. Initially, we tried resizing the images to 512x512 pixels and applied the same patch size of 256x256, but found that resizing to 1024x1024 pixels yielded better results. After resizing and patching the images from both sets, we shuffled them, using 2080 samples for the training set, 320 for validation, and 800 for testing. However, we believe that using various color stained image as reference image and performing stain normalization on the dataset for various stainings can significantly enlarge the dataset size and hence this technique can be a good augmentation technique for better generalization of the model to address dataset size problem. The comparative analysis between the stain-normalized and non-stain-normalized datasets is summarized in Table 2. Results conclude that stain normalization does not improve the performance of our model. The first row where 0 reference images are used are the observations of the non-stain normalized dataset. Second row shows the effect of using a single uniform color stain over the dataset. Lastly we used stain normalization as a data augmentation technique by using 7 different reference images.

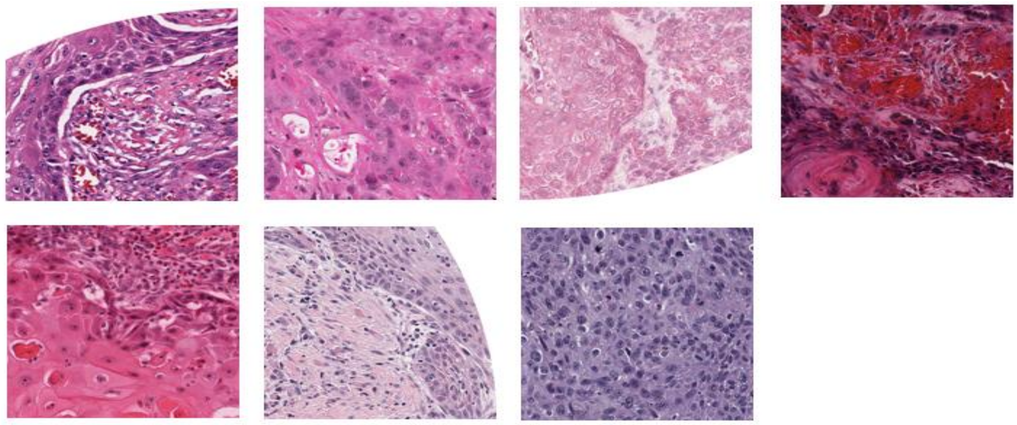


Figure 4. The seven different color stains used as reference images to stain normalize ORCA dataset. Using seven reference images increased the size of dataset by eight times.

Table 2. Observations for the ORCA dataset.

No. of reference images	Mean IoU	Dice	Accuracy	Error rate
0	82.51%	89.00%	90.54%	6.70%
1	80.89%	88.62%	89.51%	7.40%
7	80.00%	88.57%	88.57%	7.60%

4.4. Experiments on OCDC Dataset

The OCDC dataset, while relatively small, originally comprises 840 training images and 180 test images. In our study, we restructured this dataset to better suit our experimental needs by dividing it into 800 training samples, 100 validation samples, and 120 test samples. Due to resource constraints, we resized the images from their original dimensions of 640x640 pixels to 256x256 pixels, allowing us to efficiently process the data while maintaining sufficient detail for accurate analysis. We compared the performance of UNet with EfficientNet encoder on the original dataset, the stain normalized dataset and data augmented dataset consisting of stain normalized images of different reference images. According to our analysis the performance of the model does not change significantly if the dataset is stain normalized using any reference images As with the ORCA dataset, we employed the same data augmentation technique, applying seven different staining colors for stain normalization trained with UNet EfficientNet. Fig 4 shows the reference images used for stain normalization. We present the results in Table 3 for performance of model to analyse the effect of stain normalization. Similar to the results of ORCA dataset the first row where 0 reference images are used are the observations of the non stain normalized dataset. Second row shows the effect of using a single uniform color stain over the dataset. Lastly we used stain normalization as a data augmentation technique by using 6 different reference images.

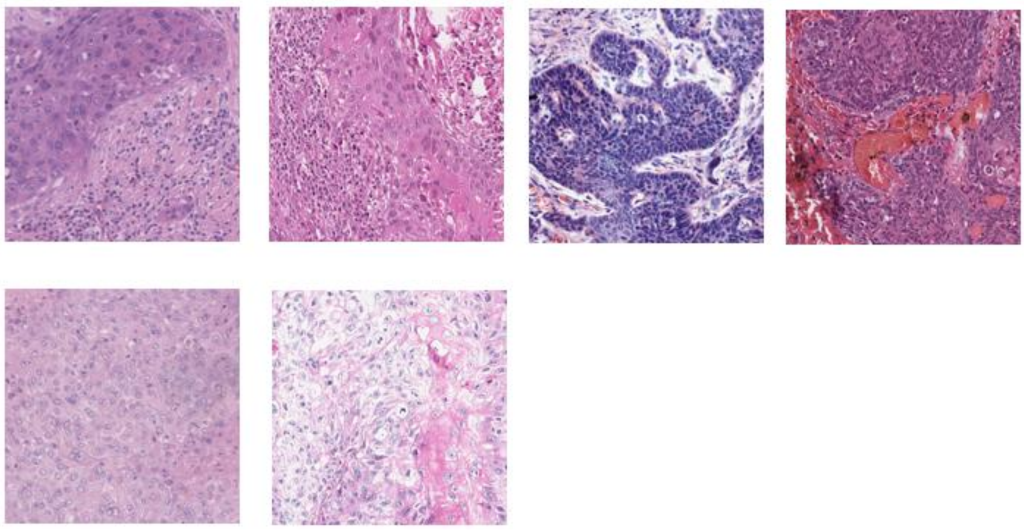


Figure 5. The seven different color stains used as reference images to stain normalize ORCA dataset. Using seven reference images increased the size of dataset by eight times.

Table 3. Observations for the OCDC dataset.

No. of reference images	Mean IoU	Dice	Accuracy	Error rate
0	87.00%	93.00%	98.06%	2.00%
1	86.25%	92.62%	97.80%	2.30%
6	88.33%	93.80%	98.08%	1.95%

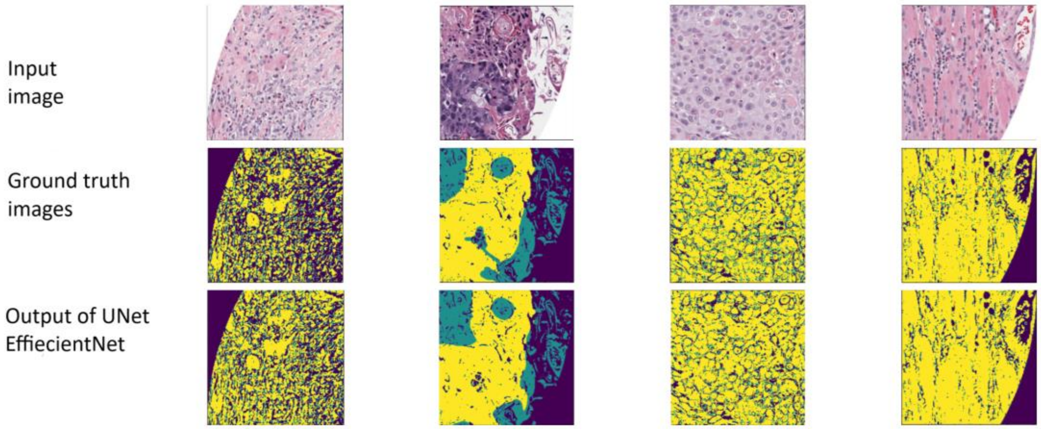


Figure 6. Output of UNet EfficientNet on ORCA test samples.

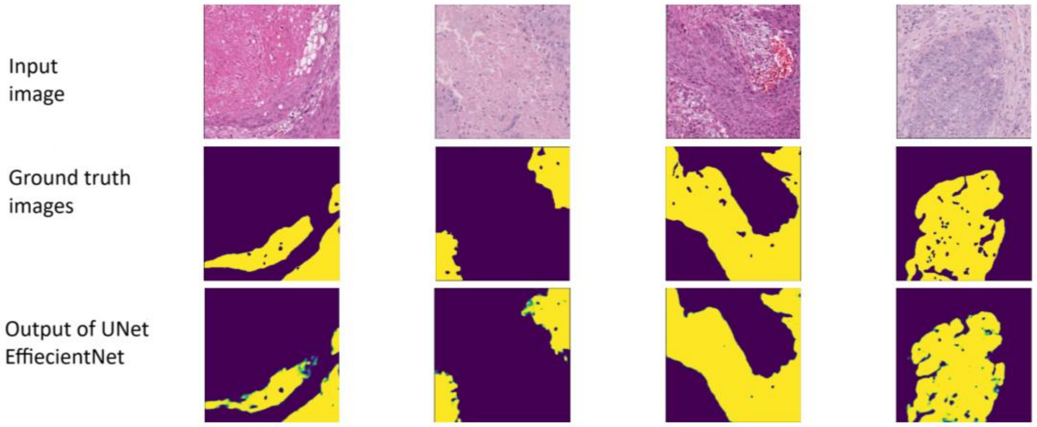
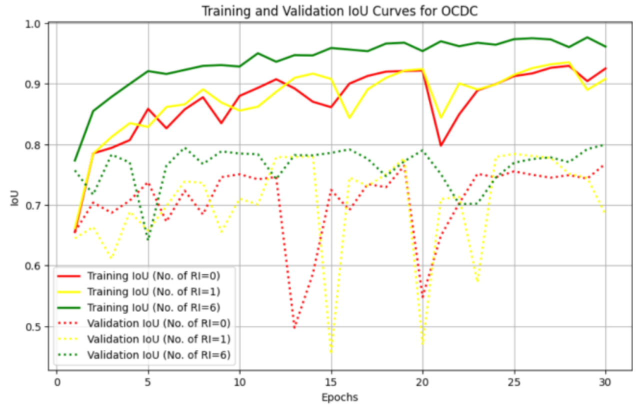
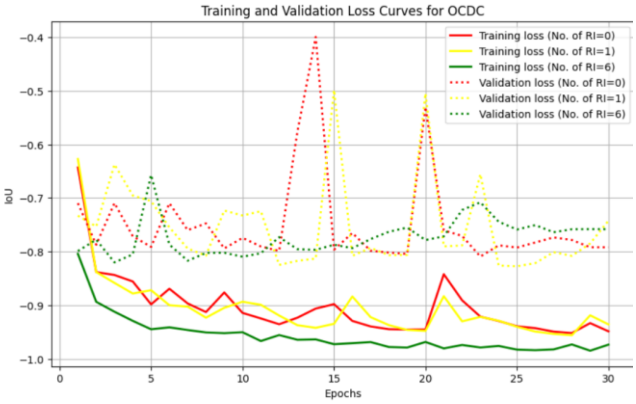


Figure 7. Output of UNet EfficientNet on OCDC test samples.

The text continues here (Figure 2 and Table 2).

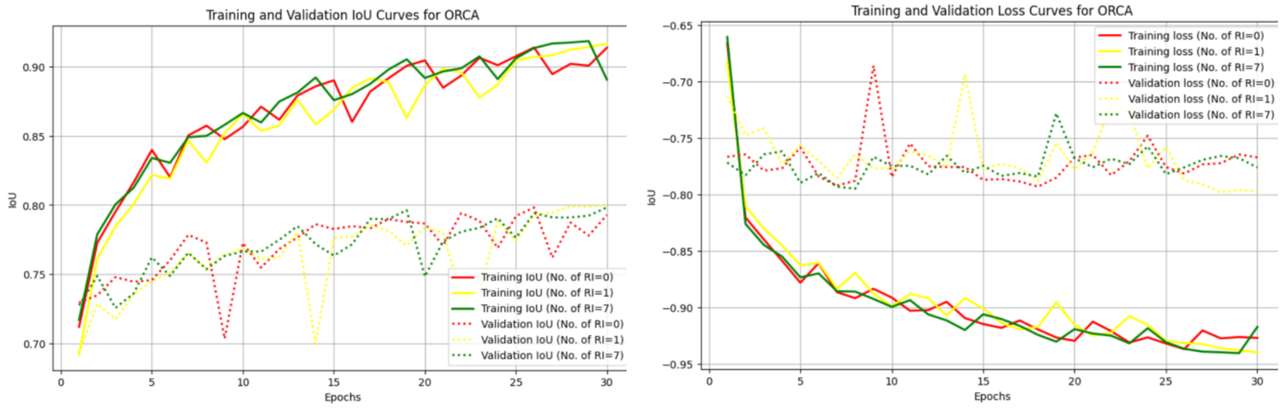


(**a**)



(**b**)

Figure 8. (**a**) **Training and Validation IOU curve for OCDC dataset; (b) Training and Validation Loss curve for OCDC dataset.**



(**a**) (**b**)

Figure 9. (**a**) **Training and Validation IOU curve for ORCA dataset; (b) Training and Validation Loss curve for ORCA dataset.**

During the model training with OCDC dataset the training and validation IOU curve and Loss obtained is represented in Figure 8 (a) and (b) respectively. Also, the training vs validation curve for ORCA dataset, both for IOU and Loss is represented in figure 9 (a) and (b) respectively.

5. Conclusions

In this study, we investigated the impact of Reinhard's stain normalization on deep learning-based semantic segmentation of Oral Squamous Cell Carcinoma (OSCC) histopathological images. Additionally, we proposed that stain normalization techniques can serve as a form of data augmentation in whole slide image (WSI) segmentation and classification tasks. The model employed for this research is the UNet architecture with EfficientNet-B0 as the encoder, chosen for its lightweight design and strong performance. Our results indicate that stain normalization did not yield significant improvements in segmentation accuracy on the ORCA dataset. However, a slight performance enhancement was observed in the OCDC dataset, where six reference images were utilized for data augmentation.

Funding: Please add: This research was funded by Indian Council of Medical Research, grant number IIRP-2023-7778" and "The APC was funded from the grant received from Indian Council of Medical Research.

References

1. Martino, F.; Bloisi, D.D.; Pennisi, A.; Fawakherji, M.; Ilardi, G.; Russo, D.; Nardi, D.; Staibano, S.; Merolla, F. Deep Learning-Based Pixel-Wise Lesion Segmentation on Oral Squamous Cell Carcinoma Images. Appl. Sci. 2020, 10, 8285. https://doi.org/10.3390/app10228285.
2. Pennisi, A.; Bloisi, D.D.; Nardi, D.; Varricchio, S. Multi-encoder U-Net for Oral Squamous Cell Carcinoma Image Segmentation. Available online: https://www.researchgate.net/publication/362864776 (accessed on 20 January 2025).
3. Musulin, J.; Štifanić, D.; Zulijani, A.; Čabov, T.; Dekanić, A.; Car, Z. An Enhanced Histopathology Analysis: An AI-Based System for Multiclass Grading of Oral Squamous Cell Carcinoma and Segmenting of Epithelial and Stromal Tissue. Cancers 2021, 13, 1784. https://doi.org/10.3390/cancers13081784.
4. Albishria, A.; Shah, S.J.; Lee, Y.; Wang, R. OCU-Net: A Novel U-Net Architecture for Enhanced Oral Cancer Segmentation. Available online: https://arxiv.org/abs/2310.02486 (accessed on 20 January 2025).
5. La Rosa, G.R.M.; Gattuso, G.; Pedullà, E.; Rapisarda, E.; Nicolosi, D.; Salmeri, M. Association of Oral Dysbiosis with Oral Cancer Development (Review). Oncol. Lett. 2020, 20, 11441. https://doi.org/10.3892/ol.2020.11441.
6. Kavyashree, C.; Vimala, H.S.; Shreyas, J. A Systematic Review of Artificial Intelligence Techniques for Oral Cancer Detection. Smart Healthc. Syst. 2024, 1, 100006. https://doi.org/10.1016/j.shs.2023.100006.
7. Veeraraghavan, V.P.; Daniel, S.; Dasari, A.K.; Reddy Aileni, K.; Patil, C.; Patil, S.R. Harnessing Artificial Intelligence for Predictive Modelling in Oral Oncology: Opportunities, Challenges, and Clinical Perspectives. Adv. Med. Oncol. 2024, 1, 100437. https://doi.org/10.1016/j.amo.2023.100437.

8. Kumar, Y.; Gupta, S.; Singla, R.; Hu, Y.C. A Systematic Review of Artificial Intelligence Techniques in Cancer Prediction and Diagnosis. Arch. Comput. Methods Eng. 2021, 28, 3949–3982. https://doi.org/10.1007/s11831-021-09648-w.
9. Falkeholm, L.; Grant, C.A.; Magnusson, A.; Möller, E. Xylene-Free Method for Histological Preparation: A Multicentre Evaluation. Lab. Investig. 2001, 81, 1213–1221.
10. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale. Available online: https://arxiv.org/abs/2010.11929v2 (accessed on 20 January 2025).
11. AL Qurri, A.; Almekkawy, M. Improved UNet with Attention for Medical Image Segmentation. Sensors 2023, 23, 8589. https://doi.org/10.3390/s23208589.
12. Wang, H.; Zhu, Y.; Green, B.; Adam, H.; Yuille, A.; Chen, L.-C. Axial-DeepLab: Stand-Alone Axial-Attention for Panoptic Segmentation. Available online: https://arxiv.org/abs/2003.07853 (accessed on 20 January 2025).
13. Guo, M.-H.; Lu, C.-Z.; Hou, Q.; Liu, Z.-N.; Cheng, M.-M.; Hu, S.-M. SegNeXt: Rethinking Convolutional Attention Design for Semantic Segmentation. Available online: https://arxiv.org/abs/2209.08575 (accessed on 20 January 2025).
14. Geng, Z.; Guo, M.-H.; Chen, H.; Li, X.; Wei, K.; Lin, Z. Is Attention Better than Matrix Decomposition? Available online: https://arxiv.org/abs/2109.04553 (accessed on 20 January 2025).
15. Shao, H.; Zeng, Q.-S.; Hou, Q.; Yang, J. MCANet: Medical Image Segmentation with Multi-Scale Cross-Axis Attention. Available online: https://arxiv.org/abs/2312.08866 (accessed on 20 January 2025).
16. Hatamizadeh, A.; Tang, Y.; Nath, V.; Yang, D.; Myronenko, A.; Landman, B.; Roth, H.; Xu, D. UNETR: Transformers for 3D Medical Image Segmentation. Available online: https://arxiv.org/abs/2103.10504 (accessed on 20 January 2025).
17. Shaker, A.; Maaz, M.; Rasheed, H.; Khan, S.; Yang, M.-H.; Khan, F.S. UNETR++: Delving into Efficient and Accurate 3D Medical Image Segmentation. Available online: https://arxiv.org/abs/2212.04497 (accessed on 20 January 2025).
18. Setty, S.; Srinath, N.K.; Hanumantharaju, M.C. Development of Multiscale Retinex Algorithm for Medical Image Enhancement Based on Multi-Rate Sampling. In Proceedings of the 2013 International Conference on Signal Processing and Pattern Recognition (ICSIPR 2013), Mysore, India, 15–17 April 2013; pp. 233–236. https://doi.org/10.1109/ICSIPR.2013.6497976.
19. Hoque, M.Z.; Keskinarkaus, A.; Nyberg, P.; Seppänen, T. Retinex Model-Based Stain Normalization Technique for Whole Slide Image Analysis. Comput. Med. Imaging Graph. 2023, 102, 101901. https://doi.org/10.1016/j.compmedimag.2021.101901.
20. Reinhard, E.; Adhikhmin, M.; Gooch, B.; Shirley, P. Color Transfer between Images. IEEE Comput. Graph. Appl. 2001, 21, 34–41. https://doi.org/10.1109/38.946629.
21. dos Santos, D.F.D.; de Faria, P.R.; Loyola, A.M.; Cardoso, S.V.; Travençolo, B.A.N.; do Nascimento, M.Z. Hematoxylin and Eosin Stained Oral Squamous Cell Carcinoma Histological Images Dataset. Available online: https://arxiv.org/abs/2303.10172 (accessed on 20 January 2025).
22. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. Available online: https://arxiv.org/abs/1706.03762 (accessed on 20 January 2025).
23. Murcia-Gómez, D.; Rojas-Valenzuela, I.; Valenzuela, O. Impact of Image Preprocessing Methods and Deep Learning Models for Classifying Histopathological Breast Cancer Images. Appl. Sci. 2022, 12, 11375. https://doi.org/10.3390/app122211375.
24. Li, H.; Zhang, L.; Shen, H. A Perceptually Inspired Variational Method for the Uneven Intensity Correction of Remote Sensing Images. IEEE Trans. Geosci. Remote Sens. 2012, 50, 3625–3633. https://doi.org/10.1109/TGRS.2011.2178075.
25. Zhang, D.; Park, W.-J.; Lee, S.-J.; Choi, K.-A.; Ko, S.-J. Histogram Partition-Based Gamma Correction for Image Contrast Enhancement. In Proceedings of the 2012 IEEE International Symposium on Consumer Electronics (ISCE 2012), Seoul, Republic of Korea, 4–6 June 2012; pp. 143–147. https://doi.org/10.1109/ISCE.2012.6241687.
26. Tan, M.; Le, Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. Available online: https://arxiv.org/abs/1905.11946 (accessed on 20 January 2025).
27. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Available online: https://arxiv.org/abs/1505.04597 (accessed on 20 January 2025).
28. Ferlay, J.; Colombet, M.; Soerjomataram, I.; Mathers, C.; Parkin, D.M.; Piñeros, M.; Znaor, A.; Bray, F. Estimating the Global Cancer Incidence and Mortality in 2018: GLOBOCAN Sources and Methods. Int. J. Cancer 2019, 144, 1941–1953.
29. Vigneswaran, N.; Williams, M.D. Epidemiologic Trends in Head and Neck Cancer and Aids in Diagnosis. Oral Maxillofac. Surg. Clin. North Am. 2014, 26, 123–141.

30. D'Souza, M.J.; Gala, R.P.; Ubale, R.V.; D'Souza, B.; Vo, T.P.; Parenky, A.C.; Zughaier, S. Trends in Non-parenteral Delivery of Biologics, Vaccines and Cancer Therapies. In Novel Approaches and Strategies for Biologics, Vaccines and Cancer Therapies; Academic Press: Cambridge, MA, USA, 2015; pp. 89–122.
31. Romano, A.; Di Stasio, D.; Petruzzi, M.; Fiori, F.; Lajolo, C.; Santarelli, A.; Contaldo, M. Noninvasive Imaging Methods to Improve the Diagnosis of Oral Carcinoma and Its Precursors: State of the Art and Proposal of a Three-Step Diagnostic Process. Cancers 2021, 13, 2864.
32. Safi, A.F.; Kauke, M.; Grandoch, A.; Nickenig, H.J.; Drebber, U.; Zöller, J.; Kreppel, M. Clinicopathological Parameters Affecting Nodal Yields in Patients with Oral Squamous Cell Carcinoma Receiving Selective Neck Dissection. J. Cranio-Maxillofac. Surg. 2017, 45, 2092–2096.
33. Atanasova, K.R.; Yilmaz, O. Looking in the Porphyromonas Gingivalis Cabinet of Curiosities: The Microbium, the Host and Cancer Association. Mol. Oral Microbiol. 2014, 29, 55–66.
34. Notes, Q.I.C.A. H and E Stain Tissue Documentation. Available online: https://www.qimaging.com/support/pdfs/he_technote.pdf (accessed on 20 January 2025).
35. Ates, G.C.; Mohan, P.; Celik, E. Dual Cross-Attention for Medical Image Segmentation. Available online: https://arxiv.org/abs/2303.17696 (accessed on 20 January 2025).
36. Bejnordi, B.E.; Litjens, G.; Timofeeva, N.; Otte-Höller, I.; Homeyer, A. Stain-Specific Standardization of Whole-Slide Histopathological Images. IEEE Trans. Med. Imaging 2015, 35, 858–871. https://doi.org/10.1109/TMI.2015.2476509.
37. Ciompi, F.; Geessink, O.; Bejnordi, B.E.; de Souza, G.S.; Baidoshvili, A.; Litjens, G.; van Ginneken, B.; Nagtegaal, I.; van der Laak, J. The Importance of Stain Normalization in Colorectal Tissue Classification with Convolutional Networks. Available online: https://arxiv.org/abs/1702.05931 (accessed on 20 January 2025).
38. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. Available online: https://arxiv.org/abs/1708.02002 (accessed on 20 January 2025).
39. Sudre, C.H.; Li, W.; Vercauteren, T.; Ourselin, S.; Cardoso, M.J. Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations. Available online: https://arxiv.org/abs/1707.03237 (accessed on 20 January 2025).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.