TIPTEKNO'22 MEDICAL TECHNOLOGIES CONGRESS
31 October - 2 November 2022
Xanadu Resort Hotel, Belek, Antalya

Fast and Interpretable Deep Learning Pipeline for Breast Cancer Recognition

Mahdi Bonyani*, Faezeh Yeganli†, S. Faegheh Yeganli‡
*Department of Computer Engineering, University of Tabriz, Tabriz, Iran.
†Department of Electrical and Electronics Engineering, Izmir University of Economics, Izmir, Turkey.
‡Department of Computer Engineering, Yasar University, Izmir, Turkey.
*m_bonyani96@ms.tabrizu.ac.ir, †faezeh.yeganli@ieu.edu.tr, ‡faegheh.yeganli@yasar.edu.tr

***Abstract*—Breast cancer is one of the main causes of death across the world in women. Early diagnosis of this type of cancer is critical for treatment and patient care. In this paper, we propose a fast and interpretable deep learning-based pipeline for automatic detection of the metastatic tissues in breast histopathological images. Firstly, the proposed pipeline uses multiple preprocessing and data augmentation techniques to reduce overfitting. Then, the proposed pipeline employs one - cycle policy technique in the pre-trained convolutional neural networks model in shallow and deep fine-tuning phases to find the optimal values. Finally, gradient-weighted class activation mapping (Grad-CAM) technique is utilized to produce a coarse localization map of the important regions in the image. Experiments on the PatchCamelyon dataset demonstrate the superior classification performance of the proposed method over the state-of-the-art.**

***Index Terms*—Breast Cancer, Deep Learning, Grad-CAM, Histopathological, One - Cycle.**

I. Introduction

Approximately 15 % of all cancer-related deaths among females are breast cancer. This number is increasing every year globally, and it is the main cause of death after lung cancer in women [1]. So, it is crucial to detect and diagnose breast cancer early as it can play an essential role in effective treatment planning and patient care. There are different kinds of advanced medical imaging modalities such as thermography, magnetic resonance imaging (MRI), computed tomography scan, ultrasound, and mammography. In this line, histopathological modality imaging is considered the best tool for cancer detection and diagnosis. One of the barriers to the challenging task of manual assessment of large-scale histopathological images is that cancer tissues have numerous variations in appearance with heterogeneous structures and textures. Also, there is a considerable interpathologist variability which usually affects the human interpretation and final diagnosis of pathology images. Another serious barrier in the analysis of histopathological images is the extreme shortage of experienced pathologists [1].

On the other hand, in recent years, deep learning (DL) techniques outperformed state-of-the-art methods in various fields of medical image analysis tasks, such as classification [2], detection [3], and segmentation [4]. Besides, many efforts have been made to apply DL techniques to the histopathological imaging modality [5]–[12]. However, DL models have a complex architecture that consists of many layers with hundreds of thousands of parameters that need to be trained. Especially in the medical domain, since the possible treatments are driven by an informed decision and not by a simple yes/no diagnosis of an algorithm, the automatic computer-assisted diagnosis (CAD) should be interpretable [13].

Along this line, in this paper, we develop an interpretable DL pipeline for recognition of the metastatic tissues in breast histopathological images by utilizing state-of-the-art architecture, one - cycle learning, data augmentation, and gradient-weighted class activation mapping (Grad CAM) techniques. The objective here is to detect breast cancer with better accuracy and higher trust in the existing pipeline, and this is to ultimately aid in the early detection of breast cancer to improve the chances of survival and prognosis. The experimental results on a major publicly available dataset demonstrate that the proposed method generates a comparable performance which also can be interpreted to find which areas in the images are important for the automated decision-making process.

The rest of this paper is organized as follows. Section II summarizes the related research in the prediction of breast cancer. Section III describes the process of the proposed pipeline. Section IV gives the experimental results. Finally, Section V presents the study's conclusion.

978-1-6654-5432-2/22/ $31.00 © 2022 IEEE

II. Related Work

In this section, the related research which is applied in computational histopathology from a methodological perspective is reviewed.

One of the earliest works is done by [6] in 2013, which revolutionized the entire field of digital histopathology. In that study, convolutional neural network (CNN) based pixel prediction is applied to detect mitosis in routinely stained Hematoxylin & Eosin (H&E) breast cancer histology images. Based on this, subsequent methods focused on a combination of CNNs features and handcrafted features. Due to the fact that CNNs are often complex, training of these models leads to an increase in computational run-time and also requires a larger training set to learn distinctive features. To this end, the earliest works such as [7] focused on integrating CNN with biologically interpretable handcrafted features. This type of approach showed a better performance than their CNN counterparts. In [8], the authors made several interesting observations about improving CNN performance by optimizing the hyper-parameters of the network, augmenting the training data, and fine-tuning rather than full training of the model. Albarqouni et al. [9] developed a DL model that utilizes crowd annotations (non-expert) and incorporates it into the CNN learning process via an additional crowdsourcing layer to improve model performance. In [10], an ensemble of histological hashing a class specific manifold learning was proposed for both binary and multi-class breast cancer detection. In [11], authors used a patch-based CNN classifier and integrated a majority voting method in the final classification phase to classify breast cancer histopathology images. Other recent studies like [12] tried to develop different CNN-based pipelines to classify pathology images on a patient-level and whole slide images level.

III. Methodology

A. Dataset

In this study, the modified version of the PatchCamelyon (PCam) benchmark dataset is used [14]. The used dataset consists of 220,000

training and 57,000 evaluation microscopy images with a patch size of 96×96 derived from the Camelyon challenge dataset with a positive label indicating the presence of metastatic tissue in the center region (32×32) of the image. After a careful exploratory analysis of the dataset, we have found that the negative and positive ratio in the dataset is not entirely 50/50 as there are 130,000 negative and 90,000 positive samples. The ratio is closer to the 60/40 ratio meaning that there are 1.5 times more negative images than positives. In Fig. 1, images with and without cancer tissue of the PCam dataset are illustrated.

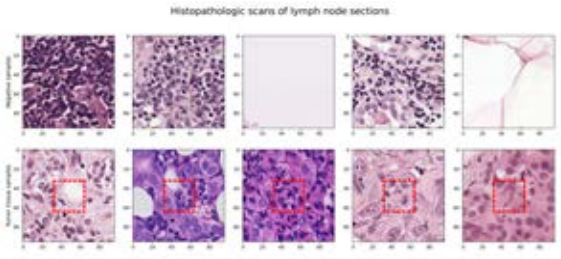

Fig. 1. Different histopathologic scans of lymph node sections of PCam dataset.

B. Interpretable Deep Learning Pipeline

In this study, the proposed pipeline consists of the following section:

- Data pre-processing.
- Model training with one - cycle policy.
- Interpretation of the model by Grad-CAM technique.

C. Data Pre-processing

In this study, the images from the PCam dataset are pre-processed and subjected to data augmentation. In the pre-processing step, first, the images are cropped to the smallest rectangular (with the size of 48×48.). Next, the intensity normalization is applied. Also, data augmentation and image normalization techniques are adopted to combat over-fitting. Here, various augmentation techniques are implemented such as horizontal flip, vertical flip, random crop, random rotation, and random lighting.

D. Model Training with One - Cycle Policy

In this study, the entire model training process is mainly done with one - cycle policy [15]. After pre-processing images, once we have the prepared data, they will be fed into the feature learning phase. One of the issues with almost every machine learning model is that there are several hyper-parameters that can have a direct effect on the performance of that model. Specifically, in DL applications where training takes a lot of time, tuning these hyper-parameters is a more of trial and error process. Learning-rate might be the most important hyper-parameter in DL models, as learning-rate decides how much gradient to be back propagated. The small learning-rate makes the model converge slowly, while the large learning-rate makes the model diverge. So, the learning-rate needs to be optimized. Here, we employ the one - cycle policy technique to find the optimal learning-rate and weight decay values. One - cycle policy results in a more organized and optimized approach for selecting hyper-parameters such as learning-rate, weight decay, and momentum. This will save a lot of time with training with sub-optimal hyper-parameters. One - cycle learning policy is based on a simpler technique that is developed in [16] which is called cyclical learning rate (CLR).

We either utilize a more extensive pre-trained deep convolutional neural network model, densely connected convolutional networks (DenseNet) [5] with transfer learning to adjust the weights to our dataset. This architecture provides state-of-the-art results on ImageNet challenge with much fewer parameters and complexity than its rivals.

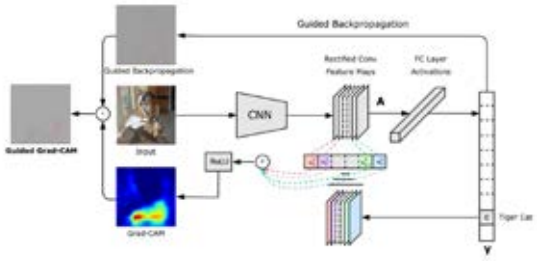

Fig. 2. Complete pipeline of Grad-CAM [17].

E. Model Interpretation by Gradient-weighted Class Activation Mapping

The Grad-CAM [17] technique can highlight discriminative regions to provide visual interpretation for model decisions in detection. It is a very effective technique for determining how to make the model interpretable to experts in the field. Here, we provide visual insights into how exactly the model is making its decision. This is achieved by utilizing the gradient information flowing into the last convolutional layer of the CNN to understand each neuron for a decision of interest and produce a coarse localization map highlighting the important regions in the image for predicting the concept. In this way, by taking the final convolutional feature map and then weighting every channel in that feature with the gradient of the class with respect to the channel, intensively the input image activates different channels with regard to how much important each channel is to the target class. To obtain the class discriminative localization map of width u and height v for any class c, first, the gradient of the score for class c, y^c should be calculated with respect to feature maps A^k of a convolutional layer, i.e. $\frac{\partial y^c}{\partial A^k}$. These gradients flowing back are global average-pooled over the width and height dimensions (indexed by i and j respectively) to obtain the neuron importance weights α_k^c for the target class as in [17]:

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \tag{1}$$

where, Z is the number of pixels in the feature map. After the calculation of α_k^c for the target class c, a weighted combination of activation maps and the ReLU operator can be performed to get the final class discriminative map as the following [17]:

$$L_{Grad-CAM}^c = ReLU\Big(\sum_k \alpha_k^c A^k\Big) \tag{2}$$

This results in a coarse heatmap of the same size as that of the convolutional feature maps. The complete pipeline of Grad-CAM is illustrated in Fig. 2.

IV. Implementation Details

Since our dataset is retrieved from Kaggle, we have access only to the labels of the training set. Therefore, we split the training set into 90% training set and 10% validation set. We evaluate the performance of the model on the validation set. The training process utilizes the shallow and deep fine-tuning technique to adjust ImageNet pre-trained weights to the dataset. Due to this fact, we resized all images to 224×224 as it is the required input size for DenseNet architecture. DenseNet-169 variant is used as the base model with a drop-out value of 0.5 to reduce over-fitting. The batch size is set to 256. Also, the output layer of our pre-trained network has been modified, so that

the last dense layer outputs 2 classes which correspond to benign or malignant. And, stochastic gradient descent (SGD) is used as the training optimizer.

A. Evaluation Criteria

We evaluate the performance of the proposed pipeline based on standard metrics that are defined as:

$$Precision = \frac{TP}{(TP + FP)} \tag{3}$$

$$Recall = \frac{TP}{(TP + FN)} \tag{4}$$

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)} \tag{5}$$

$$F1 - score = \frac{(2 \times Precision \times Recall)}{(Precision + Recall)} \tag{6}$$

where TP is true positive, FP denotes false positive, TN defines true negative, and FN is false negative. Also, we also use the confusion matrix (CM) and the area under the curve (AUC) for evaluating the performance of the proposed pipeline.

B. Results and Discussion

In this section, we show the applied experiments with one - cycle policy technique to find an optimal value for learning-rate and weight decay. As for the weight decay that is the L penalty of the optimizer, the author of [15] suggests selecting the largest value that allows us to train with a high learning-rate. In order to do that, a grid search for weight decays with values $1e-2$, $1e-4$, and $1e-6$ is adjusted. The result is illustrated in Fig. 3. Here, $1e-4$ seems like the largest weight decay that gets to a low loss and has the highest learning-rate before shooting up. Also, we select $2e-2$ as our initial learning-rate for the first phase of the fine-tuning process.

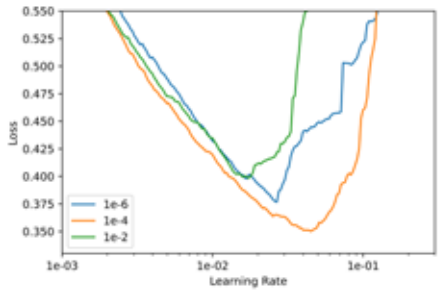

Fig. 3. Grid search on weight decay.

In the first phase of the fine-tuning process, the last dense layer is adjusted and the rest of the model is frozen which is called shallow fine-tuning. To show the advantages of one - cycle policy, training the head of the model is done over only 12 epochs. The result for this phase is shown in Fig. 4. The best accuracy of the validation set is 95.67 %.

In Fig. 4, it is obvious that starting with a low learning-rate and gradually increasing it to a higher value results in a good performance. The higher rate has a regularizing effect as it makes the model skip sharp and narrow local minima and settle for a wider and more stable one. This effect can be seen in the loss plot: as the learning-rate increases in the first half of a cycle, the loss rises often. These increases drive the model out of local minima. However, this phenomenon will pay off in the end when the learning-rates decreased (second half of the cycle).

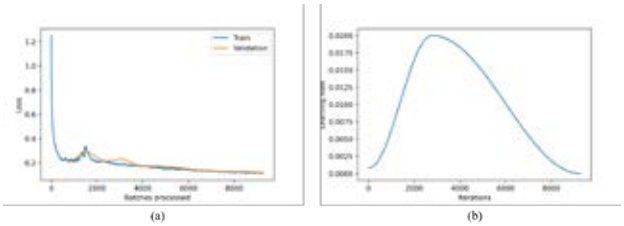

Fig. 4. a) Loss and b) learning-rate of shallow fine-tuning.

In the deep fine-tuning, all the trainable parameters in the bottom layers can be unfrozen and trained at a slower learning-rate. These layers can recognize general shapes and patterns because they were trained on the ImageNet dataset. To avoid drastically altering the weights, a lower learning-rate is adopted. Here, the minimum and maximum learning-rate of one - cycle policy is chosen as ($4e-5$,$4e-4$). In this range, the loss value is at its lowest value and right before it shoots up. With deep fine-tuning, the performance becomes better and on the validation set, the proposed pipeline reaches an accuracy of 97.11 %. The result of deep fine-tuning is shown in Fig. 5. The validation result shows that with reaching the end of the cycle, the loss of validation set separates from training performance. This means that the model has started over-fitting during the small learning-rates (end of the cycle). with further training, the model starts over-fitting and only learns the features of the training set. Therefore, using one - cycle policy not only helps us to reach a good performance in only 20 epochs but also tells us where to stop training.

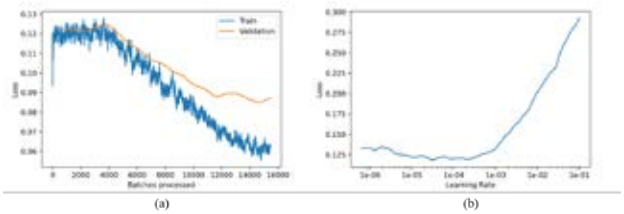

Fig. 5. a) Loss and b) learning-rate of deep fine-tuning.

The CM of both shallow and deep fine-tuning phases is depicted in Fig. 6. It is shown that the model has learned to distinguish tumor and non-tumor samples in both phases. However, by applying the deep fine-tuning and adjusting bottom layers with a much lower learning-rate, the model reached a boost in the overall performance. 0 indicates negative samples (non-tumor) and 1 indicates positive samples (tumor). In this case, reaching a higher AUC shows the better capability of the model at predicting non-tumor samples as benign and tumor samples as malignant. As it is shown in Fig. 7, the model has reached the AUC score of 99.46 %.

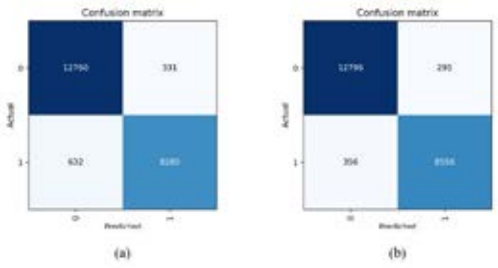

Fig. 6. Confusion matrix of (a) shallow and (b) deep fine-tuning.

Due to the fact that the utilized dataset is being accessed from the Kaggle challenge, the performance of the model can be only compared to [12]. In Table 1, the evaluation criteria of the model with deep fine-tuning on the validation set is shown and compared with the recent work. For clarity, we have featured the best results in bold. It is clear that by using one - cycle policy for training and tuning hyper-parameters, the proposed method outperformed an ensemble of four different CNN architectures that are used in [12] in terms of both evaluation metrics and computational run-time.

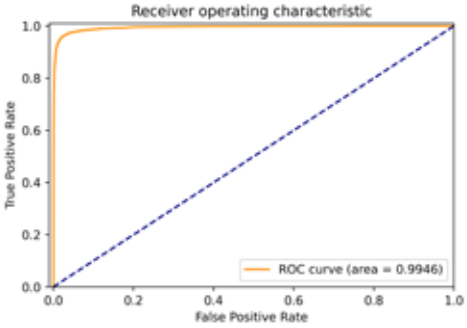

Fig. 7. ROC curve and AUC score.

TABLE I
PERFORMANCE OF COMPARISON MODELS ON THE PCAM DATASET IN TERMS OF PRECISION (%) RECALL (%), ACCURACY (%), F1-SCORE (%), AND EPOCHS. THE HIGHEST RESULTS IS SHOWN IN BOLD.

Reference	Precision	Recall	Accuracy	F1 - score	Epochs
The proposed pipeline	**96.67** %	**96.00** %	**97.11** %	**96.33** %	**32**
Kassani et al. [12]	95.70 %	95.27 %	94.64 %	95.50 %	1000

To make our classification pipeline interpretable, the visualization of the learned features is illustrated in Fig. 8. The Grad-CAM is used as the tool for generating a coarse localization map to highlight the locations for the classification decision. Fig. 8 shows the activation maps of the predicted category so if the label is the tumor, the visualization shows all the locations where the model believes the tumor patterns to be.

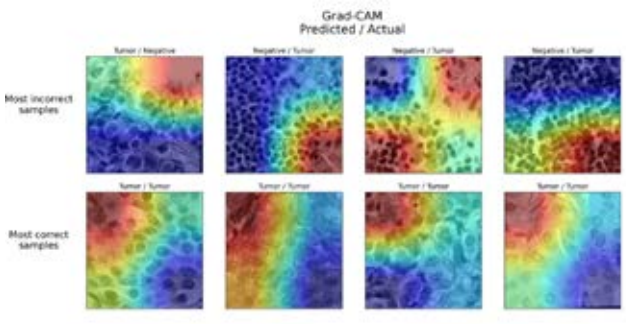

Fig. 8. Activation maps of the predicted class.

V. CONCLUSION

In this paper, we present a fast and interpretable DL pipeline for recognition of the metastatic tissues in breast histopathological images. The proposed pipeline mainly utilizes the one-cycle policy technique to find the optimal learning-rate and weight decay values. Besides, Grad-CAM technique is employed to interpret what the model has learned. Also, the pre-trained DenseNet-169 network is utilized as the pipeline's model. And, the data augmentation technique is performed to reduce over-fitting. The proposed model is evaluated on the modified PCam dataset. Experimental results demonstrate the superior classification performance of the proposed pipeline over the state-of-the-art. As the proposed pipeline becomes more robust, it might be useful to the healthcare system and pathologists. As future direction, it may be possible to achieve increased performance by introducing a novel model.

REFERENCES

[1] S. Sharma, and R. Mehra, "Conventional Machine Learning and Deep Learning Approach for Multi- Classification of Breast Cancer Histopathology Images—a Comparative Insight," *J. of Digi. Imaging*, Vol. 3, no. 3, pp. 632–654, 2020.

[2] S. Mardanisamani, F. Maleki, S. H. Kassani, S. Rajapaksa, et al., "Crop Lodging Prediction from UAV-Acquired Images of Wheat and Canola Using a DCNN Augmented with Handcrafted Texture Features," *In Proceed. of of the IEEE/CVF Conf. on Comp. Vision and Patt. Recog. Work.*, pp. 0–0, 2019.

[3] P. Herent, B. Schmauch, P. Jehanno, O. Dehaene, et al., " Detection and Characterization of MRI Breast Lesions Using Deep Learning," *Diag. and Inter. Imaging*, Vol. 100, No. 4, pp. 219–225, 2019.

[4] F. Lateef, and Y. Ruichek, " Survey on Semantic Segmentation Using Deep Learning Techniques," *Neuro.*, Vol. 338, pp. 321–348, 2019.

[5] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," *In Proceed. of the IEEE Conf. on Comp. Vision and Patt. Recog.*, pp. 4700–4708, 2017.

[6] D. C. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis Detection in Breast Cancer Histology Images with Deep Neural Networks," *In Inter. Conf. on Med. Image Compu. and Comp.-Assi. Inter.*, pp. 411–418, Springer 2013.

[7] F. Xing, Y. Xie, and L. Yang, "An Automatic Learning-Based Framework for Robust Nucleus Segmentation," *IEEE Tran. on Med. Imaging*, Vol. 35, No. 2, pp. 550–566, 2015.

[8] Z. Gao, L. Wang, L. Zhou, and J. Zhang, "HEp-2 Cell Image Classification with Deep Convolutional Neural Networks," *IEEE J. of Biom. and Heal. Info.*, Vol. 21, No. 2, pp. 416–428, 2016.

[9] S. Albarqouni, C. Baur, F. Achilles, V. Belagiannis, et al., "Aggnet: Deep Learning from Crowds for Mitosis Detection in Breast Cancer Histology Images," *IEEE Tran. on Med. Imaging*, Vol. 35, No. 5, pp. 1313–1321, 2016.

[10] S. Pratiher, and S. Chattoraj, " Diving Deep onto Discriminative Ensemble of Histological Hashing & Class-Specific Manifold Learning for Multi-Class Breast Carcinoma Taxonomy," *In ICASSP 2019-2019 IEEE Inter. Conf. on Acou., Spee. and Sign. Proc. (ICASSP)*, pp. 1025–1029, 2019.

[11] K. Roy, D. Banik, D. Bhattacharjee, and M. Nasipuri, "Patch-based System for Classification of Breast Histology Images Using Deep Learning," *Comput. Med. Imaging and Graph.*, Vol. 71, pp. 90–103, 2019.

[12] S. H. Kassani, P. H. Kassani, M. J. Wesolowski, K. A. Schneider, et al., " Classification of Histopathological Biopsy Images Using Ensemble of Deep Learning Networks," *arXiv preprint arXiv:*, pp. 1909–11870, 2019.

[13] M. A. Naji, A. Aghagolzadeh, and M. Ezoji, "Ensemble of CNN for Multi-Focus Image Fusion," *Info. Fusion*, Vol. 51, pp. 201–214, 2019.

[14] B. S. Veeling, J. Linmans, J. Winkens, T. Cohen, et al., "Rotation Equivariant CNNs for Digital Pathology," *In Inter. Conf. on Med. Image Compu. and Comp.-Assi. Inter.*, pp. 210–218, Springer 2018.

[15] L. N. Smith, "A Disciplined Approach to Neural Network Hyper-Parameters: Part 1–Learning Rate, Batch Size, Momentum, and Weight Decay," *arXiv preprint arXiv:*, pp. 1803–09820, 2018.

[16] L. N. Smith, "Cyclical Learning Rates for Training Neural Networks," *In 2017 IEEE Winter Conf. on Appl. of Comp. Vision (WACV)*, pp. 464–472, 2017.

[17] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, et al., "Grad-cam: Visual Explanations from Deep Networks via Gradient-Based Localization ," *In Proceed. of the IEEE Inter. Conf. on Comp. Vision*, pp. 618–626, 2017.