Tissue Artifact Segmentation and Severity Classification in Whole Slide Images

Galib Muhammad Shahriar Himel
Dept. of Computer Science,
American International University-Bangladesh
galib.muhammad.shahriar@gmail.com

***Abstract*—Pathological specimens are scanned using a whole slide imaging (WSI) scanner to create digital images for monitor-based diagnosis and analysis. The slide needs to be rescanned if the image quality is occasionally unsatisfactory because of focus-error or noise. An evaluation method for reference less quality was previously proposed, although some artifacts (such as tissue folds and air bubbles) were identified as false positives. Those artifacts must be disregarded when deciding whether or not rescanning is required because slide preparation, not scanning, is what causes them. By differentiating between the sources of quality degradation—the focus-error or noise created by the scanner and the artifact occurred during the slide preparation—this research suggests a strategy for a more practical approach to assess WSI quality. In the method, two UNet based architecture: DoubleUNet and ResUNet++ is used to segment the area in which the artifact is located. Besides the segmentation, the artifact severity is also analyzed using ensemble learning. To do the ensemble learning, at first transfer learning was implemented using various pre-trained model and then, from the generated model best models are selected for ensemble learning. In case of artifact segmentation above 97% accuracy is achieved and in the case of severity analysis 99.99% accuracy is gained which is almost perfect.**

Keywords*—tissue artifact, whole slide image, digital pathology, artifact segmentation, ensemble learning

I. Introduction

The development of a whole slide imaging (WSI) scanner made it possible to digitize pathological works [1], [2]. Glass slide specimens are converted into high-resolution digital images by the scanner for use in computational pathol- ogy and pathology practice. In hospitals and pathology labs, it enables us to undertake picture analysis and diagnosis using automated algorithms or human viewing on a computer screen. It pushes the boundaries of light microscopy and has great promise not just for research and education but also for primary and secondary diagnoses. However, it is important to take into account a number of factors before using a WSI scanner as a standard medical device, including image quality, color variability, inadequate consistency of slide preparation, and image format. In this research, we set our focus on mainly tissue artifact area segmentation and artifact severity analysis which are primarily base on image quality. In research labs, hospitals, medical colleges and pathology departments a whole Slide Imaging System would be used to scan biological specimens on a regular basis. WSI scanners may produce images of insufficient quality. Images of poor quality can seriously hinder diagnosis and render analysis ineffective. The operator is currently required by the WSI system to guarantee the accuracy of the scanned image. However, such manual evaluation is expensive and unreliable since it is susceptible to fatigue, inconsistent results, and prejudice. In order to provide access to high-quality photographs, it is important to analyze the quality of scanned images automatically. Failures during scanning might result in focus blur or noise being introduced into the WSI scanner, which is one of the main reasons of image quality loss. For scanning the glass slides, focus points can be manually or automatically chosen. When focus points are chosen from an area with a different focus depth than the normal tissue area, the surrounding area is rendered out of focus. The independent random value that differs from adjacent image pixels is known as noise. Fig. 1 demonstrates the noise, focus blur and under graded tissue area.

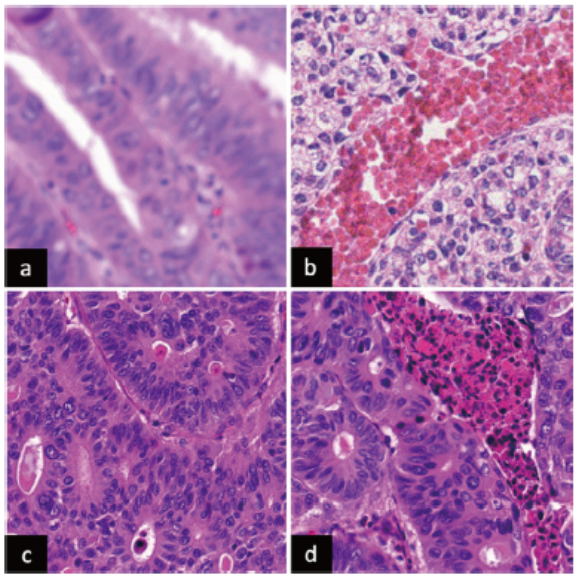

Fig. 1. Example of focus blur (a), noisy (b) and normal tissue (c) and (d) areas collected from [1].

It is important to rescan the slide if the focus error or noise significantly affects the scanned image.Hashimoto et al. previously suggested a reference less quality evaluation approach (RQM) to assess the scanned image's quality in order to identify slides that needed to be rescanned. The major causes of quality failure in the WSI scanner are sharpness and noise measurement, which is how this approach assesses the quality of images. Tissue artifacts are produced during the glass slide preparation stage, which is another reason why image quality fails

979-8-3503-0050-5/23/$31.00 ©2023 IEEE

[9]. The two main artifacts discovered in WSI are a tissue fold and an air bubble [10], as illustrated in Fig. 2 Tissue artifacts deceive analysis and diagnosis by concealing or changing important information. Furthermore, focusing errors are brought on by such objects [11]. They present various focus depths, and if focus locations are chosen from these artifacts, the surrounding area is rendered out of focus. Tissue artifacts like air bubbles and tissue fold would also be found using the prior RQM approach [8]. If a slide has a scanning problem, rescanning it can restore the image's quality but not remove tissue artifacts. It is useless to rescan the slide if the WSI quality is low due simply to tissue artifacts. For both analysis and quality assessment, artifacts should be found and disregarded. Therefore, to apply the But, simply removing all the artifacts is not a feasible solution. There can be such slides those contain minor artifacts while contain major tissue specimen which can be proven important to identify disease. So, those kind slides can generate important features for machine learning. But, due to presence of minor artifacts if those images are disregarded then again the same problem which is wrong diagnosis of disease can occur. That is why the severity of the artifact is to be determined before removing from the diagnosis process. In this research, we suggest adding the tissue artifact detection stage and the severity analysis of the artifacts, as shown in fig. 3. The suggested method uses a UNet-based image segmentation methodology to identify artifacts, then classifies the images based on their severity before deciding whether or not to exclude them from the diagnosis process.

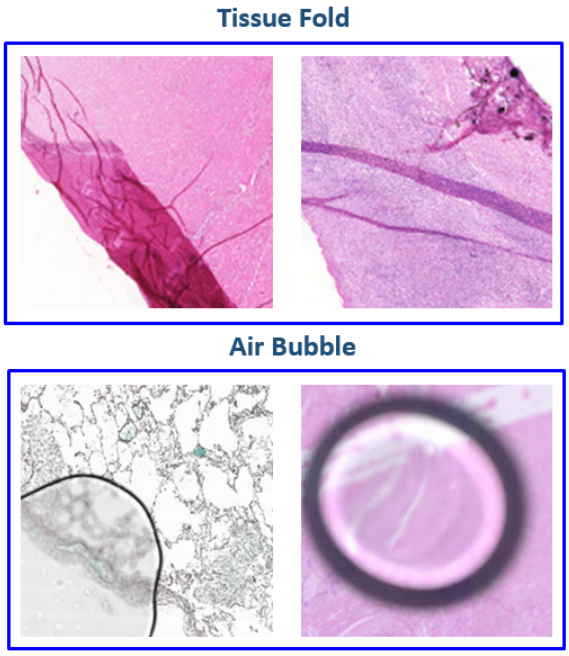

Fig. 2. Example of tissue fold and air bubble artifacts, collected from [2].

II. Literature Review

Although reduced-reference methods are available in some situations, methods for evaluating image quality can be roughly categorized into two groups: those that use reference images and those that do not. However, because a perfect reference picture is not available, reference-based evaluation cannot be used to identify a quality problem in the WSI scanner system. For the purpose of evaluating the quality of applications involving cameras, displays, video trans- mission, and codecs, references-less quality evaluation methods (RQMs) have been created [12][13][14][15][16][17][18][19][20]. Focus inaccuracy and noise are the key problems in WSI applications for man- aging the quality of digital pathology [8][21][22][23]. The RQMs have been used in this way, taking specific distortion into account. The method put out by Hashimoto et al. based on sharpness and noise assessment predicts both a sub- jective quality score as well as an objective score, such as a root mean square error, because computational image analysis is a crucial issue in both visual diagnostics and digital pathology. In order to gauge the reproducibility of the scanner, Shrestha et al. provided a method to assess the WSI's quality [21]. This approach, however, is unsuitable for practical applications because it depends on a reference image for evaluation. Another technique used deep learning to find out-of-focus areas, but it ignored noise-induced distortion [22]. Because it is effective and simple, the system suggested in this work is based on the method developed by Hashimoto et al. To distinguish between the tissue artifacts and scanning errors, it is vital to take into account the source of degradation while using the WSI scanning procedure in digital pathology. In several areas of computer vision, object detection or categorization in an image is carried out. Both supervised and unsupervised methods exist. The supervised method works well for the proposed system's air-bubble and tissue- fold detection because the unsupervised method could mix up artifacts with other anomalies that could have clinical significance. Template matching is a well-known and established supervised method for object detection. Statistical techniques like SVM and neural network classifiers have also been frequently used to identify objects in an image. Handcrafted features have been employed in this type of supervised classification, including local binary patterns, the gray level co-occurrence matrix (GLCM), and scale-invariant feature transform. It is well known that employing the sequential feature selection strategy when choosing the features for a classifier can increase performance [24][25]. More recently, non-handcrafted features that CNN gathered from the data were made possible by the development of deep learning technology. CNN can also be applied directly to an image. It has been demonstrated that a deep learning strategy, like CNN, performs better than a handmade feature based approach, while requiring a larger training dataset. When applied to a small dataset, it frequently becomes over fitted [26]. Transfer learning is typically employed in such instances to alleviate the problem of data limitation [27][28]. A network can avoid overfitting by using dropout and data augmentation [29][30]. In this work, we compared the segmentation method based on DoubleUNet and ResUNet++ with a VGG19 model as decoder. As for image classification process for severity analysis, we used various transfer learning methods for base model selection and lastly used several meta learners for stacking based ensemble learning. In histopathology, object detection using segmentation and image classification methods were used in previous researches. A review paper on importance of artifact detection and possibility of wrong disease diagnosis due to presence of artifact is published by Syed Ahmed Taqi at el.in the year 2018 [31]. In the year 2019, Kay R.J. Oskal at el. Published a paper regarding a UNet based tissue

segmentation method in WSI system [32]. An automated blur detection method is proposed by Xavier et al. in the year 2013 [33]. Samar khan published a paper regarding the probability of wrong diagnosis due to artifact in the year 2014 [34]. Ozan Oktay et al. introduced a unique attention gate (AG) model for medical imaging in 2018 that automatically learns to focus on target structures of various sizes and forms. Models trained with AGs intuitively learn to emphasize prominent features that are helpful for a particular task while suppressing irrelevant regions in an input image. Because of this, they can do away with the need for explicit external tissue/organ identification modules in cascaded convolutional neural networks (CNNs) [35]. A SVM based approach was taken by Hossain Md. Shakhawat in the year 2020 to evaluate image quality in WSI systems [7]. In the year 2019, Morteza Babaie et al. proposed a unique approach that mimerges the svm with dense layer 201 model to enhance the feature extraction process for medical images [36]. In the year, 2009, Pinky A. Bautista had done a research regarding artifact detection in WSI systems using basic object detection model [37]. In the year 2015, Hang Wu proposed a method regarding detecting blur artifacts in WSI [38]. Sonal Kothari published a research article on eliminating tissue fold artifact in the year 2013 [39]. Wei Lu et al. in the year 2013 researched on the presence of air bubble artifact in slides [40]. Pinky A. Bautista again in the year 2010 published another paper regarding enhancing the visualization and detection of tissue fold artifacts in whole slide imaging system [41]. So, from the literature, it is apparent that, continuous researches are being done by the researchers to further improve the accuracy of the detection of artifacts and also how to eliminate the artifacts automatically using different approaches. Existing methods only detects the tissue artifacts but does not analyze the severity of a particular artifact. But, severity analysis is important otherwise, significant amount of tissue can be eliminated from the specimen. Till now no work has been done to analyze the severity of the artifact. That's why we focused our research on mainly two things: 1. Improve the accuracy of artifact segmentation. 2. Analyze the artifact severity.

III. Methodology

The WSI scanner is used for scanning whole slide images. It is hard to determine the region of interest using human eye. So various approaches are being taken to enhance the automatic artifact detection. Among those approaches there are several machine learning approaches applied for advanced image processing. Pathological specimens are a bit complex to extract meaningful image data from. Because from a lot of images which are apparently similar, important dis- tinguishable features have to be extracted. For this purpose, we have proposed a method which will extract necessary information which will be used for success- ful feature extraction; ultimately leading to efficient pathological analysis and disease diagnosis. In our proposed method, we have applied image annotation, data pre-processing, segmentation, classification and visualization respectively in an organized manner.

A. Overview of The Proposed Method

In the preliminary stage, the images are captured using WSI system. The whole image will be divided into many small parts and segmentation is done for each part. The proposed method takes input images then detect the artifact using UNet based image segmentation approach. This segmentation process will identify the tissue artifact and will highlight that part while separating it from the background. After the segmentation process, the artifacts are being classified in terms of severity. For this classification task, Ensemble learning is applied. En- semble learning requires several methods. For selecting those methods transfer learning is applied using different combinations of parameters. After the severity classification being done, the final output is generated and sent for visualization. In the visualization screen, those identified artifacts along with their severity level will be shown and pathologists will be able to successfully pinpoint the affected region.

B. Artifact Detection Method

For, artifact detection we have used UNet based architecture for image segmen- tation. At first we used basic UNet. But the results were not good. Then we used 'DoubleUNet' and 'ResUNet++' as they have become quite popular in recent times for medical image segmentation tasks. DoubleUnet architecture uses VGG-19 model as encoder. The first layer of UNet generates an output which is multiplied with the primary input and that multiplied output is given as input to the second layer of UNet. The primary output and the secondary output concatenate and generates the final output. The ResUNet++ architecture follows a basic ResNet architecture and builds a bridge between the ResNet and UNet architecture. In simple words, it is a UNet architecture which follows ResNet architecture for segmentation.

C. Artifact Severity Classification

After the artifacts are being identified using segmentation techniques, the images need to be classified. For classification, we primarily used transfer learning using various pre-trained model. After getting the results we have compared the results according to the accuracy and loss values. Then we selected the best models to ensemble the results using stacking method. For implementing ensemble learning, we tested 10 different meta learners: Logistic Regression, K Nearest Neighbor, Support Vector Machine, Decision Tree, Random Forest Ada Boost Classifier, XGB Classifier, Gradient Boost Regressor, Gradient Boost Classifier and Gaussian Naïve Bias.

Severity analysis requires extensively complex calculation process for selecting differentiable features for image classification. The transfer learning models we used are not able to give perfect results from every perspective. Each model focuses on particular type of feature extraction policy. That's why, to combine those strategies, ensemble learning is used. Within the ensemble learning itself there are approaches those are specific to dif- ferent areas. Meta learners used in stacking policy acts in different ways. That's why several meta learners assessed to find out which meta learner generates the best results.

IV. Experimment

A. Dataset

In total 26 slides are used in this experiment, which included the tissues of different organs including kidney, liver, heart, brain etc. So the dataset is organ independent.

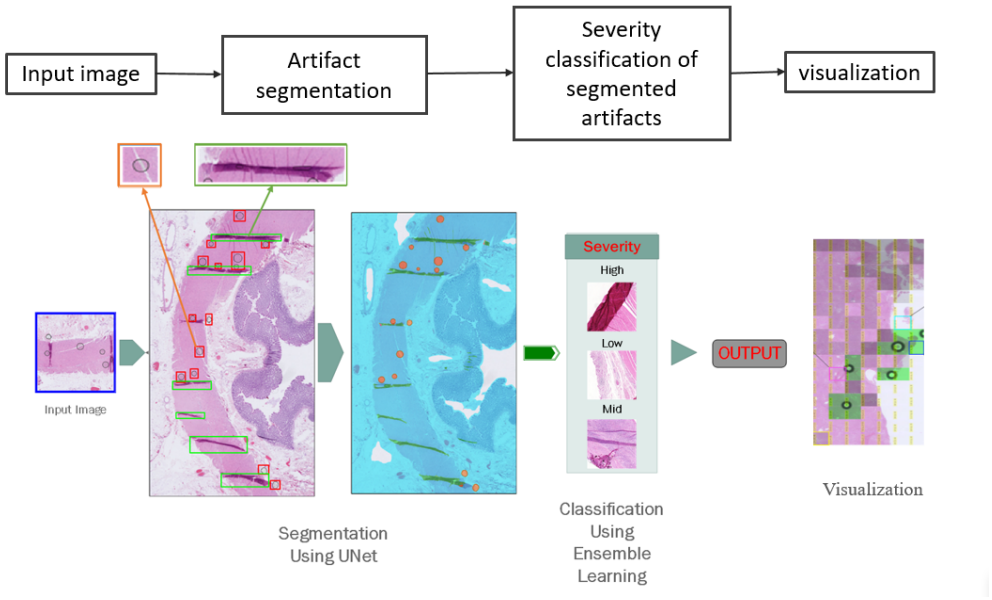

Fig. 3. Architecture of the proposed system.

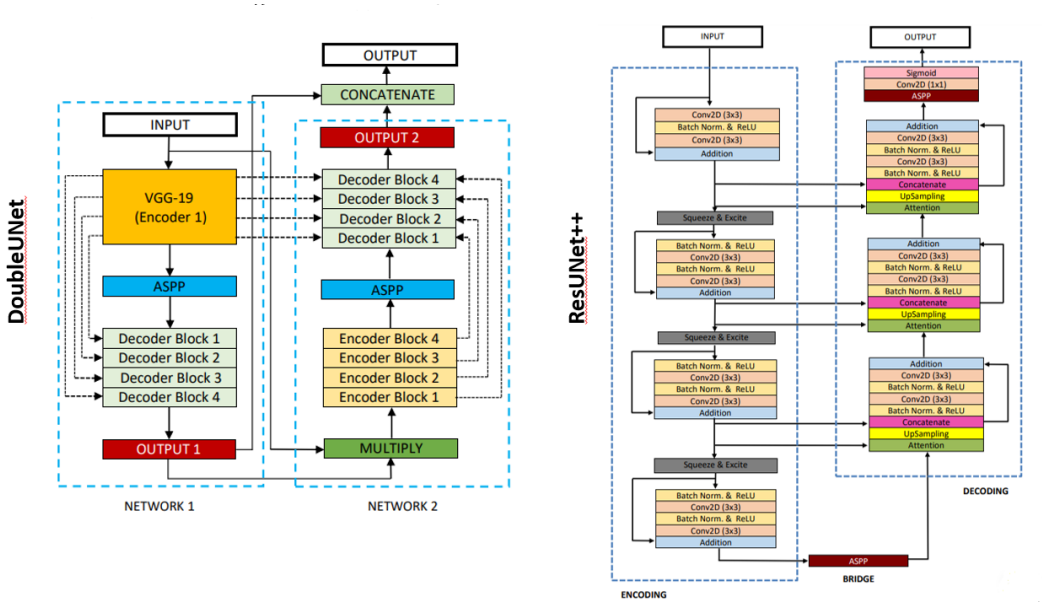

Fig. 4. Architecture of the artifact segmentation method.

Each category consists of 600 images. Data augmentation was applied for data balancing. Every images were segmented manually in adobe Photoshop by us- ing binary masking function. Each category consists of 600 images. Data augmentation was applied for data balancing. Every images were segmented manually in adobe Photoshop by using binary masking function. Taking images from the same source the dataset was categorized in three categories:High severity, Low severity and Mid severity. The dataset was again augmented for balancing purpose. This time the dataset contains 1380 images. This data annotation was done by certified expert.

B. Experimental Setup

Our experiment is divided in two parts: a) artifact segmentation; b) severity analysis/ classification. For the segmentation experiment the hyper parameters were tuned in different combinations

TABLE I. HYPER PARAMETER TUNING FOR SEGMENTATION EXPERIMENT

Train images	480
Valid images	60
Test images	60
Batch size	8
UNet architecture	**DoubleUNet,** ResUNet++
Optimizer	Adam, Adamax, **RMSprop**, SGD
Loss function	**Dice Coef Loss**, binary cross entropy
Learning Rate	Starting from $\mathbf{1e}^{-4}$
ReduceLROnPlateau	**monitor='val_loss', factor=0.1, patience=4**
EarlyStopping	**monitor='val_loss', patience=10**

Different epochs were tried but the result was not always uniformly tuned and that's why we used early stopping function to prevent overfitting and gain a reliable value for segmentation.

For classification task we used the following combinations described in table 2 and table 3.

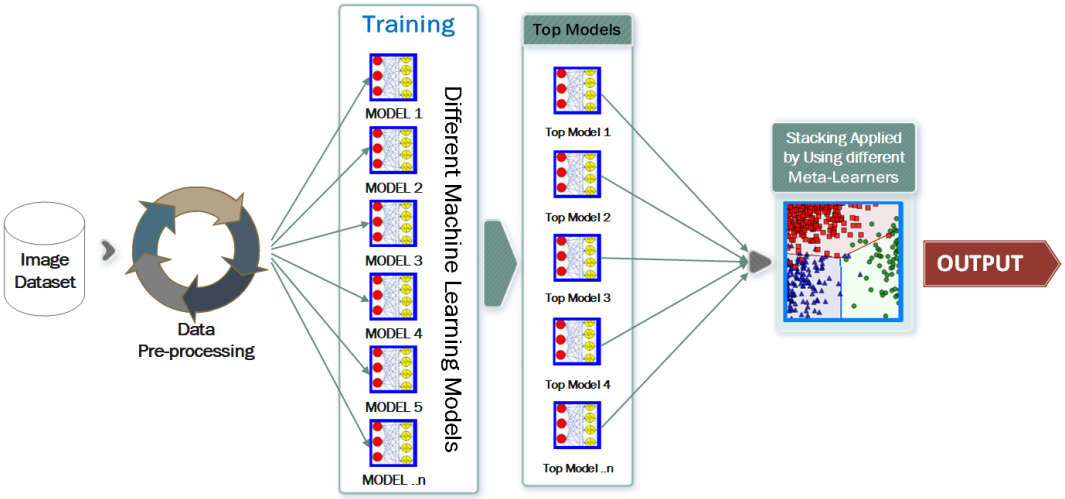

Fig. 5. Architecture of the artifact severity classification method.

TABLE II. HYPER PARAMETER OPTIMIZATION

Hyper parameter	Optimization Space
Epoch	25
Batch size	32
Learning rate	0.0001
Optimizer	'Adam', 'Adamax', 'RMSprop'
Loss function	[Categorical cross entropy], [Kullback Leibler Divergence]
Class mode	categorical

For implementing ensemble learning, we had to run transfer learning several times using different pre-trained model in various combinations. We used the most famous 10 pre-trained model from the official website of 'keras'.

V. RESULTS AND DISCUSSION

A. Artifact Segmentation Result

Table 4 represents the hyper-parameter combination for the best result for tissue fold artifact using ResUNet++ architecture. RMSprop optimizer performs better than any other optimizers. The learning rate was set to 0.0001. But, the 'ReduceLROnPlateau' function reduces the learning rate at every step where validation loss increases. The training was stopped when lr was $\mathbf{1e}^{-9}$.

Dice coef loss is defined as negative value that's why the loss values are negative.

Figure 10 represents the graph of mean IOU, Loss curve, precision curve, recall curve, dice-coefficient and the IOU. The last two graphs which are the dice-coefficient and IOU curve defines the actual performance of the model. In the next page, figure 11 represents some of the example comparison of the above mentioned model.

Table 5 represents the hyper-parameter combination for the best result for tissue fold artifact using Double-UNet++ architecture. RMSprop optimizer performs better than any other optimizers. The learning rate was set to 0.0001 and was consistent throughout the whole training process. The result was achieved at epoch 50.

Figure 12 represents the graph of mean IOU, Loss curve, precision curve, recall curve, dice-coefficient and the IOU. The last two graphs which are the dice- coefficient and IOU curve defines the actual performance of

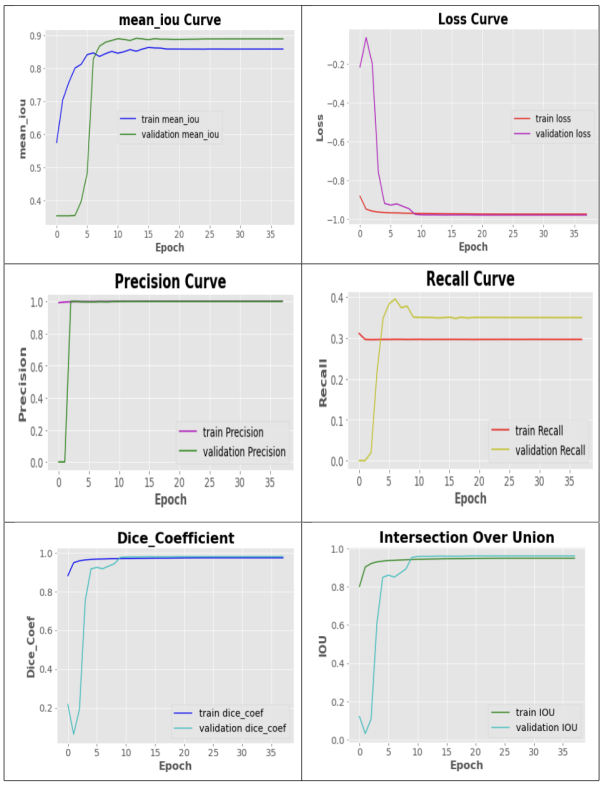

Fig. 6. Figure 10: Best results of ResUNet++ on Tissue-fold dataset.

the model. In the next page, figure 13 represents some of the example comparison of the above mentioned model.

Table 6 represents the hyper-parameter combination for the best result for air bubble artifact using ResUNet++ architecture. RMSprop optimizer performs better than any other optimizers. The learning rate was set to 0.0001. But, the 'ReduceLROnPlateau' function reduces the learning rate at every step where validation loss increases. The training was stopped when lr was $\mathbf{1e}^{-13}$.

Figure 15 represents the graph of mean IOU, Loss curve, precision curve, recall curve, dice-coefficient and the IOU. The last two graphs which are the dice- coefficient and IOU curve defines the actual performance of the model. In the next page, figure 16 represents some of

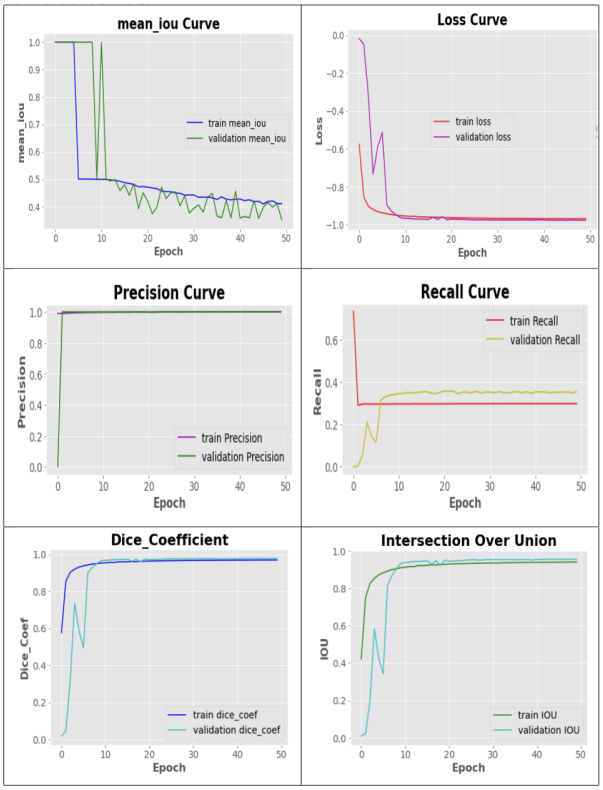

Fig. 7. Best results of DoubleUNet on Tissue-fold dataset.

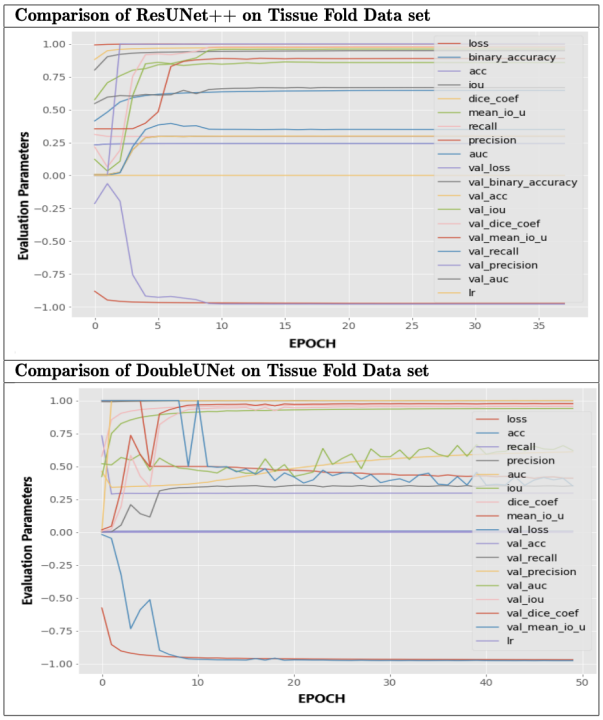

Fig. 8. Comparison graph for ResUnet++ and DoubleUNet for Tissue-fold dataset.

the example comparison of the above mentioned model.

Table 7 represents the hyper-parameter combination for the best result for air bubble artifact using DoubleUNet++ architecture. RMSprop optimizer per- forms better than any other optimizers. The learning rate was set to 0.0001 and was consistent throughout the whole training process. The result was achieved at epoch 50.

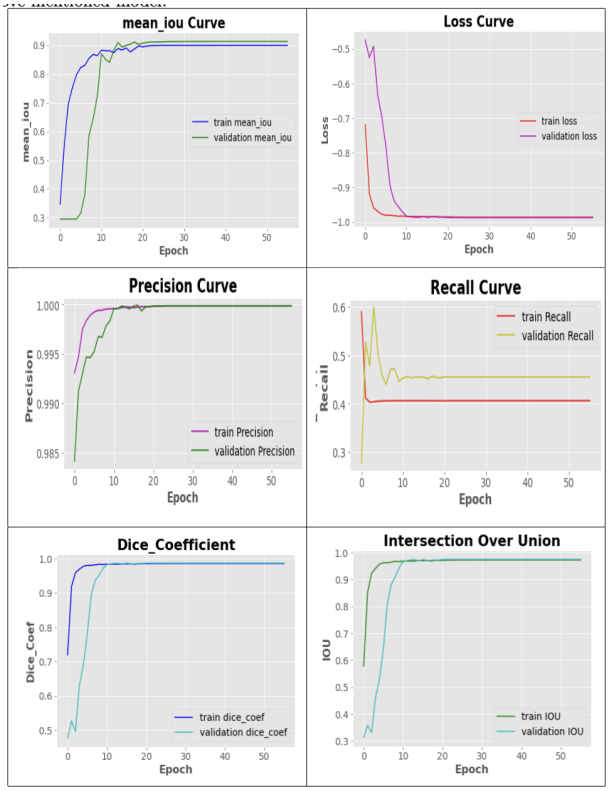

Fig. 9. Best results of ResUNet++ on Air-bubble dataset.

Figure 17 represents the graph of mean IOU, Loss curve, precision curve, recall curve, dice-coefficient and the IOU. The last two graphs which are the dice- coefficient and IOU curve defines the actual performance of the model.

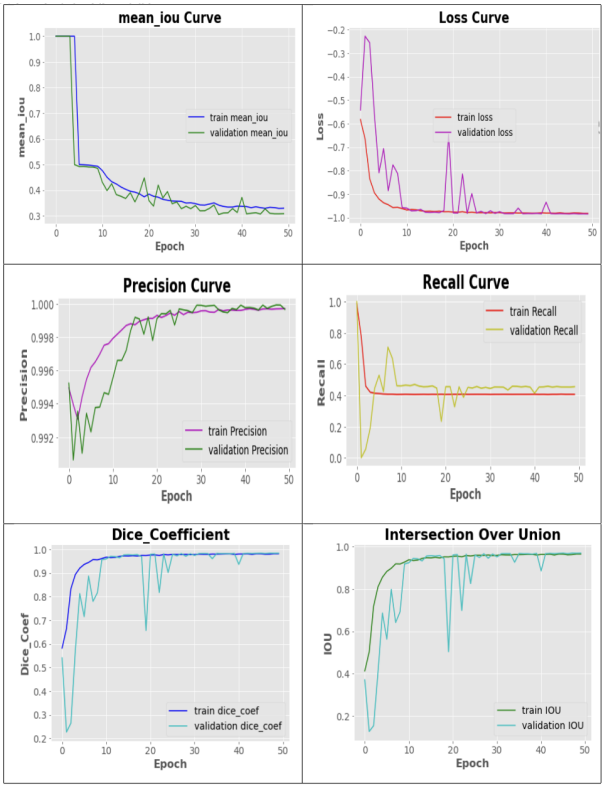

Fig. 10. Best results of DoubleUNet on Air-bubble dataset.

TABLE III. ACCURACY AND IOU FOR TISSUE FOLD DATASET

UNet	Optimizer	AVERAGE Test IOU	Test Accuracy (IOU Threshold = 90%)	Test Accuracy (IOU Threshold = 85%)
Double UNet	**Adamax**	96.54 %	90.00 %	95.00 %
	Adam	97.18 %	95.00 %	98.33 %
	RMSprop	98.02 %	98.33 %	100 %
	SGD	28.24 %	0 .00%	0.00 %
ResUNet++	**Adamax**	95.74 %	88.33 %	96.66 %
	Adam	96.95 %	96.66 %	100 %
	RMSprop	97.72 %	96.66 %	100 %
	SGD	79.62 %	36.66 %	50.00 %

TABLE IV. ACCURACY AND IOU FOR AIR BUBBLE DATASET

UNet	Optimizer	AVERAGE Test IOU	Test Accuracy (IOU Threshold = 90%)	Test Accuracy (IOU Threshold = 85%)
Double UNet	**Adamax**	99.08 %	100 %	100 %
	Adam	91.35 %	68.33 %	81.66 %
	RMSprop	99.11 %	100 %	100 %
	SGD	45. 96 %	0.00 %	0.00 %
ResUNet++	**Adamax**	97.43 %	96.66 %	98.33 %
	Adam	98.80 %	100 %	100 %
	RMSprop	98.15 %	96.66 %	96.66 %
	SGD	28.29 %	0.00 %	0.00 %

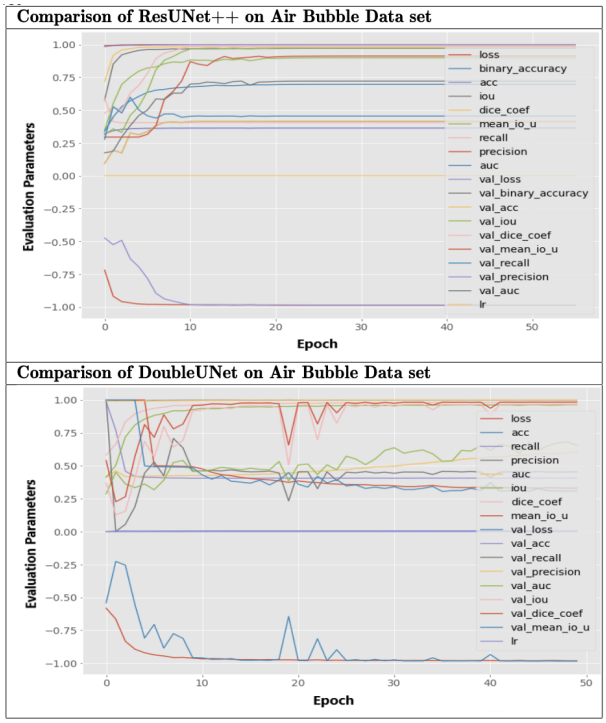

Fig. 11. Comparison graph of ResUNet++ and DoubleUNet for Air-bubble dataset.

After doing all the combinations, it was found that "RMSprop" performs better than the other three optimizers. And SGD generates the poorest result. So, in the next steps, for K-fold cross validation, "RMSprop" is used for both 'DoubleUNet' and 'ResUNet++'.

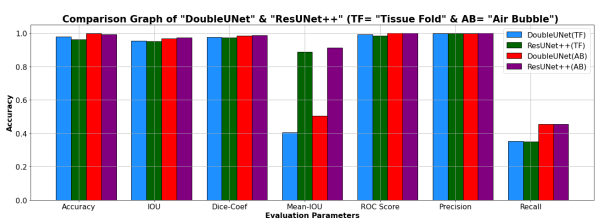

Fig. 12. Comparison graph of ResUNet++ and DoubleUNet.

Over all The DoubleUNet out performs the ResUNet++. But the results are almost similar. Only the mean IOU value is greater in case of ResUNet++. But, Overall accuracy of DoubleUnet is better. For the segmentation experiment, we can conclude that, our model works better for air-bubble dataset. Both models generate better result for air- bubble. While running 6-fold cross validation, in one or two cases, the result of ResUNet++ was better but overall the DoubleUNet works better which was explicitly clear throughout the experiment. But one interesting findings was for both dataset resUNet++ generates better Mean-IOU than the DoubleUNet when the actual assessment criteria for our result was IOU and dice-coefficient. Another finding from our segmentation experiment is that, the precision is over 99.9% while the recall is between 33%-45%. It means that, false positive rate is greater than false negative.

B. Severity Analysis/ Classification Result

This Table represents Methodology for base model selection from transfer learning approaches. Applying the methodology describing the Table, we have selected the best models which are described in the following section.

VI. CRITICAL ANALYSIS

We have so far covered the results and related discussion on both the segmenta- tion and classification tasks. In terms of the segmentation part, the segmentation accuracy is assessed based on some particular IOU threshold value instead of pixel wise accuracy. If the value is above 90% we considered it to be as a suc- cessful identification case. Based on this assessment criterion the segmentation result we got is quite good which is over 97% (while using DoubleUNet) for air bubble and over 96% (while using DoubleUNet) for tissue fold. Due to manual binary masking, some masking imperfection occurred as it couldn't have been avoided altogether. But, if perfect binary masking was possible, then we could get even better results in terms of segmentation accuracy. As for classification task, the levels of severity were determined by human eye. For this, the result can very if the annotation is done by another human. But, looking at the result, which is 99.99%, we can easily say

TABLE V. 6-FOLD CROSS VALIDATION FOR TISSUE FOLD DATASET.

	Fold no.	Training Set	Testing Set	Test Acc (IOU Threshold = 90%)
DoubleUNet	Fold 1	A, B, C, D, E	F	98.33 %
	Fold 2	B, C, D, E, F	A	96.66 %
	Fold 3	C, D, E, F, A	B	95.00 %
	Fold 4	D, E, F, A, B	C	98.33 %
	Fold 5	E, F, A, B, C	D	100 %
	Fold 6	F, A, B, C, D	E	100 %
Average				**97.96%**
ResUNet++	Fold 1	A, B, C, D, E	F	95.00 %
	Fold 2	B, C, D, E, F	A	96.66 %
	Fold 3	C, D, E, F, A	B	96.66 %
	Fold 4	D, E, F, A, B	C	95.00 %
	Fold 5	E, F, A, B, C	D	98.33 %
	Fold 6	F, A, B, C, D	E	96.66 %
Average				**96.38 %**

TABLE VI. 6-FOLD CROSS VALIDATION FOR AIR BUBBLE DATASET

	Fold no.	Training Set	Testing Set	Test Acc (IOU Threshold = 90%)
DoubleUNet	Fold 1	A, B, C, D, E	F	100 %
	Fold 2	B, C, D, E, F	A	100 %
	Fold 3	C, D, E, F, A	B	98.33 %
	Fold 4	D, E, F, A, B	C	100 %
	Fold 5	E, F, A, B, C	D	100 %
	Fold 6	F, A, B, C, D	E	100 %
Average				**99.72 %**
ResUNet++	Fold 1	A, B, C, D, E	F	96.66 %
	Fold 2	B, C, D, E, F	A	100 %
	Fold 3	C, D, E, F, A	B	100 %
	Fold 4	D, E, F, A, B	C	98.33 %
	Fold 5	E, F, A, B, C	D	100 %
	Fold 6	F, A, B, C, D	E	100 %
Average				**99.16 %**

TABLE VII. METHODOLOGY FOR BASE MODEL SELECTION FROM TRANSFER LEARNING APPROACHES

Pre-Trained Model	Optimizer	Loss Function	Validation Loss	Test Accuracy	ROC Score
1, 2, ... N	Best Optimizer Selection	Best Loss Function Selection	Value Comparison	Value Comparison	Value Comparison

TABLE VIII. HYPER PARAMETER OPTIMIZATION SPACE

Model	XCEPTION
Epoch	25
Batch size	32
Learning rate	0.0001
Optimizer	RMSprop
Loss function	Categorical cross entropy (CCE)

that, our model will definitely achieve 99% accuracy in terms of severity classification even if done by different experts providing that the binary masking should be close to perfect and the data annotation process should follow standard annotation techniques. Another thing is that, for severity analysis, regression is better than classification as no hard and fast rules applies to the severity labels, but in that case, the manual observation will be required even after visualization of the artifacts and human decision will be necessary. In our case, to avoid that scenario we implemented classification instead of regression.

VII. CONCLUSION

In this research, we worked on improving the tissue artifact segmentation ac- curacy to further enhance the detection mechanism of the whole slide imaging scanner. We have managed to achieve our goal using DoubleUNet and Re- sUNEt++ architecture; the DoubleUnet architecture performs better in most of the cases. We have also proposed an ensemble learning based classification model which analyze the severity of the detected artifact. Our model was tested on a dataset annotated by experts in this field. This will help the WSI to take decision on removing artifacts from diagnosis process according to the severity detected by our model. This proposed model will enhance the WSI experience and lead the digital pathology to a new level, making the diagnosis process easier than ever before and also more accurate. Our future plan is to apply regression for severity analysis instead of classification and also to reconstruct the artifact containing region and to check whether the reconstructed area can be used in meaningful diagnosis or not. If these can be done successfully, the digital pathology will be elevated to a new step and the future of pathological image analysis will be brighter than ever before.

REFERENCES

[1] Shakhawat, H., Nakamura, T., Kimura, F., Yagi, Y. & Yamaguchi, M. Automatic quality evaluation of whole slide images for the practical use of whole slide imaging scanner. *ITE Transactions On Media Technology And Applications*. **8**, 252-268 (2020)

[2] Hossain, M., Nakamura, T., Kimura, F., Yagi, Y. & Yamaguchi, M. Practical image quality evaluation for whole slide imaging scanner. *Biomedical Imaging And Sensing Conference*. **10711** pp. 203-206 (2018)

[3] Bindhu, P., Krishnapillai, R., Thomas, P. & Jayanthi, P. Facts in artifacts. *Journal Of Oral And Maxillofacial Pathology: JOMFP*. **17**, 397 (2013)

[4] Ekundina, V. & Eze, G. Common artifacts and remedies in

TABLE IX. TOP 6 MODELS DESCRIPTION

Model name	Combination			Loss	Accuracy
	Pretrained Model	Optimizer	Loss function		
Model 1	XCEPTION	RMSprop	CCE	0.01031494	0.99761903
Model 2	VGG19	RMSprop	CCE	0.02167457	0.99761903
Model 3	MobileNet	RMSprop	KLD	0.00927471	0.99761903
Model 4	MobileNetV2	RMSprop	CCE	0.00240695	0.99761903
Model 5	DenseNet121	RMSprop	CCE	0.00346601	0.99761903
Model 6	MobileNetV2	RMSprop	KLD	0.02067174	0.99761903

TABLE X. ENSEMBLE LEARNING COMBINATIONS USING 10 META LEARNERS.

Models	Accuracy (%)									
	LR	KNN	SVM	DT	RF	ABC	XGB	GBR	GBC	GNB
Top 2	99.99	99.99	99.99	99.82	99.99	99.99	99.99	99.99	99.99	99.99
Top 3	99.99	99.99	99.99	99.99	99.99	99.99	99.99	99.99	99.99	99.99
Top 4	99.99	99.99	99.99	99.64	99.82	99.99	99.82	99.82	99.99	99.99
Top 5	99.99	99.99	99.99	99.64	99.99	99.99	99.82	99.99	99.82	99.99
Top 6	99.99	99.99	99.99	99.82	99.99	99.99	99.99	99.82	99.82	99.82

histopathology (a review). *African Journal Of Cellular Pathology*. pp. 1-7 (2015)

[5] LB, T. & Zaki, S. A review of artifacts in histopathology J. *Oral Maxillof. Pathol*. **22**, 279 (2018)

[6] Wright, A., Dunn, C., Hale, M., Hutchins, G. & Treanor, D. The effect of quality control on accuracy of digital pathology image analysis. *IEEE Journal Of Biomedical And Health Informatics*. **25**, 307-314 (2020)

[7] Hossain, M., Syeed, M., Fatema, K. & Uddin, M. The Perception of Health Professionals in Bangladesh toward the Digitalization of the Health Sector. *International Journal Of Environmental Research And Public Health*. **19**, 13695 (2022)

[8] Pantanowitz, L., Szymas, J., Yagi, Y. & Wilbur, D. Whole slide imaging for educational purposes. *Journal Of Pathology Informatics*. **3**, 46 (2012)

[9] Kumar, N., Gupta, R. & Gupta, S. Whole slide imaging (WSI) in pathology: current perspectives and future directions. *Journal Of Digital Imaging*. **33**, 1034-1040 (2020)

[10] Hossain, M., Syeed, M., Fatema, K., Hossain, M. & Uddin, M. Singular Nuclei Segmentation for Automatic HER2 Quantification Using CISH Whole Slide Images. *Sensors*. **22**, 7361 (2022)

[11] Rosenberg, A., Palmer, M., Merlino, L., Troost, J., Gasim, A., Bagnasco, S., Avila-Casado, C., Johnstone, D., Hodgin, J., Conway, C. & Others The application of digital pathology to improve accuracy in glomerular enumeration in renal biopsies. *PLoS One*. **11**, e0156441 (2016)

[12] Mariani, L., Bomback, A., Canetta, P., Flessner, M., Helmuth, M., Hladunewich, M., Hogan, J., Kiryluk, K., Nachman, P., Nast, C. & Others CureGN study rationale, design, and methods: establishing a large prospective observational study of glomerular disease. *American Journal Of Kidney Diseases*. **73**, 218-229 (2019)

[13] Hossain, M., Hanna, M., Uraoka, N., Nakamura, T., Edelweiss, M., Brogi, E., Hameed, M., Yamaguchi, M., Ross, D. & Yagi, Y. Automatic quantification of HER2 gene amplification in invasive breast cancer from chromogenic in situ hybridization whole slide images. *Journal Of Medical Imaging*. **6**, 047501 (2019)

[14] Kanwal, N., Pérez-Bueno, F., Schmidt, A., Engan, K. & Molina, R. The Devil is in the Details: Whole Slide Image Acquisition and Processing for Artifacts Detection, Color Variation, and Data Augmentation: A Review. *IEEE Access*. **10** pp. 58821-58844 (2022)

[15] Kothari, S., Phan, J. & Wang, M. Eliminating tissue-fold artifacts in histopathological whole-slide images for improved image-based prediction of cancer grade. *Journal Of Pathology Informatics*. **4**, 22 (2013)

[16] Bautista, P. & Yagi, Y. Detection of tissue folds in whole slide images. *2009 Annual International Conference Of The IEEE Engineering In Medicine And Biology Society*. pp. 3669-3672 (2009)

[17] Kanwal, N., Fuster, S., Khoraminia, F., Zuiverloon, T., Rong, C. & Engan, K. Quantifying the effect of color processing on blood and damaged tissue detection in Whole Slide Images. *2022 IEEE 14th Image, Video, And Multidimensional Signal Processing Workshop (IVMSP)*. pp. 1-5 (2022)

[18] Babaie, M. & Tizhoosh, H. Deep features for tissue-fold detection in histopathology images. *European Congress On Digital Pathology*. pp. 125-132 (2019)

[19] Palokangas, S., Selinummi, J. & Yli-Harja, O. Segmentation of folds in tissue section images. *2007 29th Annual International Conference Of The IEEE Engineering In Medicine And Biology Society*. pp. 5641-5644 (2007)

[20] Kothari, S., Phan, J., Stokes, T. & Wang, M. Pathology imaging informatics for quantitative analysis of whole-slide images. *Journal Of The American Medical Informatics Association*. **20**, 1099-1108 (2013)

[21] Foucart, A., Debeir, O. & Decaestecker, C. Artifact identification in digital pathology from weak and noisy supervision with deep residual networks. *2018 4th International Conference On Cloud Computing Technologies And Applications (Cloudtech)*. pp. 1-6 (2018)

[22] Janowczyk, A., Zuo, R., Gilmore, H., Feldman, M. & Madabhushi, A. HistoQC: an open-source quality control tool for digital pathology slides. *JCO Clinical Cancer Informatics*. **3** pp. 1-7 (2019)

[23] Assessment of Residual Breast Cancer Cellularity after Neoadjuvant Chemotherapy using Digital Pathology [Data set]. (https://doi.org/10.7937/TCIA.2019.4YIBTJNO)

[24] Jha, D., Riegler, M., Johansen, D., Halvorsen, P. & Johansen, H. Doubleu-net: A deep convolutional neural network for medical image segmentation. *2020 IEEE 33rd International Symposium On Computer-based Medical Systems (CBMS)*. pp. 558-564 (2020)

[25] Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. *International Conference On Medical Image Computing And Computer-assisted Intervention*. pp. 234-241 (2015)

[26] Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. *ArXiv Preprint ArXiv:1409.1556*. (2014)

[27] Mohammed, M., Mwambi, H., Mboya, I., Elbashir, M. & Omolo, B. A stacking ensemble deep learning approach to cancer type classification based on TCGA data. *Scientific Reports*. **11**, 1-22 (2021)

[28] Su, R., Liu, X., Xiao, G. & Wei, L. Meta-GDBP: a high-level stacked regression model to improve anticancer drug response prediction. *Briefings In Bioinformatics*. **21**, 996-1005 (2020)

[29] Hossain, M., Raihan, M., Hossain, M., Syeed, M., Rashid, H. & Reza, M. Aedes Larva Detection Using Ensemble Learning to Prevent Dengue Endemic. *BioMedInformatics*. **2**, 405-423 (2022)

[30] Shakhawat, H., Hanna, M., Ibrahim, K., Serrette, R., Ntiamoah, P., Edelweiss, M., Brogi, E., Hameed, M., Yamaguchi, M., Ross, D. & Others Automatic Grading of Invasive Breast Cancer Patients for the Decision of Therapeutic Plan. *Artificial Intelligence For Disease Diagnosis And Prognosis In Smart Healthcare*. **7** pp. 123 (2023)

[31] Shakhawat, H., Hossain, S., Kabir, A., Mahmud, S., Islam, M. & Tariq, F. Review of Artifact Detection Methods for Automated Analysis and Diagnosis in Digital Pathology. *Artificial Intelligence For Disease Diagnosis And Prognosis In Smart Healthcare*. pp. 177-202 (2023)

[32] Hossain, M., Hasan, N., Samad, M., Shakhawat, H., Karmoker, J.,

Ahmed, F., Fuad, K. & Choi, K. Android Ransomware Detection From Traffic Analysis Using Metaheuristic Feature Selection. *IEEE Access*. **10** pp. 128754-128763 (2022)

[33] Sifat, S., Hossain, M., Tonny, S., Majumder, B., Mahajabin, R. & Shakhawat, H. Android ransomware attacks detection with optimized ensemble learning. *International Conference On Cybersecurity, Cybercrimes, And Smart Emerging Technologies*. pp. 41-53 (2022)

[34] Shatil, S., Tabassum, M., Ahmed, M., Karim, F., Rahman, M. & Hossain, M. Polyps Segmentation using DoubleU-Net from Colonoscopy Images of Gastrointestinal Tract.

[35] Hossain, M., Salsabil, U., Syeed, M., Rahman, M., Fatema, K. & Uddin, M. SmartPoultry: Early Detection of Poultry Disease from Smartphone Captured Fecal Image. *2023 20th International Joint Conference On Computer Science And Software Engineering (JCSSE)*. pp. 345-350 (2023)

[36] Hossain, M., Shahriar, G., Syeed, M., Uddin, M., Hasan, M., Hossain, M. & Bari, R. Tissue Artifact Segmentation and Severity Assessment for Automatic Analysis Using WSI. *IEEE Access*. **11** pp. 21977-21991 (2023)