Classification of Precancerous Colorectal Lesions via ConvNeXt on Histopathological Images

Mehmet Nergiz

Abstract—In this translational study, the classification of precancerous colorectal lesions is performed by the ConvNeXt method on MHIST histopathological imaging dataset. The ConvNeXt method is a modernized ResNet-50 architecture having some training tricks inspired by Swin Transformers and ResNeXt. The performance of the ConvNeXt models are benchmarked on different scenarios such as 'full data', 'gradually increasing difficulty based data' and 'k-shot data'. It is shown that the ConvNeXt model outperformes almost all the other studies in the literature which are applied on MHIST by using ResNet models, vision transformers, weight distillation, self-supervised learning and curriculum learning strategy based on different scenarios and metrics. The ConvNeXt model trained with 'full data' yields the best result with the score of 0.8890 for accuracy, 0.9391 for AUC, 0.9121 for F1 and 0.7633 for cohen's cappa. The power of ConvNeXt is found as promising for classifying precancerous histopathological images and may be a good base line for miscellaneous tasks of computational pathology field with respect to the classical convolutional neural networks and vision transformers..

Index Terms—**ConvNeXt, CNN, Vision Transformer, Colorectal Cancer, Histopathology.**

I. INTRODUCTION

COLORECTAL CANCER is the reported as the world's third most common cancer type and it's at the fourth rank in terms fatality whereas there are only 3.94 pathologists per 100000 people in United States [1]–[3]. As a result of this imbalanced burden, classifying colorectal polyps which grow inside the colon lining and may transform to colonic cancer if not treated early enough is one of the time and energy consuming tasks in pathology [4]. The classification task of the colorectal polyps in which the inter-rater variability is significantly high mainly targets to identify hyperplastic polyps (HPs) and sessile serrated adenomas (SSAs) which are respectively benign and precancerous lesions to be treated immediately [4].

MEHMET NERGİZ, is with Department of Computer Engineering University of Dicle, Diyarbakır, Turkey,(e-mail: mnergiz@dicle.edu.tr).

https://orcid.org/0000-0002-0867-5518

The sort of challenges like high-resolution and variable size images as well as costly annotations which do not have clear guidelines constitute a barrier for the researchers that have limited resources to enter the field and apply different approaches and deep learning models. In order to lower that barrier, Wei et al. proposed MHIST which is planned to be an environment where different deep learning models to be benchmarked easily and then applied to large scale datasets acting just like a petri dish for histopathology image analysis [4], [5]. MHIST by having a limited amount of images can be classified as having a low data regime and is also a good fit for benchmarking few-shot learning which is a concept to define the ability of a machine learning model to learn from a restricted amount of data [6], [7]. The idea of few-shot learning on histopathology imaging is worth to be analyzed since collecting, preprocessing and annotating steps of this particular medical field is very costly and time consuming [8], [9]. Additionally, the histopathological imaging datasets which will be proposed in the future for different or rare diseases will probably start with a small number of examples and some of the large scaled current state of the art deep learning methods may not be tuned well for them.

This article is organized as follows: Related Works section includes a review of recent works using ConvNeXt method or MHIST imaging dataset. Material section gives the details about the MHIST dataset. Method section explains the details and tricks proposed by the ConvNeXt method. Results section presents the benchmarking scenarios and the obtained results comparatively. Discussion section presents some analyses about the obtained results for each proposed benchmarks and the compares this study with respect to the related literature. Conclusion boils down the results and figures out the future work.

A. Related Work

In most of the histopathology images, the dissemination and progression of diseases show a gradually increasing pattern and the levels of difficulty of classification changes on time [10]. MHIST has a natural range of difficulty based on the majority voting of the annotation of the pathologists. Thus, Wei et al. applied curriculum learning strategy in which the training examples are presented in an increasing level of difficulty and a better performance is obtained compared to the random sampling of training examples [4], [5], [10]–[12]. More and more, self-supervised learning based approach which exploits and learns from a relatively large scale

unlabeled in domain data and then fine-tuned in downstream data is used by Wang et al., Srinidhi and Martel [12], [13]. ResNet based deep learning models are used as a gold standard Convolutional Neural Networks (CNN) method for MHIST dataset whereas Wang et al. not only utilized CNN but also added a token-aggregating and excitation module and a transformer to their model [13]. Tasdemir et al., applied ConvNeXt models on data the Colorectal Polyps of UniToPatho , EBHI and Kayseri City Hospital [14]. Alternatively, Zhang et al. proposed a cross knowledge transfer among 9 different histopathological image datasets which is basically composed of downstream fine-tuning and weight distillation procedures aiming to increase the performance of vanilla ResNet-18 model [15].

To sum up, this study proposes the following contributions:

1- The ConvNeXt models are benchmarked on MHIST and almost all the other methods are outperformed with respect to various evaluation criteria like Accuracy, F1 score, AUC and Cohen's Kappa.

2- The ConvNeXt models are benchmarked to the subsets of MHIST having k-shots of images and different levels of difficulties.

3- The AUC of the ConvNeXt model on 'full data' scenario is at least 8.53% higher than of the baseline CNNs such as AlexNet, DenseNet, SqueezeNet, Inception v3 and VGG16.

II. MATERIAL

This dataset is obtained from scanning 328 Formalin Fixed Paraffin-Embedded (FFPE) WSI of colorectal polyp tissues of the patients at the Dartmouth- Hitchcock Medical Center [4]. Each scanned WSI which is scanned by an Aperio AT2 scanner at 40x resolution is diagnosed as sessile serrated adenomas (SSAs) or hyperplastic polyps (HPs) as shown in Fig. 1., 3152 image tiles which are having 224x224 resolution extracted from the obtained images. The permission for releasing and using this dataset is accepted by the Dartmouth-Hitchcock Health IRB [4].

The annotation process is performed by seven board certified gastrointestinal pathologists at the Dartmouth Hitchcock Medical Center [4]. Each histopathological image tile is classified as SSA or HP with respect to World Health Organization criteria from 2019 [16]. The label class of each image is gathered individually from each of the pathologists and the final gold standard labels are assigned based on the majority vote of the seven labels, as commonly performed at the most of the studies in the literature [6]. In this dataset, 16.7% of the images have 4/7 agreement which implies that there is a discordance among the annotators and a nontrivial challenge on the colorectal polyp classification task. The average Cohen's Kappa of the performance of the per-pathologist is 0.45 [17]. In this study, aligned with the other studies, to propose different training scenarios, four different levels of difficulty are defined based on the agreement of the annotators such as: very easy(VE) (7/7 agreement), easy(E) (6/7 agreement), hard(H) (5/7 agreement) and very hard(VH) (4/7 agreement) [17].
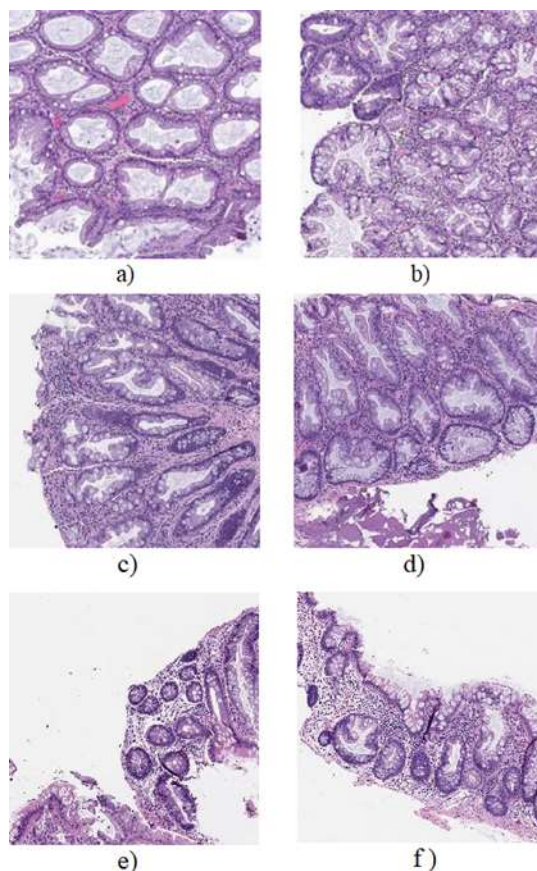

Fig.1. Two characteristic images for both of the classes of MHIST dataset. a) HP (benign) b) SSA (precancerous) c) HP d) SSA

III. METHOD

The main flowchart of the proposed study is given in Fig. 2. The first procedure is to design benchmarking data for 'full data', 'gradually increasing difficulty based data' and 'k-shot data' and applying ConvNeXt-L to these benchmarking datasets. Then, different variants of ConvNeXt is applied on the 'full data'. Afterwards, the behavior of ConvNeXt is observed on 'k-shot data' by applying Grad-Cam.

A. ConvNeXt

The Vision Transformer (ViT) based methods have begun to be addressed as one of the best image classifiers since 2020. The ViT methods have no image related inductive bias and are not much different from the Natural Language Processing (NLP) Transformers except for patchifying the input images into a sequence of smaller patches and their global attention design suffers from the quadratic computational complexities during handling high resolution images [18]. The Swin Transformer (ST) methods overcome that computational complexity issue by introducing sliding windows and can easily exceed the standard ResNet

Fig.2. The Main flowchart of the study
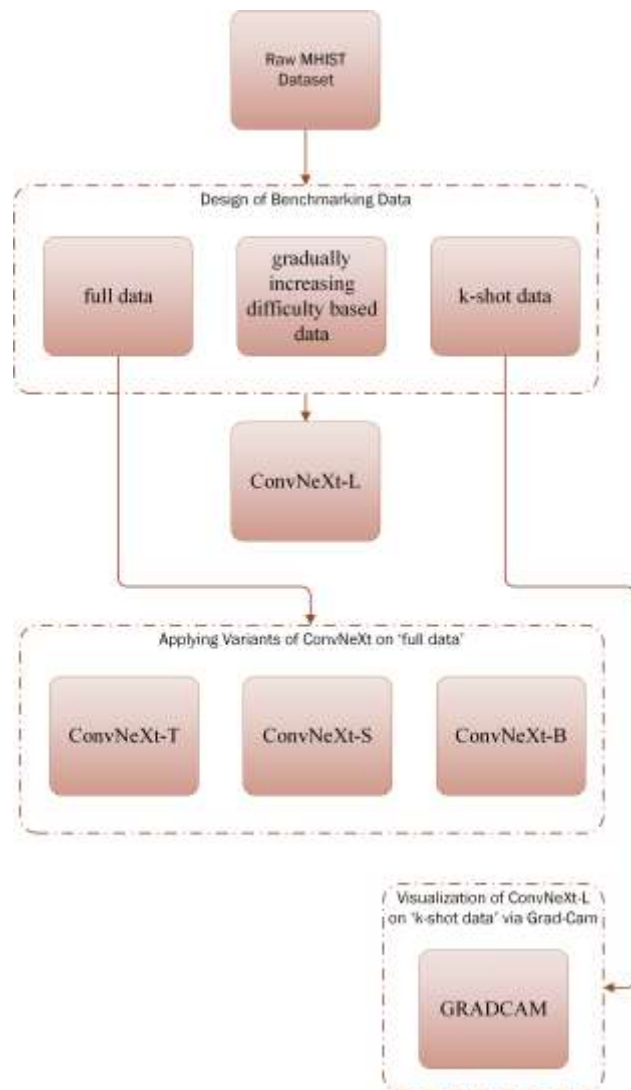
models [19]. ConvNeXt method, as a manifestation of ConvNets for 2020s, is the name of the set of pure convolutional methods (ConvNet) modified by taking into account the novel features of ViT and ST [20]. In this study, the ConvNeXt method, which is specifically the implementation of the modernization of the standard ResNet-50 model towards the design of ST methods, is used for the classification of precancerous colorectal lesions on MHIST image dataset [20]. The gradual steps of the modernization of ResNet-50 are outlined in Fig. 3 and explained in detail as below [20]:

1) *Training procedure*
   The training procedure of ConvNeXt is similar to ST and Data-Efficient Image Transformers (DeiT) [19]–[21]. The ConvNeXt model used in this study is pretrained on ImageNet-22K for 90 epochs with a warmup of 5 epochs and fine-tuned on ImageNet-1K for 30 epochs with no warmup. AdamW optimizer, cosine learning rate schedule, a learning rate of 5e-5, data augmentation methods like RandAugment, Cutmix, RandomErasing and Mixup [22], and regularization techniques such as Label Smoothing and Stochastic Depth are used [20].

2) *Macro Design*
   - Stage Compute Ratio: The number of blocks in each stage of ResNet-50 is converted from (3, 4, 6, 3) to (3, 3, 9, 3) like in ST [20].
   - Stem Cell: The stem cell of standard ResNet which contains a 7x7 convolution layer having stride 2 and a max pool is replaced by a patchifying layer using a 4x4 non-overlapping, stride 4 convolution [20].

3) *ResNeXtify*
   The ResNeXt method, like ViT architectures, has grouped convolution concept in its bottleneck block to separate the spatial and channel mixing. The ConvNeXt method uses depthwise convolution that is a sort of grouped convolution having groups as many as the number of the channels. Additionally, the network width of ConvNeXt is increased from 64 to 96 similar to ResNeXt as shown in Fig. 4 [20], [23].

4) *Inverted Bottleneck*
   Each Transformer block is designed as having an inverted bottleneck in which the dimension of the hidden layer is 4 times wider than its input. Inspired by this design, the block shape of ResNet-50 is inverted from a) to b) as in Fig. 4 [20]

5) *Large Kernel Sizes*
   - Moving up depthwise conv layer: The depthwise conv layer is moved up in order to apply this more complex operation to fewer channels as shown in Fig. 4 b) to c) [20].
   - Increasing the kernel size: The kernel sizes are increased to 7x7 [20].

6) *Micro Design*
   - Changing Rectified Linear Unit (ReLU) with Gaussian Error Linear Unit (GELU): The ReLU activation functions are replaced with the GELU that is used by the most of the sophisticated Transformers [20].
   - Fewer activation functions: Except for the one block which is shown in Fig. 5, all the other GeLU layers are removed from the ConvNeXt [20].
   - Fewer normalization layers: Except for the one layer that is shown in Fig. 5, all the other Batch Normalization (BN) layers are eliminated from the ConvNeXt [20].
   - Substituting BN with Layer Normalization (LN) : The BN is replaced by the simpler LN that is used in Transformers [20].
   - Separate down sampling layers: The down sampling strategy of ResNet which is using 3x3 convolution layer with stride 2 is changed to a separate 2x2 convolution

layer with stride 2 which is plugged between each stage [20].

In this study, the ConvNeXt-L model variant having C = (192; 384; 768; 1536) channels and B = (3; 3; 27; 3) blocks in each stage is used for the classification of precancerous colorectal lesions on MHIST image dataset with a batch size of 8. The Google Colab envoriment having Nvidia K80 / T4 GPU, 12 GB Memory and Intel(R) Xeon(R) 2.3 GHz CPU is used for the benchmarking.
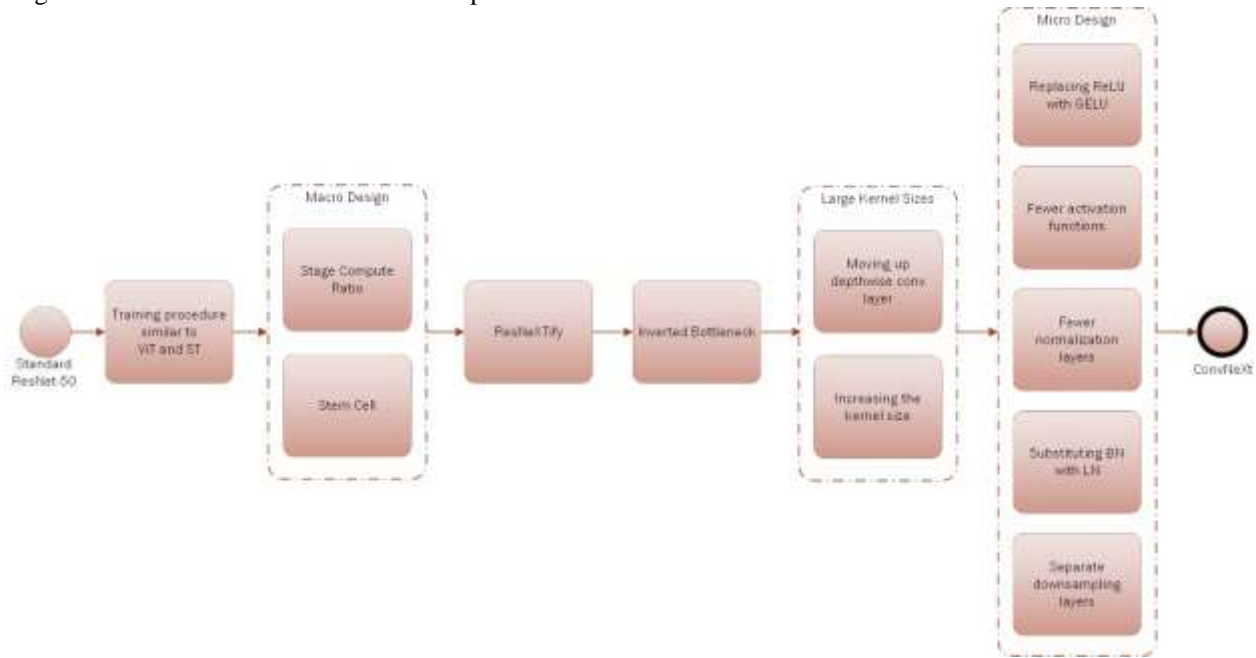

Fig.3. The Modernization Steps of ResNet-50 towards ST to finally get the ConvNeXt model [20]
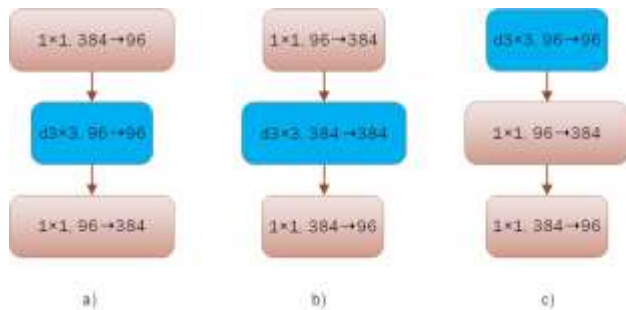

Fig 4. The Bottleneck Inversion and depthwise convolution borrowed from ResNeXt [20], [23] a) ResNeXt block   b) Inverted bottleneck block c) Moved up depthwise spatial convolution block

Fig.5. Block and layer architectures of ResNet, ST and ConvNeXt models [19], [20], [23]

IV. RESULT

In this study, three different benchmarking scenarios such as 'full data',' gradually increasing difficulty based data' and 'k-shot data' are proposed and the results of ConvNeXt method are shared in Table 1-3. The 'full data' scenario is based on using the whole training and testing data as originally provided by MHIST dataset. The 'gradually increasing difficulty based data' scenario is set as constituting new data subsets from the original training data by firstly including only VE level data and then adding E, H and VH levels respectively. The 'k-shot data' scenario is proposed as the subsets of the original dataset having only k members for each image class. The results of Baseline CNN models using 'full data' such as AlexNet, DenseNet, SqueezeNet, Inception v3 and VGG16 are given in Table 4 and Fig. 13. The confusion matrices of k-shots data results and ROC AUC curves of gradually increasing difficulty based data are given in Fig. 10 and 11. The different variants of ConvNeXt is applied on the 'full data' and its results are given in Fig. 12 and Table 5.

TABLE I
ConvNeXt BENCHMARKING ON 'FULL DATA'

Max ACC	Max F1	Max Cohen Kappa	Max ROC-AUC
0,8890	0,9121	0,7633	0,9391

TABLE II
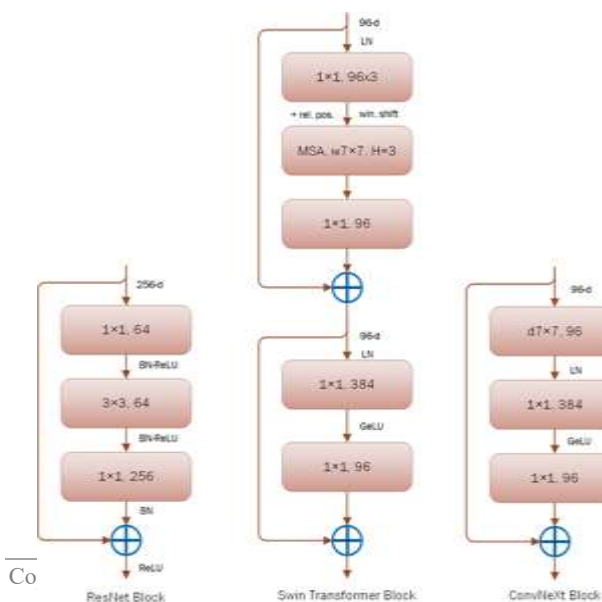
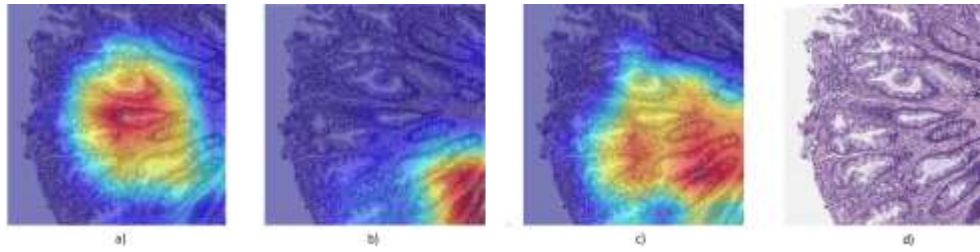| ConvNeXt BENCHMARKING ON 'K-SHOT DATA'								
K-SHOT	Max ACC	Max F1	Max Cohen Kappa	Max ROC-AUC				
					400	0,7861	0,8088	0.5750	0,8941
40	0,7052	0,7377	0.4128	0,8101				
100	0,7554	0,7805	0.5151	0,8617				

Fig. 6 Grad-Cam Visualizations of positive predictions of ConvNeXt model on an image selected as HP by 7 annotators. a)   40-shots   b) 100-shots   c) 400-shots   d) Original Image
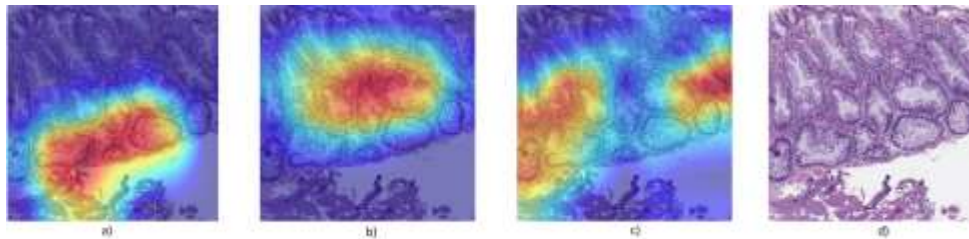

Fig. 7 Grad-Cam Visualizations of negative predictions of ConvNeXt model on an image selected as SSA by 6 annotators. a) 40-shots   b) 100-shots   c) 400-shots   d) Original Image
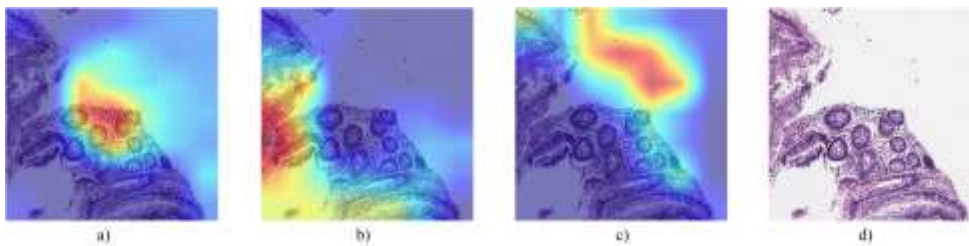

Fig. 8 Grad-Cam Visualizations of positive predictions of ConvNeXt model on an image selected as HP by 5 annotators. a)   40-shots   b) 100-shots   c) 400-shots   d) Original Image
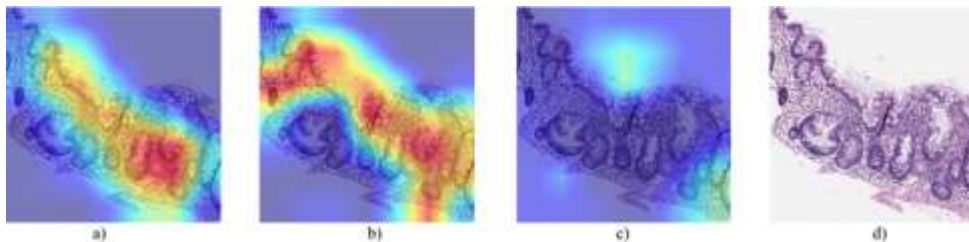

Fig. 9 Grad-Cam Visualizations of positive predictions of ConvNeXt model on an image selected as SSA by 7 annotators. a) 40-shots   b) 100-shots   c) 400-shots   d) Original Image
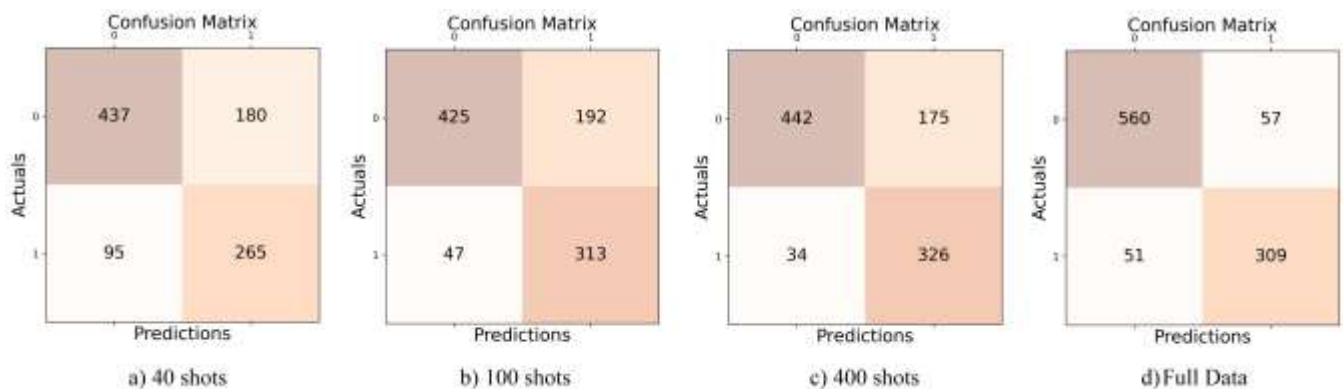

Fig. 10. The Confusion Matrices of ConvNeXt on k-shots data and 'full data' scenarios. a)   40 shots b) 100 shots c) 400 shots d) Full Data
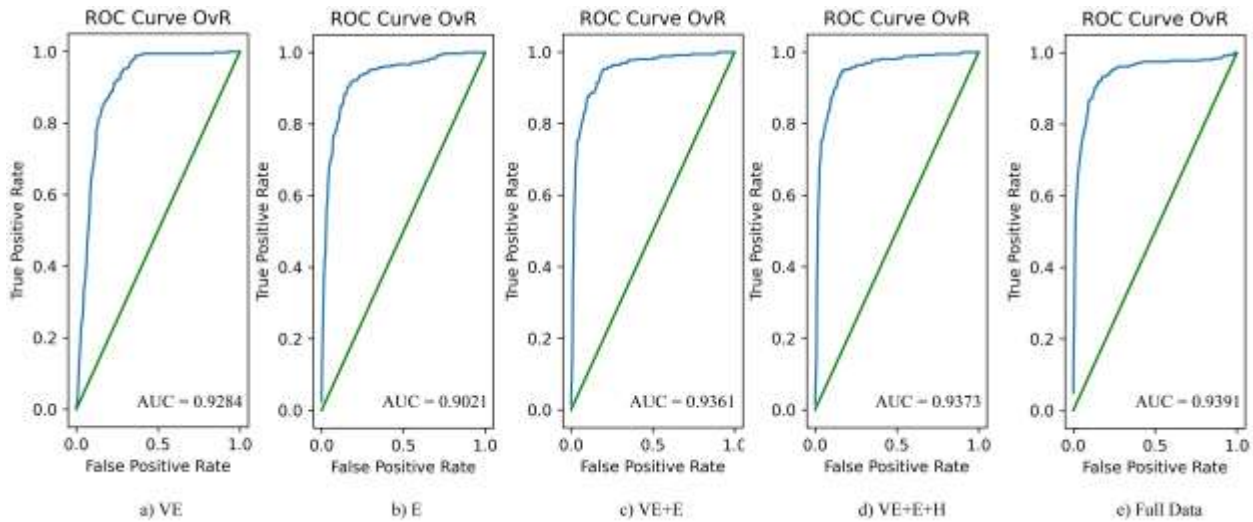
Fig. 11. The ROC graphs of ConvNeXt on difficulty based data and 'full data' scenarios. a) VE b) E c) VE+E d) Full Data
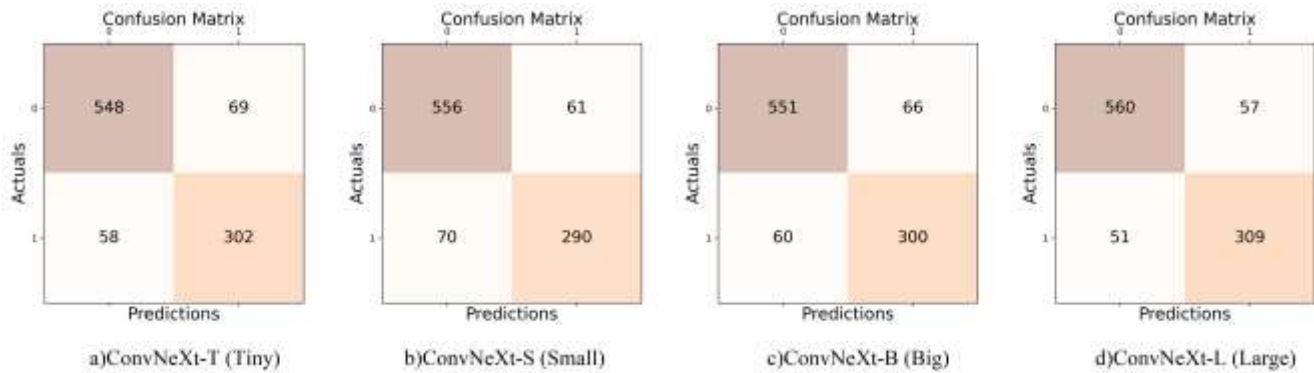

Fig. 12. The Confusion Matrices of different ConvNeXt variants on 'full data' scenario. a) Tiny b) Small c) Big d) Large

TABLE III
ConvNeXt BENCHMARKING ON 'GRADUALLY INCREASING
DIFFICULTY BASED DATA'

DIFF	Max ACC	Max F1	Max Cohen Kappa	Max ROC-AUC
E	0,8690	0,8947	0,7214	0,9284
VE	0,8209	0,8450	0,6372	0,9021
VE+E	0,8843	**0,9067**	0,7547	0,9361
VE+E+H	0,8865	0,9072	0,7588	0,9373
VE+E+H+VH	0,8890	0,9121	0,7633	0,9391

TABLE IV
BASELINE CNNS BENCHMARKING ON 'FULL DATA'

Model	Max ACC	Max F1	Max Cohen Kappa	Max ROC-AUC
AlexNet	0,7292	0,6461	0,3356	0,8243
DenseNet	0,7211	0,6275	0,3081	0,8439
SqueezeNet	0,7005	0,6184	0,2736	0,8049
Inception v3	0,7131	0,6488	0,3346	0,8538
VGG16	0,7041	0,6011	0,2635	0,8268

TABLE V
ConvNeXt BENCHMARKING ON 'FULL DATA' FOR DIFFERENT
MODELS

Model	Max ACC	Max F1	Max Cohen Kappa	Max ROC-AUC
Tiny	0,8700	0,8962	0,7225	0,9326
Small	0,8659	0,8946	0,7104	0,9298
Big	0,8710	0,8930	0,7239	0,9305
Large	0,8890	0,9121	0,7633	0,9391

V. DISCUSSION

The comparative table of the state of the art studies applied on MHIST database in literature is summarized on Table 6. The balanced values of ACC, AUC and F1-score values and high Cohen's Kappa of this study is promising with respect to all related studies. The Kappa metric analysis on the Fig. 7 of Wei et al. states that the average Cohen's Kappa among the annotations of 7 seven pathologists is 0.45 in the moderate range of 0.41-0.60 [4]. Nearly all the Cohen's Kappa results of all scenarios and models of this study which are calculated using majority voted ground truth labels and predicted labels as input to the cohen_kappa_score() function of sklearn library are over 0.60. Additionally, in this study, 40,100 and 400 shot data settings are curated from all difficulty levels by selecting equally and randomly are shown in Table 2. These results show that the generalization capability of ConvNeXt method on few shot subsets is promising. The 400-shots of not pretrained model of Wei et al. achieved 0.793 whereas the 40-shots of this study achieved 0.8101 for AUC [4]. The 100-shots of pretrained model of Wei et al. achieved 0.837 whereas the 40-shots of this study achieved 0.8617 for AUC [4].
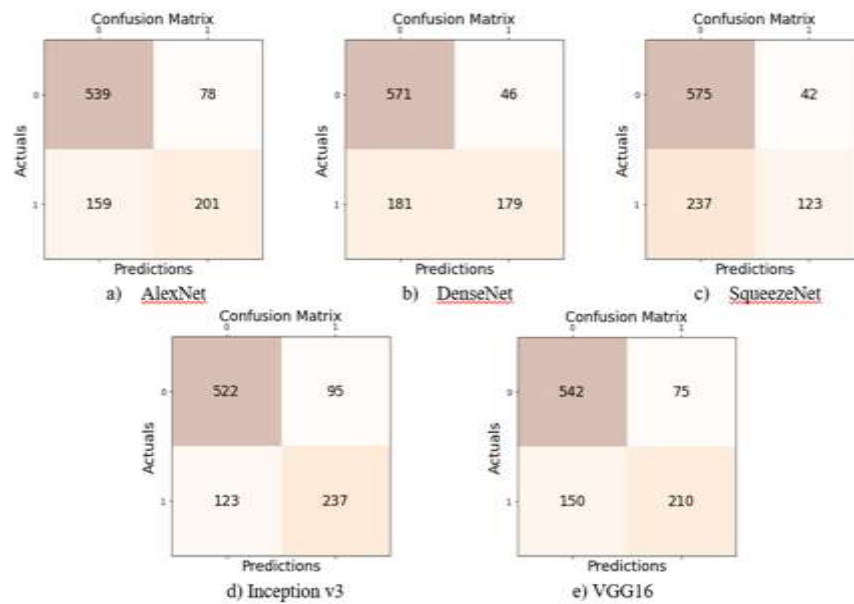
Fig.13. The Confusion Matrices of baseline CNN models on 'full data' scenario

TABLE VI
COMPARISON OF THIS STUDY AND THE STATE OF ART METHODS
APPLIED ON MHIST

Study		ACC	AUC	F1	Cohen Kappa
Wei et. al [5]		-	0,936	-	0,473
Wei et. al [4]		-	0,927	-	-
Srinidhi and Martel [12]		0,815	**0,887**	-	-
Wei et. al [10]		-	0,862	-	-
Wang et. al [13]	CNN	0,8095	0,9188	0,7957	-
	CNN+TRANS	0,8302	0,9378	0,8277	-
	CNN+TRANS+TAE	0,8651	**0,9476**	0,8586	-
Zhang [15]		0,8536	-	-	-
ConvNeXt (This Study)		**0,8890**	**0,9391**	**0,9121**	**0,7633**

This study exceeds the AUC results of the Srinidhi and Martel as can be seen on Table 6. The ConvNeXt model on 'full data' scenario of this study outperforms the AUC of all the ablated model settings of Wang et al. and reached very close to the performance of their best model setting which is 'CNN(ResNet50)+Transformer+TAE module pretrained as SSL' as can be observed in Table 6 (0.9391 versus 0,9188, 0.9378 and 0.9476) [13]. This study outperforms also the ACC and AUC of both pure transformers like ViT and T2T-ViT-24 as well as hybrid CNN-transformer paradigms such as VT-ResNet and BoTNet-50 [13]. On the other hand, it is worth to notice that Wang et al. used millions of histopathology images and trained 100 epochs on 32 Nvidia V100 GPUs during 200 hours. This sort of GPU capacity is very hard to reach for small research groups. On the other hand, this study based on ConvNeXt model can work with very few images and does not have high time complexity.

The superior ACC performance of this study is reasonable and more reliable since the SSA/HP class imbalance ratio of the MHIST dataset is 58%. The performance results and confusion matrices of Baseline CNN methods such as AlexNet, DenseNet, SqueezeNet, Inception v3 and VGG16 models benchmarked on 'full data' scenario are shown in Table 4 and Fig. 13. As it can be clearly understood from the

Table 1 and 4, the AUC value of the ConvNeXt model on 'full data' scenario are 8.53 higher than of the Inception v3 which is reported as the best baseline CNN. The ACC of the ConvNeXt model on 'full data' scenario is 15.98 higher than of the AlexNet as the baseline model having highest accuracy.

The last but not least, the Grad-Cam method is used for visually observing the behavior of the ConvNeXT model in Fig. 6,7,8 and 9 [24]. On the Grad-Cam images, it is observed that the effective regions of the images are more specialized as the number of shots are increased. Additionally, the effective regions of the images are observed as mostly observed as overlapping which supports the consistency of the ConvNeXT model across different shots of images.

The comparison of the change of training losses of the 40, 100 and 400 shots with respect to the full data is shown in Fig. 14. It is observed the loss of the smaller amount of shots starts from higher values as it is expected but they tend to reach to the plateau as the the curve of full data as epoch 40. On the other hand, the comparison of the change of training losses of the VE,E, VE+E and VE+E+H shots with respect to the full data is shown in Fig. 15. It is observed the loss of the easier subsets of the dataset starts from lower values as it is expected but they tend to reach to the plateau as the curve of full data as epoch 40. More comprehensive datasets may give further insights about the ConvNext method.

VI. CONCLUSION

In this study, the ConvNeXt method is applied on MHIST which is a histopathological imaging dataset having low data regime for the classification of precancerous colorectal lesions. ConvNeXT outperformes the other state of the art methods, which are applied on MHIST with respect to various metrics of different benchmarked training scenarios such as 'full data', 'gradually increasing difficulty based data' and 'k-shot data'. It is observed that the ConvNeXT method is capable of generalization even from a low data regime MHIST histopathological dataset using only limited number

of images. The ConvNeXt method also shows superior performance on the full data with respect to CNNs and ViTs with the score of 0.889, 0.9391, 0.9121 and 0.7633 for accuracy, AUC, F1 and cohen's cappa. The preferred ConvNeXT-L model is also found as superior with respect to the other variants of ConvNeXt such as Tiny, Small and Big ones in terms of all performance metrics. The ConvNeXt method is found as giving promising results on the computational pathology field and may be used in our future works for different datasets or deep learning based tasks as an alternative to CNNs and ViTs.
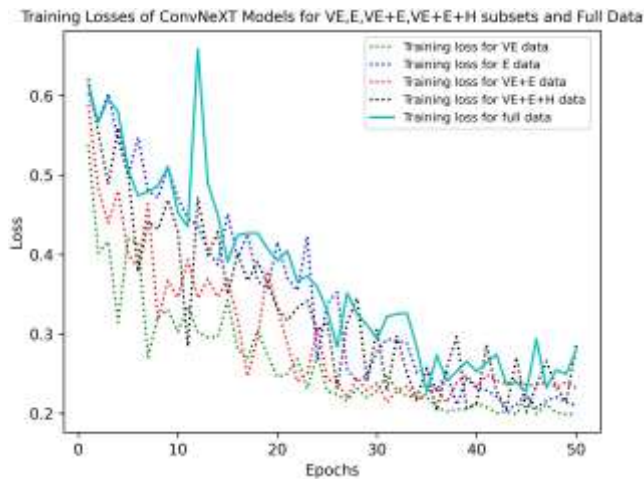

Fig.14. The comparison of the training losses of ConvNeXT models for 40, 100, 400 shots and full data
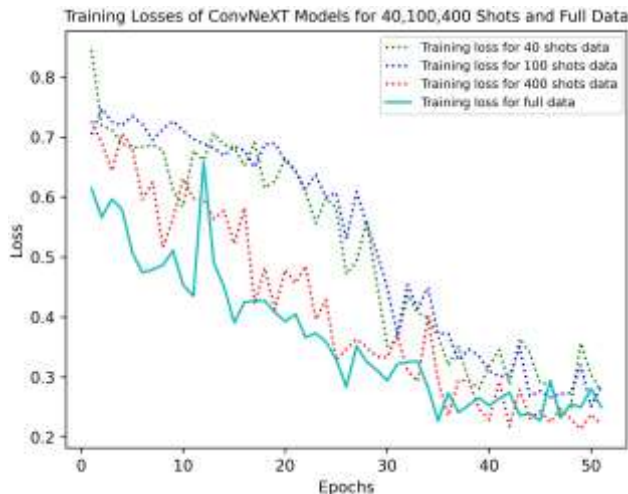

Fig.15. The comparison of the training losses of ConvNeXT models for VE, E, VE+E, VE+E+H subsets and full data

ACKNOWLEDGMENT

REFERENCES

[1] S. Famitha and M. Moorthi, "Intelligent and novel multi-type cancer prediction model using optimized ensemble learning," *Comput. Methods Biomech. Biomed. Engin.*, 2022, doi: 10.1080/10255842.2022.2081504.

[2] D. M. Metter, T. J. Colgan, S. T. Leung, C. F. Timmons, and J. Y. Park, "Trends in the us and canadian pathologistworkforces from 2007 to 2017," *JAMA Netw. Open*, vol. 2, no. 5, pp. 1–11, 2019, doi: 10.1001/jamanetworkopen.2019.4337.

[3] I. Mármol, C. Sánchez-de-Diego, A. P. Dieste, E. Cerrada, and M. J. R. Yoldi, "Colorectal carcinoma: A general overview and future perspectives in colorectal cancer," *Int. J. Mol. Sci.*, vol. 18, no. 1, 2017, doi: 10.3390/ijms18010197.

[4] J. Wei *et al.*, "A Petri Dish for Histopathology Image Analysis," Jan. 2021, [Online]. Available: http://arxiv.org/abs/2101.12355

[5] J. Wei *et al.*, "Learn like a Pathologist: Curriculum Learning by Annotator Agreement for Histopathology Image Classification," Sep. 2020, [Online]. Available: http://arxiv.org/abs/2009.13698

[6] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, "Generalizing from a Few Examples," *ACM Comput. Surv.*, vol. 53, no. 3, pp. 1–34, 2021, doi: 10.1145/3386252.

[7] V. Dumoulin *et al.*, "Comparing Transfer and Meta Learning Approaches on a Unified Few-Shot Classification Benchmark," Apr. 2021, [Online]. Available: http://arxiv.org/abs/2104.02638

[8] X. X. Yin, S. Hadjiloucas, Y. Zhang, and Z. Tian, "MRI radiogenomics for intelligent diagnosis of breast tumors and accurate prediction of neoadjuvant chemotherapy responses-a review," *Comput. Methods Programs Biomed.*, vol. 214, p. 106510, 2022, doi: 10.1016/j.cmpb.2021.106510.

[9] D. Pandey, X. Yin, H. Wang, and Y. Zhang, "Accurate vessel segmentation using maximum entropy incorporating line detection and phase-preserving denoising," *Comput. Vis. Image Underst.*, vol. 155, pp. 162–172, 2017, doi: 10.1016/j.cviu.2016.12.005.

[10] J. Wei, L. Torresani, J. Wei, and S. Hassanpour, "Calibrating Histopathology Image Classifiers using Label Smoothing," Jan. 2022, [Online]. Available: http://arxiv.org/abs/2201.11866

[11] Y. Bengio, umontrealca Jérôme Louradour, R. Collobert, and J. Weston, "Curriculum Learning."

[12] C. L. Srinidhi and A. L. Martel, "Improving Self-supervised Learning with Hardness-aware Dynamic Curriculum Learning: An Application to Digital Pathology." [Online]. Available: https://github.com/srinidhiPY/

[13] X. Wang *et al.*, "TransPath: Transformer-Based Self-supervised Learning for Histopathological Image Classification," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12908 LNCS, pp. 186–195, 2021, doi: 10.1007/978-3-030-87237-3_18.

[14] S. B. Yengec-tasdemir, "Classi cation of Colorectal Polyps from Histopathological Images using Ensemble of ConvNeXt Variants Classification of Colorectal Polyps from Histopathological Images using Ensemble of ConvNeXt Variants," pp. 0–26, 2022.

[15] R. Zhang *et al.*, "HistoKT: Cross Knowledge Transfer in Computational Pathology," Jan. 2022, [Online]. Available: http://arxiv.org/abs/2201.11246

[16] T. Who and T. Who, "The 2019 WHO classification of tumours of the digestive system," pp. 182–188, 2020, doi: 10.1111/his.13975.

[17] N. A. C. S. Wong, L. P. Hunt, M. R. Novelli, N. A. Shepherd, and B. F. Warren, "Observer agreement in the diagnosis of serrated polyps of the large bowel," *Histopathology*, vol. 55, no. 1. pp. 63–66, 2009. doi: 10.1111/j.1365-2559.2009.03329.x.

[18] M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio, "Transfusion: Understanding Transfer Learning for Medical Imaging," Feb. 2019, [Online]. Available: http://arxiv.org/abs/1902.07208

[19] Z. Liu *et al.*, "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows," pp. 9992–10002, 2022, doi: 10.1109/iccv48922.2021.00986.

[20] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," 2022, [Online]. Available: http://arxiv.org/abs/2201.03545

[21] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and H. Jégou, "Training data-efficient image transformers & distillation through attention," 2020, [Online]. Available:

http://arxiv.org/abs/2012.12877

[22]   H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "MixUp: Beyond empirical risk minimization," *6th Int. Conf. Learn. Represent. ICLR 2018 - Conf. Track Proc.*, pp. 1–13, 2018.

[23]   S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 5987–5995, 2017, doi: 10.1109/CVPR.2017.634.

[24]   R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 336–359, 2020, doi: 10.1007/s11263-019-01228-7.

BIOGRAPHIES

MEHMET NERGİZ He graduated from the Computer Engineering Department of Boğaziçi University in 2008. He received his M.S and Ph.D. degrees in 2013 and 2017, respectively from the Department of Electrical and Electronics Engineering at Dicle University. He worked as a Postdoctoral Research Fellow on medical image analysis and artificial intelligence at Harvard Medical School - Massachusetts General Hospital for one year between 2019 - 2020. His research field is machine learning, deep learning, federated learning, computer vision and signal processing. He currently works as an Assistant Professor and also the Head of Computer Engineering Department, the Director of Remote Distance Center and the Advisor of the Rector at the Dicle University.