Pan Organ Cancer Detection

1st Prof. Amit Sethi
Dept. of Electrical Engineering,
Indian Institute of Technology, Bombay

2nd Prof. Ayan Bhattacharya
Dept. of Mathematics
Indian Institute of Technology, Bombay

3rd Prajyot Kore
Dept. of Mathematics
Indian Institute of Technology, Bombay

4th Shruti Sawant
Dept. of Electrical Engineering,
Indian Institute of Technology, Bombay

***Abstract*—Whole Slide Imaging (WSI) has revolutionized pathology by digitizing entire slides at high resolution, offering remote accessibility, aiding computer-aided analysis, and enhancing archiving. In this paper, we present a patch-wise approach for tumor detection in Multiorgan WSI images sourced from The Cancer Genome Atlas (TCGA) repository. Leveraging pre-trained models such as MobileNet-V3-Small and KimiaNet, we fine-tune and train with histopathology images to achieve robust performance. Our methodology includes extensive preprocessing, data augmentation, and hyperparameter tuning to mitigate overfitting and enhance model generalization. Through experiments on diverse organ datasets, including breast, colon, prostate, and thyroid, we demonstrate the effectiveness of our approach. KimiaNet, utilizing a unique "high-cellularity mosaic" methodology, and DenseNet architecture outperforms traditional architectures, showcasing superior performance in tumor/non-tumor classification tasks. Our study underscores the importance of iterative model development and the efficacy of leveraging diverse datasets for enhanced performance in digital pathology.**

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

WSI (Whole Slide Imaging) revolutionizes pathology by digitizing entire slides at high resolution, replacing traditional microscopy. Offering remote accessibility, aiding computer-aided analysis, and enhancing archiving, WSI is pivotal for diagnosis, education, and research. This paper presents an approach for tumor detection in Multiorgan WSI images **sourced from formalin-fixed paraffin-embedded human pathology samples available via The Cancer Genome Atlas (TCGA) repository.** To do this we utilized pretrained model MobileNet-V3-Small, and KimiaNet that employs the topology of the DenseNet with four dense blocks, fine-tuned and trained with histopathology images in different configurations. We used kimiaNet a pretrained model for histopathology images they used more than 240,000 image patches with 1000 1000 pixels acquired at 20 magnification through our proposed "high cellularity mosaic" approach to enable the usage of weak labels of 7,126 whole slide images of formalin-fixed paraffin-embedded human pathology samples publicly available through the The Cancer Genome Atlas (TCGA) repository Many studies, however, have shown that deep features, i.e., high-level embeddings in a properly trained deep network, can outperform handcrafted features in most applications [1]. As a result, many different convolutional architectures have been trained and introduced to provide features either directly or through transfer learning [2] [3]. This has created new questions in computer vision research, and consequently in medical image analysis as to which network topology is most suitable for a given task. Is transfer learning sufficient to solve specific problems? What challenges and benefits does training an entire network from scratch entail?

[4] [5] [6]

II. Background

The literature on the applications of deep learning in digital pathology is diverse and contains a multitude of approaches [7] [4] [8] panella et al., 2019). As we are focused on image repre sentation, in this section we only review recent works that have used, fine-tuned or trained deep networks for different purposes in digital pathology [7] [9]

Pre-Trained Networks– Two early examples of off-the shelf feature extractors are Overfeat [10] and DeCaf [11] used successfully in breast cancer classification by capturing and combining different fully connected (FC) layers [12]. Combining Inception (V3) features (extracted from multi magnification pathology images) with a fully connected layer has been used for binary classification in breast cancer metas tasis analysis [13] (Liu et al., 2017). Feature selection from pre-trained deep networks has also been performed in many pathology tasks like HEp-2 protein classification [14]. Using features from pre trained networks on pathology domains is another way to classify the pathology problems [3]. For instance, features from DeepLoc [15] have been used to classify protein subcellular local ization [16]. As well, deep features from pre trained networks have been successfully used for image search [17] [18] KimiaNet, a novel system developed based on the topology of DenseNet, introduces a unique methodology called cellMosaic. This approach utilizes weak or soft labels on extensive datasets like TCGA, which lack pixel-level annotations. Leveraging cellMosaic has notably enhanced KimiaNet's performance, surpassing that of DenseNet for diagnostic imaging tasks involving TCGA datasets.

Fine-Tuned Networks– It is generally expected that fine tuning should increase the accuracy of a pre-trained network or the expressiveness of its features in pathology image clas sification [19]. [20]. fine-tuned the last 2 convolutional blocks

of a VGG-19 model initialized with ImageNet weights. More specifically, they utilized an notated patches of size 1024 by 1024 at 20 magnification from a neuropathology dataset with 13 distinct classes for transfer learning. Eventually, they reported a 95% accuracy for the classification task. [21] also fine-tuned a VGG-19 model utilizing an extended version of the latter dataset with 74 distinct classes annotated employing the World Health Organization (WHO) tomour classification scheme. Ultimately, they reported the training validation accuracy of 66% over the images spanning the 74 trained classes. Trained Networks– [22] Coudray et al. (2018) trained an Inception V3 model on 1176 lung WSIs from the TCGA dataset. They extracted patches of size 512x512 pixels from Lung Adenocarcinoma (LUAD), Lung Squamous Cell Carci noma (LUSC), and healthy lung tissue slides. Subsequently, they predicted diagnosis associated with any given WSI by assessing the majority vote among the patch level predictions of all tiles of the same slide. Finally, they reported around 97% accuracy for both 20 and 5 magnifications

III. Dataset and Preprocessing

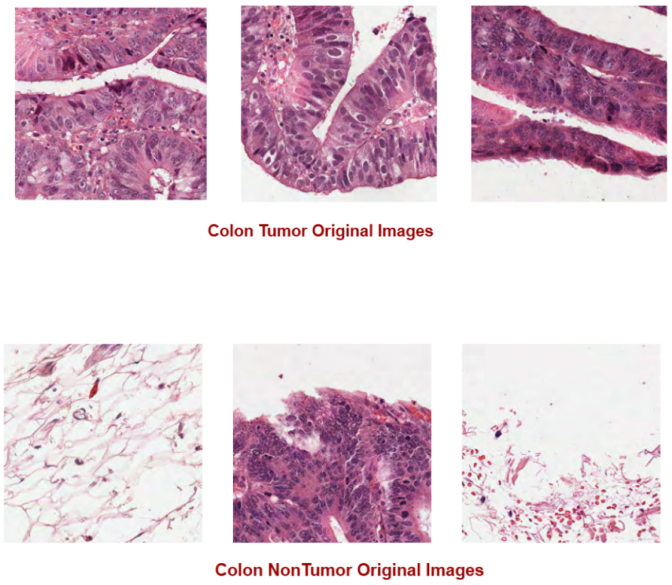


Fig. 1. Samples from Colon

The dataset initially lacked suitability for training, as WSIs were labeled, not the patches. To address this: we labelled data patchwise, we initially had, seven organs names Breast, Thyroid, Prostate, Colon, Ovary and Lung. samples from colon organ is as shown in Fig.1, We had 4 whole slides for breast, 4 whole slides for colon, 4 whole slide for prostate, 4 whole slides for Thyroid and we considered 3 slides for training and 1 slide for testing.

Preprocessing-

In our study, we employed preprocessing techniques aimed at mitigating overfitting and enhancing the generalization performance of our models. We utilized random sampling to enrich the diversity and resilience of our training dataset.

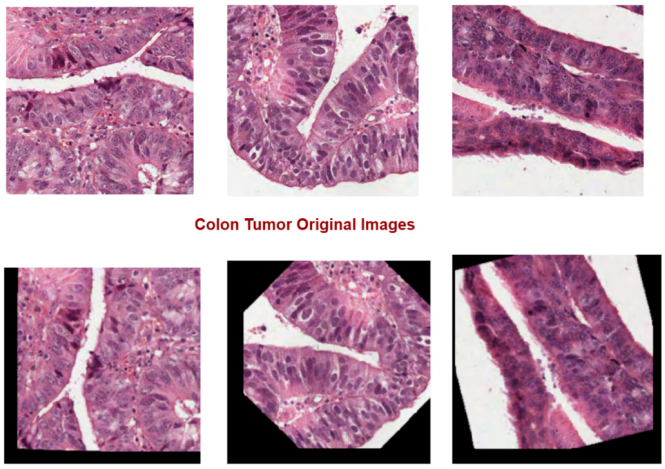


Fig. 2. Data augmentation for Colon sample

Through random selection from each slides, augmentation techniques like rotation, scaling, and flipping, this approach introduces variability into the dataset.Also, we did color normalization. By doing so, we aim to mitigate overfitting by broadening the model's exposure to diverse data instances. This strategy becomes particularly crucial when working with a limited number of slides from which patches are extracted for training, as it helps alleviate the model's dependence on specific patterns inherent in the original dataset.

For our preprocessing pipeline, we implemented color normalization specifically tailored for stained whole slide images (WSI). This technique aims to standardize the color distributions across the WSI dataset, ensuring consistency in color representations despite potential variations in staining protocols or imaging conditions.

TABLE I
Summary of Dataset

Organ	Train Tumor patches	Train Non-Tumor patches	Test Tumor patches	Test Non-Tumor patches
Colon	588	2149	53	156
Thyroid	449	7819	75	1245
Prostate	341	2635	47	144
Breast	520	7291	225	1547

Focused on training at the patch level, as the patch dataset contained both healthy and diseased patches from the same WSI, potentially causing confusion during deep network training.

IV. Methodology

Our method was to try out different architecture to validate results of varieties of architecture on different organs as our main aim is to develop a model which will able to and different dataloader.

A. Architecture

1) MobileNet-V3-Small:: MobileNet v3 Small is a lightweight convolutional neural network architecture optimized for mobile devices, offering high efficiency and accuracy in image classification tasks through depthwise separable

convolutions and efficient inverted residuals. we considered MobileNet-v3-Small model with custom Fully Connected layer.

2) KimiaNet: [23] proposed a new network, namely KimiaNet, that employs the topology of the DenseNet with four dense blocks, fine-tuned and trained with histopathology images in different configurations. they used more than 240,000 image patches with pixels acquired at 20 magnification through our proposed "high-cellularity mosaic" approach to enable the usage of weak labels of 7126 whole slide images of formalin-fixed paraffin-embedded human pathology samples publicly available through The Cancer Genome Atlas (TCGA) repository. We tested KimiaNet using three public datasets, namely TCGA, endometrial cancer images, and colorectal cancer images by evaluating the performance of search and classification when corresponding features of different networks are used for image representation. We used KimiaNet for tumor/ nontunmor classification with customized fully connected layer.

B. Balanced Dataloader:

The initial dataloader's limitation in only generating samples from certain slides restricted the model's exposure to a diverse range of data.(Fig. 3) As a result, the model may have struggled to generalize well to unseen samples or slides, leading to potential performance degradation, especially when faced with variations present in real-world data By implementing the new dataloader that randomly selects samples from each slide (Fig .4), we introduced greater diversity into the training dataset. The introduction of random sampling from all slides has significantly contributed to the overall improvement in model performance.

Fig. 3. Initial Dataloader

```
Image Name: Colon_slide3_nontumor_patch100_cn.png, Class: NonTumor
Image Name: Colon_slide3_nontumor_patch101_cn.png, Class: NonTumor
Image Name: Colon_slide3_nontumor_patch102_cn.png, Class: NonTumor
Image Name: Colon_slide3_nontumor_patch103_cn.png, Class: NonTumor
Image Name: Colon_slide3_nontumor_patch104_cn.png, Class: NonTumor
Image Name: Colon_slide3_nontumor_patch105_cn.png, Class: NonTumor
Image Name: Colon_slide3_nontumor_patch106_cn.png, Class: NonTumor
Image Name: Colon_slide3_nontumor_patch107_cn.png, Class: NonTumor
Image Name: Colon_slide3_nontumor_patch108_cn.png, Class: NonTumor
Image Name: Colon_slide3_nontumor_patch109_cn.png, Class: NonTumor
Image Name: Colon_slide3_nontumor_patch10_cn.png, Class: NonTumor
Image Name: Colon_slide3_nontumor_patch110_cn.png, Class: NonTumor
Image Name: Colon_slide3_nontumor_patch111_cn.png, Class: NonTumor
Image Name: Colon_slide3_nontumor_patch112_cn.png, Class: NonTumor
Image Name: Colon_slide3_nontumor_patch113_cn.png, Class: NonTumor
Image Name: Colon_slide3_nontumor_patch114_cn.png, Class: NonTumor
Image Name: Colon_slide3_nontumor_patch115_cn.png, Class: NonTumor
Image Name: Colon_slide3_nontumor_patch116_cn.png, Class: NonTumor
Image Name: Colon_slide3_nontumor_patch117_cn.png, Class: NonTumor
Image Name: Colon_slide3_nontumor_patch118_cn.png, Class: NonTumor
Image Name: Colon_slide3_nontumor_patch119_cn.png, Class: NonTumor
Image Name: Colon_slide3_nontumor_patch11_cn.png, Class: NonTumor
Image Name: Colon_slide3_nontumor_patch120_cn.png, Class: NonTumor
Image Name: Colon_slide3_nontumor_patch121_cn.png, Class: NonTumor
Image Name: Colon_slide3_nontumor_patch122_cn.png, Class: NonTumor
Image Name: Colon_slide3_nontumor_patch123_cn.png, Class: NonTumor
```

C. Mitigating Class Imbalance

Acknowledging our dataset's class imbalance, with a surplus of non-tumor patches compared to tumor, we employed **weighted cross-entropy loss** to address this disparity effectively. Additionally, in our evaluation metrics, we prioritized **Receiver Operating Characteristic Area Under Curve (ROC-AUC)** to ensure robust performance evaluation and mitigate the impact of class imbalance on our model's effectiveness.

Fig. 4. Balanced Dataloader

```
Breast_slide0_tumor_patch136_cn.png
Breast_slide1_nontumor_patch647_cn.png
Breast_slide1_tumor_patch144_cn.png
Breast_slide2_nontumor_patch691_cn.png
Breast_slide1_tumor_patch172_cn.png
Breast_slide0_nontumor_patch990_cn.png
Breast_slide0_tumor_patch18_cn.png
Breast_slide2_nontumor_patch1039_cn.png
Breast_slide0_tumor_patch102_cn.png
Breast_slide1_nontumor_patch1285_cn.png
Breast_slide2_tumor_patch83_cn.png
Breast_slide2_nontumor_patch2170_cn.png
Breast_slide1_tumor_patch193_cn.png
Breast_slide2_nontumor_patch2458_cn.png
Breast_slide1_tumor_patch77_cn.png
Breast_slide2_nontumor_patch133_cn.png
Breast_slide1_tumor_patch110_cn.png
Breast_slide1_nontumor_patch98_cn.png
Breast_slide1_tumor_patch41_cn.png
Breast_slide1_nontumor_patch2177_cn.png
Breast_slide0_tumor_patch133_cn.png
Breast_slide2_nontumor_patch1329_cn.png
Breast_slide2_tumor_patch69_cn.png
Breast_slide2_nontumor_patch1474_cn.png
Breast_slide1_tumor_patch43_cn.png
```

D. Hyperparameter Tunning

we conducted an exhaustive grid search over hyperparameters. We explored learning rates ranging from 1e-7 to 1e-2, considered different combinations of trainable layers from the top 0 to 4 layers, and varied dropout rates from 0 to 0.5. This comprehensive approach allowed us to identify the most effective model configuration for our task.

E. Mitigating Over-fitting

we implemented a multifaceted approach. We experimented with various models, adjusting dropout rates to introduce regularization and mitigate overfitting. Additionally, we incorporated early stopping mechanisms and applied L2 regularization to constrain model complexity, collectively enhancing generalization performance and addressing overfitting concerns.

V. Experiments and Results

Due to time and resources restriction we performed hyperparameter tunning on few organs only,

A. Results for MobileNet and KimiaNet without Balanced Dataloader-

Table II presents the results obtained from MobileNet and KimiaNet models trained without balanced dataloader. It is evident from the table that these models exhibit limited generalization capabilities, indicating a challenge in effectively capturing the underlying patterns present in the dataset.

TABLE II
SUMMARY KIMIANET WITH UNBALANCED DATALOADER ON ALL ORGANS

Organ	Train Layer	Learning Rate	Dropout	Test Accuracy	AUC Score	epochs
Colon	3	1e-06	0	-	0.76	200
Breast	0	1e-06	0	-	0.57	200
Prostate	0	1e-06	0	-	0.59	200
Thyroid	4	1e-06	0.5	-	0.56	200

B. Results for MobileNet and KimiaNet with Balanced Dataloader-

One of the models we explored was KimiaNet, where we experimented with different configurations across multiple organs. For instance, in colon classification, we observed that training with 3 layers, a learning rate of 1e-06, and a dropout rate of 0 yielded promising results with an AUC score of 0.76. However, in breast classification, the performance was comparatively lower, with an AUC score of 0.57. (Table II)

Similarly, we evaluated the MobileNet-V3 architecture, focusing on colon classification. Among the configurations tested, training with 4 layers, a learning rate of 1e-07, and a dropout rate of 0.2 exhibited the most favorable performance, achieving an AUC score of 0.62.(Table II)

However, despite these efforts, the performance of the models fell short of our expectations, indicating the need for further optimization.

TABLE III
TOP 5 BEST MODELS FOR COLON FROM MOBILENET-V3-SMALL WITH BALANCED DATALOADER, WITH EPOCH 250 AND EARLY STOPPING

LR	Dropout	Layers	AUC	Accuracy
0.001	0.5	4	0.8954	84.68
0.001	0	4	0.8898	83.25
0.001	0.2	2	0.8839	82.77
1e-06	0.5	2	0.8562	82.77
0.001	0	0	0.8595	79.90

TABLE IV
TOP 5 BEST MODELS FOR BREAST FROM MOBILENET-V3-SMALL WITH BALANCED DATALOADER, WITH EPOCH 250 AND EARLY STOPPING

LR	Dropout	Layers	AUC	Accuracy
0.001	0.2	0	0.9818	89.84
0.001	0	4	0.9816	89.89
0.001	0.5	0	0.9811	89.72
1e-06	0.5	0	0.9786	89.61
1e-06	0.5	2	0.9757	88.37

In light of these findings, we refined our approach by implementing Balanced Dataloader techniques in our subsequent experiments. This strategic adjustment aimed to enhance model generalization and performance by introducing variability in the training data.

Our revised methodology yielded significant improvements in classification accuracy and AUC scores across both colon and breast classification tasks. For instance, in the latest experiments with the MobileNet-v3-Small model for colon classification, our top-performing model achieved an AUC of 0.8954 and an accuracy of 84.68%, representing a substantial enhancement over our previous results.

Similarly, our experiments with DenseNet models for colon classification showcased impressive performance, with the top-performing model achieving an AUC of 0.9869 and an accuracy of 94.25

Moreover, the introduction of Balanced Dataloader techniques played a pivotal role in augmenting model performance, underscoring the importance of data preprocessing strategies in medical image analysis.

TABLE V
5 PERFORMANCE METRICS OF MOBILENET-V3-SMALL-BREAST WITH BALANCED DATALOADER ON COLON

Model Specification			AUC		Accuracy	
LR	DR	L	Breast	Colon	Breast	Colon
0	1E-06	0	0.981	0.857	89.729	30.144
0	1E-06	2	0.977	0.923	89.391	28.708
0	1E-06	4	0.982	0.884	89.898	30.144
0.2	1E-06	0	0.982	0.924	89.842	30.144
0.2	1E-06	2	0.977	0.912	88.657	27.273

TABLE VI
5 PERFORMANCE METRICS OF MOBILENET-V3-SMALL-RANDOM-COLON ON BREAST

Model			AUC		Accuracy	
LR	DR	L	Colon	Breast	Colon	Breast
0	1E-06	0	0.78	0.92	71.29	89.05
0	1E-06	2	0.75	0.93	69.37	91.19
0	1E-06	4	0.84	0.91	80.38	88.71
0.2	1E-06	0	0.82	0.93	76.55	88.99
0.2	1E-06	2	0.82	0.93	70.33	90.85
0.2	1E-06	4	0.80	0.90	77.99	88.54

TABLE VII
TOP MODELS FOR COLON AND BREAST FROM KIMIANET-REGULARIZED-BALANCED WITH WEIGHT-DECAY = 1E-5

Organ	LR	Dropout	Layers	AUC	Accuracy
Colon	1e-06	0	2	0.98	93.77
Breast	1e-06	0.	0	0.99	96.84

Additionally, we evaluated KimiaNet models for both colon and breast classification tasks, achieving remarkable results. For colon classification, the top-performing KimiaNet model achieved an AUC of 0.9811 and an accuracy of 93.78 %. In breast classification, the KimiaNet model outperformed all others, attaining an exceptional AUC score of 0.99 and an accuracy of 96.84

Overall, our experiments underscore the iterative nature of model development and optimization in medical image classification tasks, highlighting the efficacy of meticulous experimentation and strategic adjustments in achieving superior results.

Table VIII and Table IX presents the performance metrics of MobileNet-v3-Small-Balanced models trained for Breast and tested on Colon and performance metrics of MobileNet-v3-Small-Balanced models trained for Colon tested on Breast.

From Table VIII and Table IX, it can be observed that the performance metrics of MobileNet-v3-Small-Balanced models trained for Breast and tested on Colon differ significantly from those trained for Colon and tested on Breast.

For instance, in Table VIII, where the models are trained for Breast and tested on Colon, the AUC values for Breast range from 0.75 to 0.84, while the AUC values for Colon range from 0.90 to 0.93. This indicates that the models perform relatively better on Colon compared to Breast, as reflected by the higher AUC values for Colon.

Conversely, in Table IX, where the models are trained for Colon and tested on Breast, the AUC values for Breast range from 0.78 to 0.82, while the AUC values for Colon range from 0.90 to 0.93. Here, the models exhibit higher performance on Colon, with AUC values consistently higher than those for Breast.

Additionally, considering the results from DenseNet presented in Table X, it can be observed that the AUC values for Breast range from 0.960 to 0.987, and for Colon, they range from 0.911 to 0.955. These results further emphasize the trend observed with MobileNet-v3-Small-Balanced models, with higher AUC values generally obtained for Colon compared to Breast.

These observations suggest that there may be variations in the predictive performance of both MobileNet-v3-Small-Balanced and DenseNet models depending on the organ they are trained for and tested on. Further analysis and investigation would be necessary to understand the underlying factors contributing to these differences and to optimize the model's performance across different scenarios.

C. Loss Curves for Models:

Fig. 5. Loss Curve for MobileNet-V3-Small with dropout 0.5, 4 trainable Layers with Balanced Dataloader

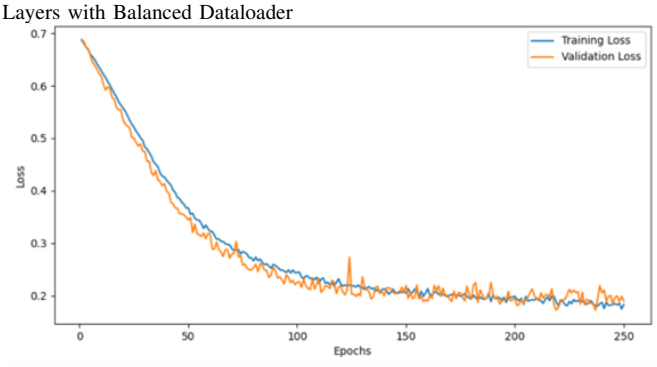
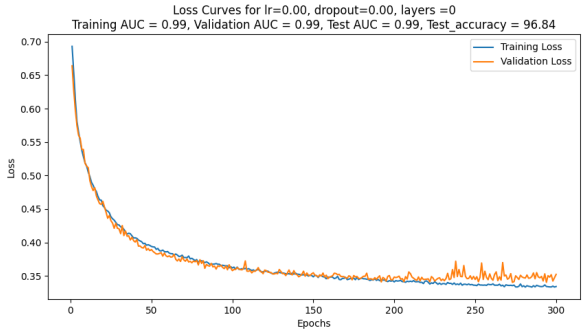


Fig. 6. Loss Curve for KimiaNet with dropout 0, 0 trainable Layers with Balanced Dataloader



VI. Summary and Conclusion

In the realm of digital pathology, the question of image representation is paramount, given the intricate texture complexity, polymorphism, and sheer size of Whole Slide Images (WSIs). Recent advancements underscore the significance of high-level embeddings in artificial neural networks for robust and expressive image representation. Pre-trained networks like DenseNet and MobileNet-v3 Small, leveraging their discrimination power honed through intensive training with millions of natural images, have gained widespread adoption in medical image analysis. However, the scarcity of labeled medical image data and the computational demands pose significant challenges for fine-tuning or training deep networks with histopathology images.

In our study, we introduce KimiaNet, a novel approach tailored to the demands of digital pathology. We propose several fine-tuned configurations of KimiaNet, alongside investigations into MobileNet-v3 Small and DenseNet architectures. KimiaNet utilizes a clustering-based mosaic structure for image representation, augmented by leveraging high cellularity to enable the utilization of Weakly Supervised Learning (WSL) with Whole Slide Image (WSI)-level labels, particularly in archives lacking pixel-level annotations. To validate our approach, we leverage three public datasets to demonstrate its efficacy.

Our primary contribution lies in leveraging a diverse, multi-organ public image repository, such as TCGA, at 20x magnification to extract large, high-resolution patches (1000x1000 pixels) for training a densely connected network with weak labels, serving as a powerful feature extractor. Additionally, our study highlights that fine-tuning a deep network on a sufficiently large number of histopathology images yields superior performance compared to using a pre-trained network model.

References

[1] M. Babaie and H. R. Tizhoosh, "Deep features for tissue-fold detection in histopathology images," in *Digital Pathology: 15th European Congress, ECDP 2019, Warwick, UK, April 10–13, 2019, Proceedings 15*. Springer, 2019, pp. 125–132.

[2] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.

[3] R. Mormont, P. Geurts, and R. Marée, "Comparison of deep transfer learning strategies for digital pathology," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018, pp. 2262–2271.

[4] M. K. K. Niazi, A. V. Parwani, and M. N. Gurcan, "Digital pathology and artificial intelligence," *The lancet oncology*, vol. 20, no. 5, pp. e253–e261, 2019.

[5] D. Yu, L. Deng, F. T. B. Seide, and G. Li, "Discriminative pretraining of deep neural networks," Jan. 12 2016, uS Patent 9,235,799.

[6] J. Burt, N. Torosdagli, N. Khosravan, H. RaviPrakash, A. Mortazi, F. Tissavirasingham, S. Hussein, and U. Bagci, "Deep learning beyond cats and dogs: recent advances in diagnosing breast cancer with deep neural networks," *The British Journal of Radiology*, vol. 91, p. 20170545, 2018.

[7] A. Janowczyk and A. Madabhushi, "Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases," *Journal of Pathology Informatics*, vol. 7, 2016.

[8] G. Campanella, M. Hanna, L. Geneslaw, A. Miraflor, V. Silva, K. Busam, E. Brogi, V. Reuter, D. Klimstra, and T. Fuchs, "Clinical-grade computational pathology using weakly supervised deep learning on whole slide images," *Nature Medicine*, vol. 25, pp. 1301–1309, 2019.

[9] M. Gurcan, L. Boucheron, A. Can, A. Madabhushi, N. Rajpoot, and B. Yener, "Histopathological image analysis: A review," *IEEE Reviews in Biomedical Engineering*, vol. 2, p. 147, 2009.

[10] A. Sharif Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "Cnn features off-the-shelf: an astounding baseline for recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2014, pp. 806–813.

[11] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell, "Decaf: A deep convolutional activation feature for generic visual recognition," in *International conference on machine learning*, 2014, pp. 647–655.

[12] F. A. Spanhol, L. S. Oliveira, P. R. Cavalin, C. Petitjean, and L. Heutte, "Deep features for breast cancer histopathological image classification," in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2017, pp. 1868–1873.

[13] G. Huang, Z. Liu, L. Van Der Maaten, and K. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[14] H. T. H. Phan, A. Kumar, J. Kim, and D. Feng, "Transfer learning of a convolutional neural network for hep-2 cell image classification," in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2016, pp. 1208–1211.

[15] J. Almagro Armenteros, C. Sønderby, S. Sønderby, H. Nielsen, and O. Winther, "Deeploc: prediction of protein subcellular localization using deep learning," *Bioinformatics*, vol. 33, pp. 3387–3395, 2017.

[16] O. Kraus, B. Grys, J. Ba, Y. Chong, B. Frey, C. Boone, and B. Andrews, "Automated analysis of high-content microscopy data with deep learning," *Molecular Systems Biology*, vol. 13, 2017.

[17] S. Kalra, C. Choi, S. Shah, L. Pantanowitz, and H. Tizhoosh, "Yottixel–an image search engine for large archives of histopathology whole slide images," *arXiv preprint arXiv:1911.08748*, 2019.

[18] S. Kalra, H. Tizhoosh, S. Shah, C. Choi, S. Damaskinos, A. Safarpoor, S. Shafiei, M. Babaie, P. Diamandis, and C. Campbell, "Pan-cancer diagnostic consensus through searching archival histopathology images using artificial intelligence," *arXiv preprint arXiv:1911.08736*, 2019.

[19] B. Kieffer, M. Babaie, S. Kalra, and H. Tizhoosh, "Convolutional neural networks for histopathology image classification: Training vs. using pre-trained networks," in *2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*. IEEE, 2017, pp. 1–6.

[20] K. Faust, Q. Xie, D. Han, K. Goyle, Z. Volynskaya, U. Djuric, and P. Diamandis, "Visualizing histopathologic deep learning classification and anomaly detection using nonlinear feature space dimensionality reduction," *BMC Bioinformatics*, vol. 19, p. 173, 2018.

[21] K. Faust, S. Bala, R. van Ommeren, A. Portante, R. Al Qawahmed, U. Djuric, and P. Diamandis, "Intelligent feature engineering and ontological mapping of brain tumour histomorphologies by deep learning," *Nature Machine Intelligence*, vol. 1, pp. 316–321, 2019.

[22] N. Coudray, P. Ocampo, T. Sakellaropoulos, N. Narula, M. Snuderl, D. Feny¨ o, A. Moreira, N. Razavian, and A. Tsirigos, "Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning," *Nature Medicine*, vol. 24, p. 1559, 2018.

[23] A. Riasatian, M. Babaie, D. Maleki, S. Kalra, M. Valipour, S. Hemati, M. Zaveri, A. Safarpoor, S. Shafiei, M. Afshari *et al.*, "Fine-tuning and training of densenet for histopathology image representation using tcga diagnostic slides," *Medical image analysis*, vol. 70, p. 102032, 2021.