ECTI Transactions on Computer and Information Technology
Journal homepage: https://ph01.tci-thaijo.org/index.php/ecticit/
Published by the ECTI Association, Thailand, ISSN: 2286-9131

An Optimal Deep Learning Approach to BCa Tissue Detection using Case Studies

Kasikrit Damkliang[1], Thakerng Wongsirichot[2], Chawisa Khongrak[3] and Piyathida Suwannarat[4]

ABSTRACT

In this paper, we present a Deep Learning (DL) model with optimized performance for breast cancer (BCa) tissue classification. A simple DL approach is applied to the analysis of invasive ductal carcinoma (IDC) tissue, which is the most common BCa subtype. Binary classification of non-IDC and IDC tissues is proposed using Convolutional Neural Networks (CNN) in the training and prediction phases. Our trained model achieved F1 and sensitivity scores of 0.88, as well as micro-average values of 0.94 for the area under the curve (AUC) of the receiver operating characteristic (ROC) curve and 0.95 for the area under the precision-recall curve. Since the file size of our model is small, it has the potential for application in real-world scenarios.

Article information:

Keywords: Breast Cancer Tissue, Deep Convolutional Neural Network, Invasive Ductal Carcinoma, Classification, Cross-validation

Article history:
Received: October 10, 2022
Revised: December 24, 2022
Accepted: January 21, 2023
Published: February 4, 2023 (Online)

DOI: 10.37936/ecti-cit.2023171.250441

1. INTRODUCTION

Nowadays, Artificial Intelligence (AI) is increasingly applied in various fields of research. A crucial element of AI is the size of the dataset, and Machine Learning (ML) is a vital tool for analyzing the acquired data. Specifically, AI plays an essential role in the medical community, helping doctors diagnose complex diseases, e.g., analyzing cancer from a breast X-ray to predict the risk of breast cancer from the picture. A high level of accuracy helps detect people at risk for early treatment at a lower cost and with a higher chance of being cured.

As an example, a pathologist requires years of work experience to rapidly diagnose conditions. In contrast, AI can be trained in a much shorter time through the use of Deep Learning (DL) algorithms. Rapid, accurate data analysis can improve diagnosis. When the deep learning network has been trained with an adequately robust training set, it is usually able to generalize the learned performance to unseen situations, obviating the need for manually input recognition features.

The accuracy of AI depends on the size and number of datasets analyzed. More extensive neural networks can be formed by exploiting DL algorithms. DL methods involve deriving suitable feature spaces solely from the data itself. Learning from training exercises will help pave the way for the generalization of the learned model to other independent test sets.

Breast cancer (BCa) is the most prevalent disease among women. It is also the leading cancer death for women worldwide. Invasive ductal carcinoma (IDC) is the most common of all BCa subtypes around the world [1]. It spreads outside the walls of the milk ducts and into the fat tissues of the breast. Cancer can spread from this point to other parts of the body through the lymphatic and blood systems. Pathologists classify cancers by assigning an aggressiveness grade to a whole mounted tissue sample. When doing so, they typically focus on the regions which contain the IDC. Automatic aggressiveness grading commonly uses a pre-processing step in which the precise extent of the IDC is delineated in a whole-slide image (WSI) [2] [3]. Janowczyk investigated IDC detection using a DL approach without handcrafted features extraction. An F1-score of 0.7648 was reported [3].

In a study that could be applied to cancerous breast histopathology analysis [4], Xu et al. [5] classified epithelial and stromal tissues from colorectal cancer analysis. Xu used Local Binary Patterns (LBP) requiring handcrafted features and compared the results obtained with an end-to-end feature extractor using deep Convolutional Neural Networks (CNN), which achieved an F1 classification score of

[1,2,3,4] The authors are with Division of Computational Science, Faculty of Science, Prince of Songkla University, Hat Yai, Songkhla, Thailand, 90112., E-mail: kasikrit.d@psu.ac.th, thakerng.w@psu.ac.th, 5910210063@psu.ac.th and 5910210211@psu.ac.th

0.85. The CNN-based approach outperformed the approach that used handcrafted feature extraction to discriminate between epithelial and stromal tissues.

In 2017, the CAMELYON16 Challenge competition was organized to detect of lymph node metastases in women with BCa [6]. Complex algorithms of both DL and ML were utilized to classify cancer metastases in WSIs.

Mercan et al. [7] proposed a framework for learning features of histopathology images at the region level. A pre-trained CNN was fine-tuned using small fixed-size samples from the regions of interest (ROIs). Their experiments achieved a 4-class classification in ROIs on the breast histopathology slides. The CNN-based method achieved an accuracy of 0.6680.

In 2020, Shahidi et al. [8] utilized a dataset of breast cancer histopathology images (BreakHis) [9]. A set of pre-trained DL models with weights derived from ImageNet [10] was fine-tuned and examined for tissue classifications, e.g., ResNeXt [11], Dual Path Network (DPN) [12], Squeeze and Excitation Network (SENet) [13], and Neural Architecture Search Network (NASNet) [14] [15]. Pre-processing included resizing, class re-balancing, normalization, image contrast enhancement, multi-resolution segmentation, stain normalization, and stain color transfer between normalized images were also performed. Shahidi et al. reported that factors such as pre-processing, data augmentation, and transfer learning methods could influence model performances. At the image patch level, the best accuracy of 0.99 for multiclass classifications was achieved by the InceptionResNetV2 architecture [16]. The model training's execution time was about 12 hours.

In 2020, Wang et al. [17] utilized an exciting dataset of the IDC tissue type of which characteristics were close to our dataset in this work. Raw pathological images were provided by the Hospital of the University of Pennsylvania. They presented a deep hybrid learning for detecting IDC. A CNN-enabled-Gated Recurrent Unit Network (GRU) architecture was proposed to solve a vanishing gradient issue. However, they did not report all trainable parameters. Their approach achieved a 0.88 F1-score and a corresponding value of 0.89 area under the curve (AUC), respectively.

For a text format, machine learning approaches are optimal for text-based feature analysis. Mojrian et al. [18] utilized the radial basis function (RBF) to analyze the Wisconsin Breast Cancer Data set (WBCD) [19] for the binary classification, (benign and malignant). They reported the coefficient determination of a model evaluation metric of 0.9374. Jabbar et al. [20] utilized the same dataset. The patient-level classification of benign and malignant was reported with a 0.97 accuracy. The ensemble approach of machine learning classifiers including, the Bayesian Network and the RBF, was proposed. Selected ten attributes of the dataset were extracted and used for the classifier development. However, Khan et al. [21] adopted a one-dimensional convolutional neural network (Conv1D) to learn a text-based dataset. Khan also utilized logistic regression (LR), decision tree (DT), random forest (RF), the voting classifier (VC), and support vector machine (SVM). They reported the F1-score of 0.96 given by DT.

To improve DL methods for practical application, we focused on an end-to-end analysis approach of pixels-based avoiding handcrafted engineering features. In contrast, the pixels-based analysis using the DL methods usually requires a deep architecture of networks and massive datasets. Therefore, the handcrafted engineering features and the end-to-end analysis using only pixels, induce time-consuming and highly complex computation.

In this work, we propose an optimal approach with satisfactory analytical performances using DL performing on raw pixels. We used a DL technique to find the optimal model for BCa tissue analysis. The ultimate goal is to develop an automatic screening tool for BCa detection that identifies whether a WSI presents evidence of an IDC. The method reduced the execution time compared to the techniques which require experts to determine the severity or age of carcinoma and those more complex approaches, which are more challenging to deploy.

The training was conducted at the image patch level using patches of 48 × 48 pixels with data augmentation. The training dataset contained patches from around 280,000 images. Predictions were made at the histology slide level to create tumor probability maps using three data sources of images previously unseen by the model.

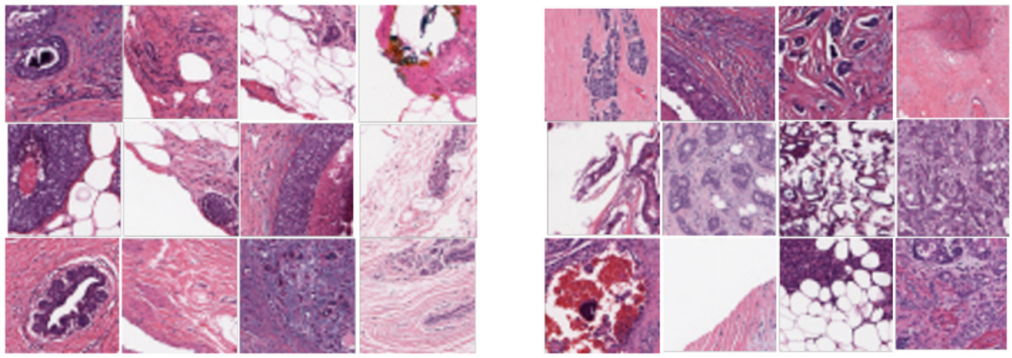
The following section presents materials and methods. Results of model performance evaluation and IDC probability maps are reported in Section 3. The discussion and conclusion are presented in Sections 4 and 5, respectively.

2. MATERIALS AND METHODS

In this section, the materials and proposed methods in this work are presented.

2.1 Dataset for Model Development

Originally, WSIs of diagnosed IDCs from 162 women at the Hospital of the University of Pennsylvania (HUP) and The Cancer Institute of New Jersey (CINJ) were digitized. Cruz-Roa et al. [22] introduced a dataset consisting of IDC tissue regions which were scanned at 40 × (0.25 µm/pixel resolution) and stained using Hematoxylin & Eosin (H&E). Due to their immense size, these WSIs were down-sampled by a factor of 16:1 to a resolution of 4 µm/pixel. Therefore, this colossal dataset with the mentioned attributes is suitable for cancerous tissue classification using DL approaches.



(a) Non-IDC (0) (b) IDC (1)

Fig.1: *Example BCa images of Non-IDC and IDC from the dataset.*

Table 1: *Original unbalanced IDC dataset is organized into three categories.*

Objective	Class	No. of image patches	Total
Training	Non-IDC	143,056	199,818
	IDC	56,762	
Validation	Non-IDC	15,889	22,201
	IDC	6,312	
Testing	Non-IDC	39,793	55,505
	IDC	15,712	
		Total	277,524

Note: IDC = Invasive Ductal Carcinoma

In 2015, Janowczyk et al. [2] [3] curated this dataset and made it available in the public domain on the Kaggle website [23]. Example patches of these BCa images are presented in Figure 1. The dataset comprised 277,524 image patches of 50 × 50 pixels were extracted from the WSIs in the Kaggle dataset. The WSIs included 198,738 non-IDC (negative) and 78,786 IDC (positive) diagnoses.

In this work, the image patches were incoherently selected, and organized into three categories, as presented in Table 1. The dataset was split 80:20 between training and testing sets. In the training set, a portion of 10% was divided into the validation set.

2.2 Datasets for Prediction

We curated relevant WSI datasets of IDC-negative and -positive BCa from various data sources. These datasets were completely unseen by our classifier (the trained model with the best weights). The attributes of the datasets were outlined in Table 2. Key features considered were the same for the prediction set as for the training set, e.g., magnification, staining, color mode, and file format.

The BCa datasets used for prediction were collected from CINJ, Case Western Reserve University (CWRU) [24] [25], and BreakHis [9]. The histological samples used to create these datasets had been H&E stained, and the images were in the PNG format digitized at the same 40X magnification in the RGB color mode. The datasets of CINJ and CWRU were provided by The Cancer Genome Atlas and Genomic Data Commons (GDC) Data Portal. The BreakHis dataset comprised microscopic image samples, which were generated from breast tissue biopsy slides stained with H&E. This dataset was built in collaboration with the P&D Laboratory, Pathological Anatomy, and Cytopathology, Parana, Brazil. The image samples were prepared for histological study and labeled by the pathologists. The samples were collected by the Surgical Open Biopsy (SOB) method. Currently, the dataset contains images of 2,480 benign and 5,429 malignant samples from 82 patients. The images measure 700 × 460 pixels in the PNG format, at different magnifying factors in the 3-channel RGB using the 8-bit depth in each channel [9] [26].

2.3 Proposed Approach

In this section, we proposed a DL method for BCa tissue classification of benign and malignant. Our goal aims to create an automatic screen tool that can detect an IDC tissue type in a slide-level prediction. We exploited an end-to-end approach for model development. The proposed method consists of training and prediction phases, as shown in Figure 2.

In preparation for training, non-IDC and IDC image patches measuring 50 × 50 raw pixels were resized to 48 × 48 pixels and then fed into the designed network of our model (details in Section 2.4). The data augmentation of these patches was processed on-the-fly during the training processes.

In 2019, Ornek et al. [27] reported that classification performance was significantly improved by image augmentation with simple enhancement methods. The parameters and configuration of the augmentation carried out in this work are reported in Table 3. Since the tissue slide images have no orientation, the image patches can be flipped horizontally and vertically. We found empirically that the model returned a better performance by using a zoom range of 25% and reflecting the fill mode. Batches of 32 patches

Table 2: *Datasets and their attributes for prediction.*

Dataset	Scanner	Mag.	Stained	Size (pixels)	Color mode	Format
CINJ	Aperio	40X	H&E	1,500 - 4,000	RGB	PNG
CWRU	Ventana	40X	H&E	1,500 - 6,000	RGB	PNG
BreakHis	Olympus	40X	H&E	700 × 460	RGB	PNG

Note: CINJ = The Cancer Institute of New Jersey, CWRU = Case Western Reserve University, BreakHis = The Breast Cancer Histopathological Image Classification

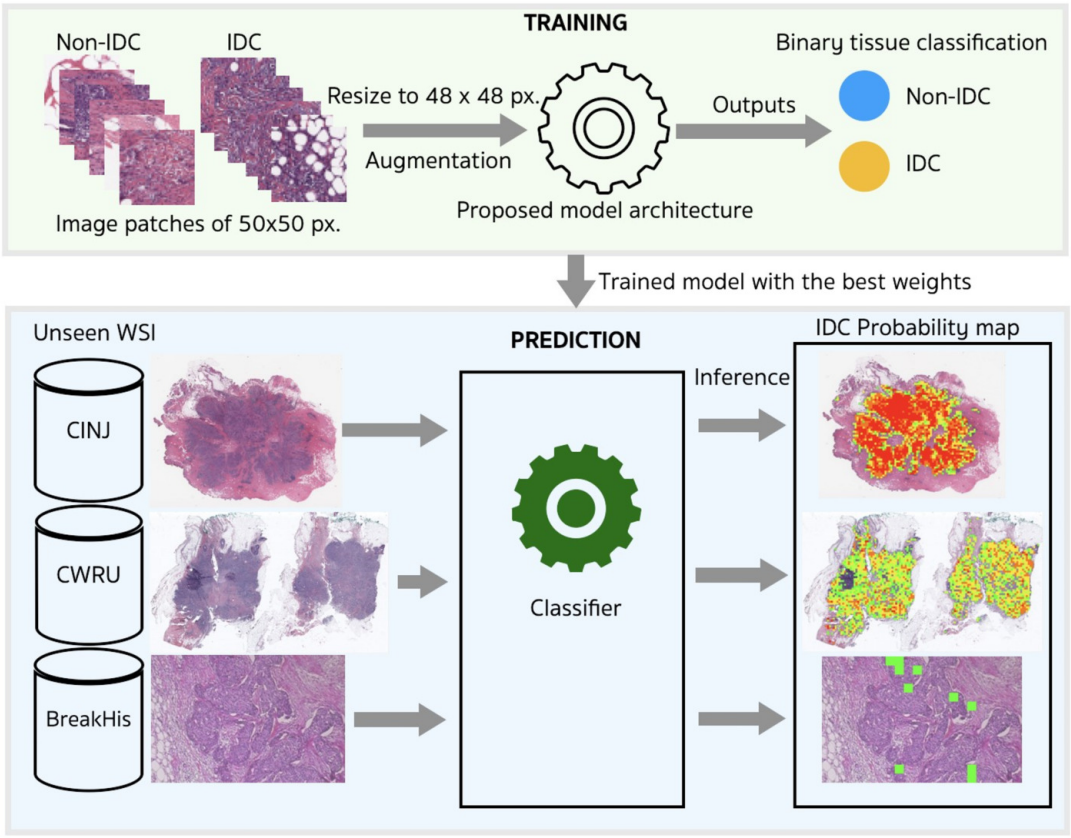


Fig.2: *Deep Learning approach for breast cancer tissue classification.*

were fed into the proposed model during the training phase.

The outputs produced by the model were binary predictions of non-IDC (0; negative) or IDC (1; positive) tissue. In the prediction phase, the trained model with the best weights was selected to be our classifier. The unseen WSIs from the three datasets were used to create IDC probability maps as our final results. The image patches were kept small to ensure smooth and fine-grained representations. A small image block of 32 × 32 pixels was selected, and a Gaussian. The filter of 3 × 3 pixels was applied for smooth operation.

2.4 Model Architecture

The structural architecture of the proposed model is shown in Figure 3. Multi-dimensional feature vectors were computed using 2D separable CNN layers with kernel sizes of 3 × 3 pixels and the Rectified Linear Unit (ReLu) activation function. The last feature vectors of samples with the dimensionality of 256 × 6 × 6 were pooled into a multilayer perceptron (MLP) using Global Average Pooling 2D (GAP2D) [28]. The dense layers of 256, 128, 64, and 32 perceptrons generated squeezed-feature vectors of *samples x number*, where the number is the corresponding perceptron units. The ReLu activation function [29], as defined in Equation 1, was applied to all the dense layers. Finally, the 32-dimensional feature vectors were fed forward into the Softmax (Equation 2) [30], the activation function layer, which normalized classification abilities over the binary classes of the tissues where z_j is the feature vector in K examples.

$$g(\vec{z}) = \max(0, z) \tag{1}$$

$$\sigma(\vec{z}) = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \tag{2}$$

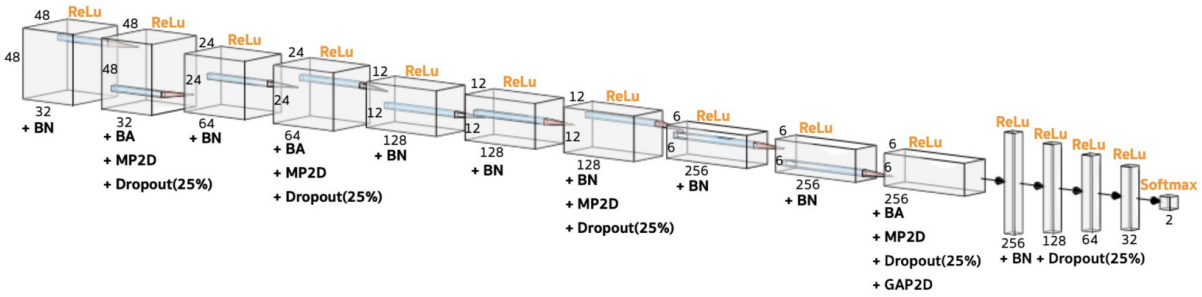


Fig.3: *Structures of the proposed model architecture presented in Figure 2: (BN = Batch Normalization, MP2D = Max Pooling 2D, GAP2D = Global Average Pooling 2D.*

Table 3: *Parameters of data augmentation and their configurations for the training patches.*

Parameter	Configuration	Description
Rescale	1.0/255.0	Rescaling factor, data will be multiplied by the value
Horizontal flip	True	Randomly flip inputs horizontally
Vertical flip	True	Randomly flip inputs horizontally
Rotation range	20	Degree range for random rotations
Width shift range	0.1	Fraction of total width to random shift
Height shift range	0.1	Fraction of total width to random shift
Zoom range	0.25	The range for random zoom
Fill mode	Reflect	Points outside the boundaries of the input are filled

Batch normalization [31] and regulation [32] rates of 25% were deployed for all layers. When the feature vectors changed the dimensionalities, e.g., from 32 × 48 × 48 to 64 × 24 × 24, max pooling 2D (MP2D) with a filter size of 2 × 2, and the same regulation rate were applied to preserve feature characteristics and help protect against overfitting.

Table 4: *Patch sizes were probed, and model performances were evaluated in the pilot study.*

Patch size (pixels)	F1 score
75 × 75	0.65
64 × 64	0.76
50 × 50	0.81
48 × 48	0.85
32 × 32	0.78

For the last CNN, GAP2D was applied to the weights before the weights were fed forward into the 256-dense layer. The total number of parameters of these weights was 344,610. The number of parameters would be more than double if the weights were processed in the flattened layer instead of GAP2D. To compensate for the unbalanced class of the dataset shown in Table 1, class weights of the training set were computed and used during the training. Table 5 lists the hyper-parameters and their configurations for the proposed model.

2.5 Model Training

The training and validation of the model were determined using many hyper-parameters. In our pilot study, patch sizes of 75, 64, 50, and 48 pixels were probed using the CNN kernel sizes of 3 × 3. We found that the size of 48 × 48 pixels gave the best validation scores due to its most effective receptive field [33] exploited by the kernel size of CNN (see Table 4).

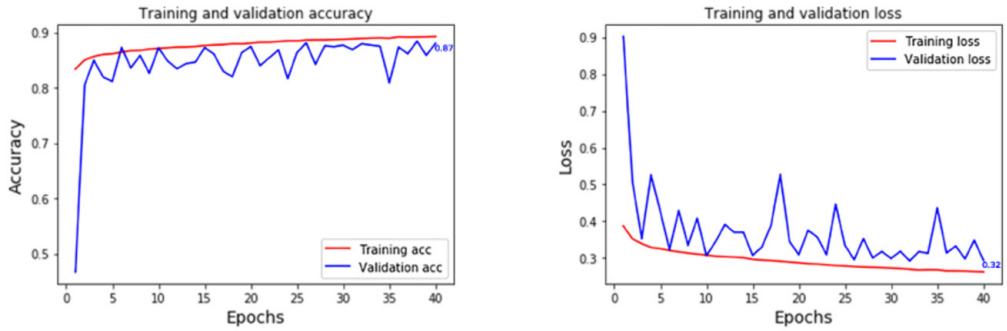
Numbers of MLP layers and their neurons of the model architecture were trialed. Batch sizes, kernel sizes, activation functions, optimizers (Adagrad [34], Adam, Adamax [35], and Gradient descent with momentum (SGD) [36]), loss functions (binary cross-entropy and categorical cross-entropy), and learning rates were also fine-tuned. Table 5 concludes the optimal hyper-parameters and their configurations for the proposed model. The data augmentation and 5-fold cross-validation were applied.

Figure 4 shows the accuracy and loss graphs of 40 epochs in the training and validation of the proposed architecture. During the training of this dataset, each fold required approximately six hours. The final validation achieves 0.87 accuracy and 0.32 loss, consecutively. The graphs show that the training process still generated some degree of overfitting. However, the charts of accuracy and loss are convergent each other, which is acceptable.

3. RESULTS

3.1 Model Performance Evaluations

The model performance was evaluated using the metrics defined in Equations 3 - 11 [37] [38]. The metrics include the accuracy, F1-score, precision (posi-



(a) Accuracy of training and validation (b) Loss of training and validation

Fig.4: *Training and validation accuracy of the proposed architecture.*

tive predictive value; PPV), negative predictive value (NPV), sensitivity (true positive rate; TPR), specificity (true negative rate; TNR), balanced accuracy (BA), false positive rate (FPR), and false negative rate (FNR). The metrics were analyzed where TP = true positive, TN = true negative, FP = false positive, and FN = false negative.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

$$F1_{score} = \frac{2TP}{2TP + FP + FN} \tag{4}$$

$$precision(PPV) = \frac{TN}{TP + FP} \tag{5}$$

$$precision(NPV) = \frac{TN}{TN + FN} \tag{6}$$

$$sensitivity(TPR) = \frac{TP}{TP + FN} \tag{7}$$

$$specificity(TNR) = \frac{TP}{TP + FN} \tag{8}$$

$$BA = \frac{TPR + TNR}{2} \tag{9}$$

$$FPR = \frac{FP}{FP + TN} = 1 - TNR \tag{10}$$

$$FNR = \frac{FN}{FN + TP} = 1 - TPR \tag{11}$$

Table 5: *Hyper-parameters and their configurations for the proposed model.*

Parameter	Configuration
Train-test ratio	0.8
Validate ratio	0.1
Patch size	48 × 48
Batch size	32
Kernel size	3 × 3
Activation function	ReLu for CNN layers and MLPs Softmax for prediction layer
Loss function	Binary cross-entropy
Optimizer	Gradient descent optimization (SGD) algorithm
ReduceLROnPlateau	Monitor on loss with factor = 0.3, patience = 3
EarlyStopping	Monitor on loss with mode = minimum, patience = 3

The model performances were evaluated at the image-patch level using the testing set that consisted of 55,505 image patches (non-IDC 39,793 and IDC 15,712) (Table 1). The classification report of the patch-level evaluation is presented in Table 6. The report provides information about Precision (PPV), Sensitivity (Recall), and F1-score. The classifier model achieved the weighted average F1-score of 0.88. Another performance evaluation was also set up. Training, validation, and testing were preprocessed to make balanced classes. The minor type was oversampled to reach the number of patches of the significant type utilizing the Imblearn python package [39]. The total images consist of 397,476 patches which is a 42.2% increase from the total images of the unbalanced dataset. Stratified 5-folds cross-validation of the training, validation, and testing was performed. The model returned the F1-score of 0.86, which means that the model has good stability compared to the F1-score of 0.88 of the unbalanced dataset.

Receiver operating characteristic curves (ROC) and the corresponding values of AUC calculated from TPR (sensitivity) and FPR (1-specificity) were evaluated. The model achieved a micro-average of 0.94 of AUC. However, Saito et al. [38] reported that the precision-recall plot is more informative than the ROC plot in the case of an imbalanced class performance evaluation such as our testing set in this work. Therefore, Figure 5 presents graphs plotted from the relationship between precision (PPV) and recall (sensitivity) computed using the binary prediction for

each tissue class. The micro-average area of the precision-recall of our classifier was 0.95, which is close to the corresponding calculated AUC of the plot of sensitivity against 1-specificity.

Table 6: *Classification report of the testing set.*

Class	Precision (PPV)	Sensitivity (Recall)	F1-score	Support
Non-IDC	0.94	0.89	0.91	39,793
IDC	0.75	0.85	0.80	15,712
Accuracy			0.88	55,505
Macro avg	0.84	0.87	0.86	55,505
Weighted avg	0.88	0.88	**0.88**	55,505

Note: IDC = Invasive Ductal Carcinoma, PPV = Positive Predictive Value

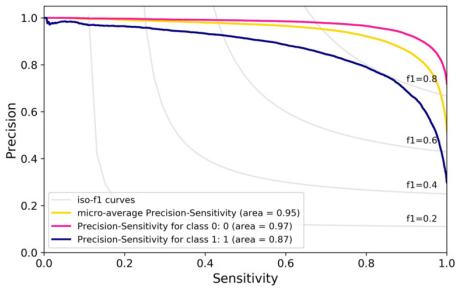


Fig.5: *Precision-recall (sensitivity) curve evaluated using the testing set (0: Non-IDC, 1: IDC).*

3.2 IDC Probability Maps

Tissue classification maps of unseen WSIs were processed by our classifier using the datasets from CINJ, CWRU, and BreakHis (Table 2). A slide was decomposed into image patches of 32 × 32 pixels. The actual coordinates of these patches were recorded. Each patch was resized to the size of 48 × 48 pixels. The size was identical in the training phase. These patches were predicted by our classifier. To construct an IDC probability map, prediction scores between 0.0 to 1.0 were calculated from the Softmax function (Equation 2) of each patch. Then, a completed score map consisting of new patches of 32 × 32 pixels was created and initialized using its corresponding prediction scores. We aligned the scores of the predicted patches using the false color representation. The lower scores were mapped into cool tone colors, whereas the higher scores were mapped into intense tone colors. Finally, the score map was composed over the original slide image at the same coordinate as the corresponding patch. A Gaussian blur filter of three with bicubic interpolation was applied to obtain fine-smooth representations of the probability maps.

Figure 6 shows examples of IDC probability maps superimposed on original WSIs from the CINJ dataset. The rows in the figure display the original image, the corresponding IDC probability map, and the ground truth binary mask of the probability map. The IDC probability maps for WSIs A, B, and C are shown in D, E, and F, respectively. The colors in the probability maps represent the degree of probability, with dark blue indicating the lowest IDC probability and red indicating the highest.

In Figure 7, we can see examples of IDC probability maps superimposed on WSIs from the CWRU dataset. At the image patch level, there are more false positive results in these probability maps compared to those produced by the classification of the CINJ dataset. However, these probability maps are still suitable for slide-level predictions.

Slide-level predictions of BreakHis images are presented in Figure 8. Images A - D represent examples of four different types of malignant carcinoma tissues. Even at the middle levels of probability, characterized by green tones, the classifier could detect the location of malignant tissues. Images E - H represent examples of four different types of normal, benign tissues of the types used to test the negative prediction performance of the optimized classifier. There was no FP result from these images. Our classifier detected them correctly.

4. DISCUSSIONS

DL algorithms were tightly coupled with the training dataset. Model generalization toward other datasets, even though in the same domain, still requires improvement, especially for applying in real-world applications. Cross-validation to improve model generalization is a future challenge. In this work, we propose a practical, end-to-end approach to an ML pipeline for data analysis. Our methods do not require handcrafted feature extraction. The attributes of the training set selected for prediction were kept as close as possible to the characteristics of the original datasets. Qualitatively, the results of our IDC probability maps at the slide level were satisfactory, with a sensitivity of 0.88 and an AUC of 0.94.

However, in practice, the classifier's performance still depends on attributes of the dataset, such as the H&E staining, 40x magnification, RGB color mode, and PNG file format. We also found that the choice whether which file format, PNG or JPG, is used for patches significantly influences the model performance, especially when applying data augmentation.

A training technique of transfer learning utilizing pre-trained models with the ImageNet database [10] WSIs was also determined in this work. VGG16 [40] [41], VGG19 [40], Xception [42], and ResNet50 [43] [44] were used and configured using the same settings as the proposed model. Compared to our experiments, the proposed model still outperforms the classification performances of the pre-trained models, as shown in Table 7.

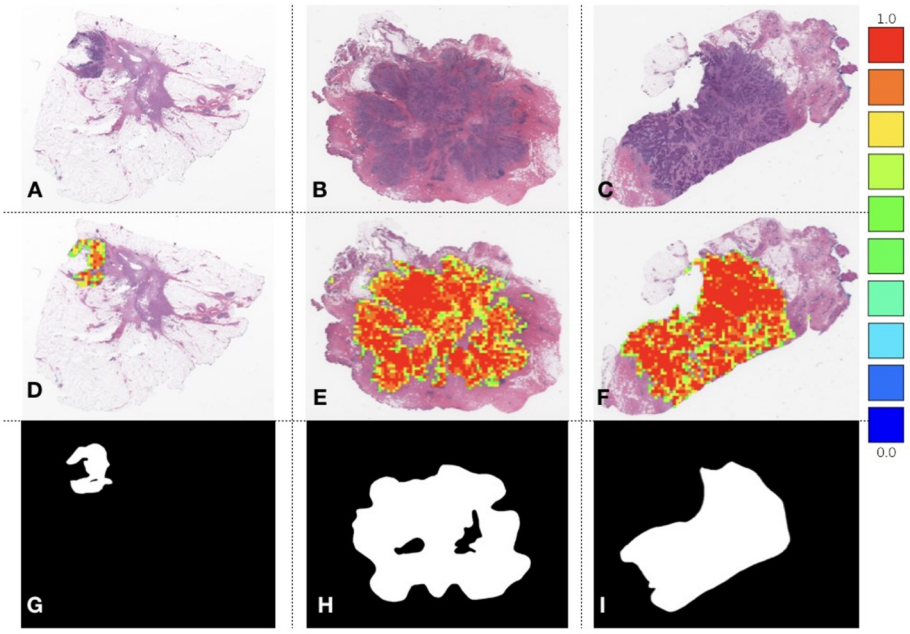


Fig.6: *(A-C) Examples of original BCa WSIs from the CINJ dataset, were previously unseen by our classifier. (D-F) Corresponding tumor probability maps of WSIs A, B, and C indicate probability by a false-color spectrum. Probability can be interpreted using the representative bar (red for the highest probability and dark blue for the lowest probability). The images (G-I) show corresponding ground truths of the BCa WSIs.*

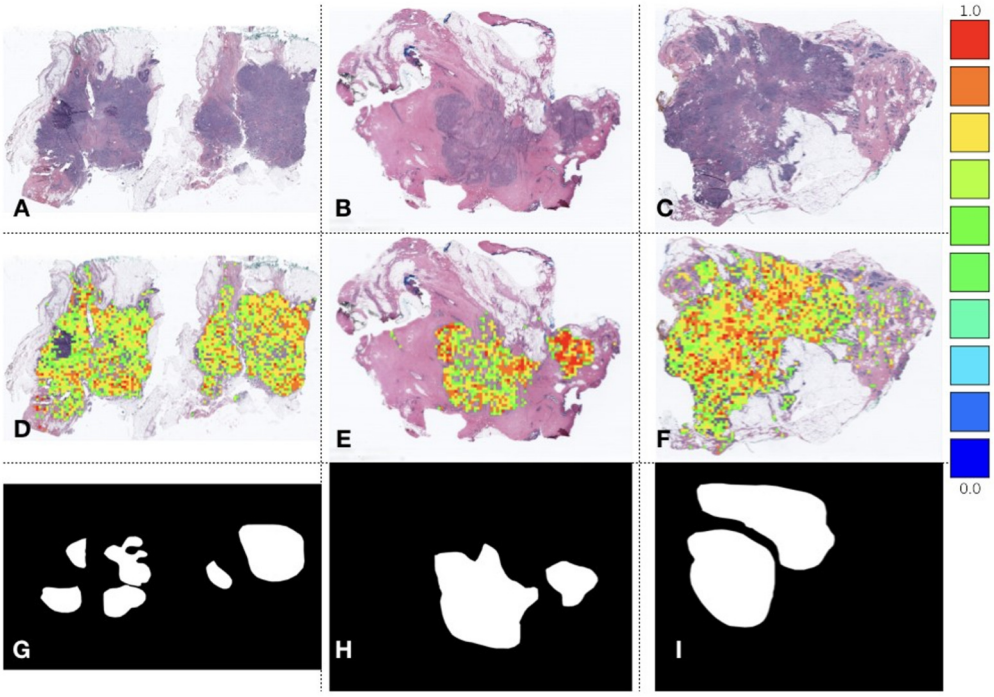


Fig.7: *(A-C) Examples of original BCa WSIs from the CWRU dataset, were previously unseen by our classifier. (D-F) Corresponding tumor probability maps of A, B, and C indicate the probability by false colors. Red represents the highest probability, and dark blue is the lowest probability. (G-I) Corresponding ground truth binary masks of the BCa WSIs.*

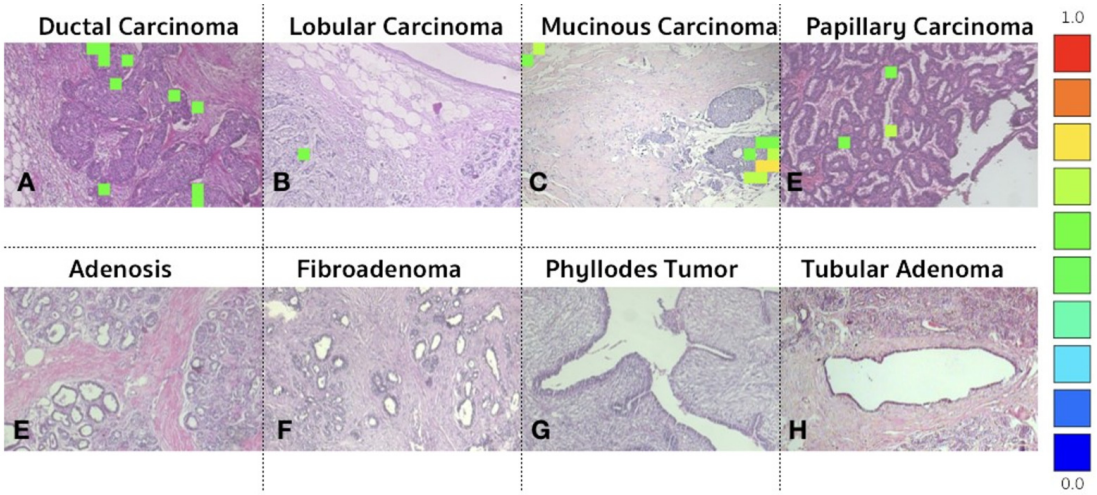


Fig.8: *Tumor probability maps of unseen BreakHis images. (A-E) Examples of four types of carcinoma tumor tissues are correctly predicted as positive. (E-H) Examples of four different types of negative tissues are correctly predicted as negative.*

Table 7: *Model performance evaluations comparing with the pre-trained models.*

Model	F1-score	Accuracy	Sensitivity	Specificity	FPR	FNR
VGG16	0.8050	0.8050	0.8140	0.7961	0.1950	0.1950
VGG19	0.7888	0.7889	0.7696	0.8082	0.2111	0.2111
Xception	0.8151	0.8151	0.8012	0.8290	0.1849	0.1849
ResNet50	0.8126	0.8127	0.7954	0.8300	0.1873	0.1873
ResNet101	0.7206	0.7212	0.7658	0.6766	0.2788	0.2788
Our model	**0.8802**	**0.8782**	**0.8907**	**0.8466**	**0.1313**	**0.1313**

Table 8: *Model performance evaluations compared with the related works on the same dataset.*

Model	F1-score	Balance Accuracy	Sensitivity	Specificity
Cruz-Roa et al. [22]	0.7180	0.8423	N/A	N/A
Janowczyk et al. [3]	0.7648	0.8468	N/A	N/A
Wang et al. [17]	0.8800	N/A	0.8560	0.8471
Present work	**0.8802**	**0.8687**	**0.8907**	**0.8466**

Our proposed classifier has been compared with other state-of-art classifiers [3] [17] [22] that utilized the same training dataset. Our proposed classifier outperformed in all significant measures, including F1-score, balance accuracy, sensitivity, and specificity, as shown in Table 8. Specifically, Cruz-Roa et al. [22] and Janowczyk et al. [3] reported 0.7180 and 0.7648 F1-scores, respectively, but our proposed classifier achieved a significantly higher F1-score of 0.8802. The achievement was similar to Wang et al. [17]. Nevertheless, the AUC of 0.94 in this work outperformed the AUC of 0.89 in [17]. In addition, our approach promotes more straightforward and explainable algorithms.

When considering executable times, InceptionResNetV2 consists of 572 layers and around 56 million parameters with a size of 215 MB [45]. Shahidi et al. [8] conducted their experiments using Inception-ResNet-V2 [16], ResNeXt-101 [11], and SENet-154 [13] on the Colab service Pro version with access to the GPU enabled (T4 or P100, and 27.4 GB of RAM). The training time required was 12 hours. Other models, e.g., NASNet [14] [15], DPN-131 [12], and ResNeXt-101(64×4d), required 24 hours for training.

The total number of trainable parameters of our model architecture was three hundred thousand. As a result, a local high-end workstation (with or without GPU enabled) can handle the model training on the user customized datasets. With our local GPU enabled (3,072 CUDA cores, 16 GB setting), our workstation required around six hours for training and one and a half minutes to create an IDC probability map at the slide-level prediction.

Used in this optimal configuration, we believe that our model has the potential for further performance improvements, such as boosting up the size of the dataset, modifying the model architecture such as enabling GRU [17], fine-tuning hyper-parameters, and real-world deployment.

5. CONCLUSIONS

Based on the feasibility of deploying our Deep Learning model for practical usage, we have proposed a simple and effective Deep Learning approach to breast cancer tissue classification. Invasive ductal carcinoma, the most common subtype of all breast cancers, was our priority for detection. The architectural design of the proposed model comprises just three hundred thousand total trainable parameters. It was quickly trained on a dataset consisting of around 280,000 image patches with the size of 50×50 pixels acquired from non-invasive and invasive ductal carcinoma tissues. During the training, each batch of images was pre-processed using several operations of data augmentation.

The model achieved an F1-score of 0.88. Three data sources were selected to evaluate the slide-level prediction. Probability maps of invasive ductal carcinomas were generated and superimposed over the whole-slide images. Even though some false positives were produced, qualitatively, the probability maps produced satisfactory results. Since we obtained a sensitivity of 90% for identifying of non-invasive ductal carcinomas, false negative prediction filtering at the slide level was acceptable.

The training and prediction configurations of our proposed approach are easy to deploy, particularly in practical applications, because the classifier only consists of 1.5 MB of trained weights. The flexible model can be easily installed and deployed on readily available, portable devices like smartphones and tablets or on regular PCs.

ACKNOWLEDGMENT

The original files for the training dataset are supported by the Center of Computational Imaging and Personalized Diagnostics (CCIPD) at Case Western Reserve University, Cleveland, Ohio, USA [3] [22] [49] and are publicly available on the Kaggle website [23]. The datasets of CINJ and CWRU are provided by The Cancer Genome Atlas on `http://cancergenome.nih.gov/` and the Genomic Data Commons (GDC) Data Portal at `https://portal.gdc.cancer.gov/`. The dataset of BreakHis was provided by the P&D Laboratory – Pathological Anatomy and Cytopathology, Parana, Brazil (`http://www.prevencaoediagnose.com.br`) and is publicly available on the Kaggle website [50].

IMPLEMENTATION NOTES

Keras with Tensorflow version 2.3.1 [46] backend was used. Keras is a Python high-level neural network API. Scikit-learn [47] was leveraged to evaluate the classification performance of the model. Scikit-image [48] was used to process tumor probability maps. The Jupyter Notebook of this work has been distributed and is publicly available at `https://www.kaggle.com/code/kasikrit/breast-cancer-tissue-detection`.

AUTHOR CONTRIBUTIONS

Conceptualization, K.D.; methodology, K.D.; software, K.D.; validation, K.D. and T.W.; formal analysis, K.D.; investigation, K.D.; data curation, K.D., C.K., P.S.; writing—original draft preparation, K.D; writing—review and editing, K.D. and T.W.; visualization, K.D.; supervision, K.D.; All authors have read and agreed to the published version of the manuscript.

References

[1] World Health Organization (WHO), "Breast cancer: prevention and control," *https://www.who.int/cancer/detection/breastcancer/en/*, January 2021.

[2] A. Janowczyk, "Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases," Case Western Reserve University, Technical Report, 2015. [Online]. Available: `http://www.andrewjanowczyk.com/deep-learning/`

[3] 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, no. 1, p. 29, Jun 2016.

[4] S. Deng, X. Zhang, W. Yan, E. I.-C. Chang, Y. Fan, M. Lai, and Y. Xu, "Deep learning in digital pathology image analysis: a survey," *Frontiers of Medicine*, vol. 14, no. 4, pp. 470–487, July 2020. [Online]. Available: `https://doi.org/10.1007/s11684-020-0782-9`

[5] J. Xu, X. Luo, G. Wang, H. Gilmore, and A. Madabhushi, "A Deep Convolutional Neural Network for segmenting and classifying epithelial and stromal regions in histopathological images," *Neurocomputing*, vol. 191, pp. 214 – 223, 2016. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0925231216001004`

[6] B. E. Bejnordi, M. Veta, P. J. van Diest et al., "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *JAMA*, vol. 318, no. 22, pp. 2199–2210, 2017. [Online]. Available: `http://dx.doi.org/10.1001/jama.2017.14585`

[7] C. Mercan, S. Aksoy, E. Mercan, L. G. Shapiro, D. L. Weaver, and J. G. Elmore, "From patch-level to ROI-level deep feature representations for breast histopathology classification," in *Medical Imaging 2019: Digital Pathology*, J. E. Tomaszewski and A. D. Ward, Eds., vol. 10956, International Society for Optics and Photon-

ics. SPIE, 2019, pp. 86–93. [Online]. Available: https://doi.org/10.1117/12.2510665

[8] F. Shahidi, S. Mohd Daud, H. Abas, N. A. Ahmad, and N. Maarop, "Breast Cancer Classification Using Deep Learning Approaches and Histopathology Image: A Comparison Study," in*IEEE Access*, vol. 8, pp. 187531–187552, 2020.

[9] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A Dataset for Breast Cancer Histopathological Image Classification," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 7, pp. 1455–1462, July 2016.

[10] S. U. Akram, T. Qaiser, S. Graham, J. Kannala, J. Heikkilä, and N. M. Rajpoot, "Leveraging unlabeled whole-slideimages for mitosis detection," *CoRR*, vol. abs/1807.11677, 2018. [Online]. Available: http://arxiv.org/abs/1807.11677

[11] S. Xie, R. B. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," *CoRR*, vol. abs/1611.05431, 2016. [Online]. Available: http://arxiv.org/abs/1611.05431

[12] Y. Chen, J. Li, H. Xiao, X. Jin, S. Yan, and J. Feng, "Dual path networks," *CoRR*, vol. abs/1707.01629, 2017. [Online]. Available: http://arxiv.org/abs/1707.01629

[13] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," *CoRR*, vol. abs/1709.01507, 2017. [Online]. Available: http://arxiv.org/abs/1709.01507

[14] B. Zoph and Q. V. Le, "Neural architecture search with reinforcement learning," *CoRR*, vol. abs/1611.01578, 2016. [Online]. Available: http://arxiv.org/abs/1611.01578

[15] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," *CoRR*, vol. abs/1707.07012, 2017. [Online]. Available: http://arxiv.org/abs/1707.07012

[16] C. Szegedy, S. Ioffe, and V. Vanhoucke, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning," *CoRR*, vol. abs/1602.07261, 2016. [Online]. Available: http://arxiv.org/abs/1602.07261

[17] X. Wang, I. Ahmad, D. Javeed, S. A. Zaidi, F. M. Alotaibi, M. E. Ghoneim, Y. I. Daradkeh, J. Asghar, and E. T. Eldin, "Intelligent Hybrid Deep Learning Model for Breast Cancer Detection," *Electronics*, vol. 11, no. 17, 2022. [Online]. Available: https://www.mdpi.com/2079-9292/11/17/2767

[18] S. Mojrian, G. Pintér, J. H. Joloudari, I. Felde, A. Szabo-Gali, L. Nádai, and A. Mosavi, "Hybrid Machine Learning Model of Extreme Learning Machine Radial basis function for Breast Cancer Detection and Diagnosis; a Multilayer Fuzzy Expert System," in *2020 RIVF International Conference on Computing and Communication Technologies, RIVF 2020, Ho Chi Minh City, Vietnam, October 14-15, 2020*. IEEE, 2020, pp. 1–7. [Online]. Available: https://doi.org/10.1109/RIVF48685.2020.9140744

[19] D. Dua and C. Graff, "UCI Machine Learning Repository," 2017. [Online]. Available: http://archive.ics.uci.edu/ml

[20] M. A. Jabbar, "Breast Cancer Data Classification Using Ensemble Machine Learning," *Engineering and Applied Science Research*, vol. 48, no. 1, pp. 65–72, Jan. 2021. [Online]. Available: https://ph01.tci-thaijo.org/index.php/easr/article/view/234959

[21] A. Bhardwaj, M. M. Khan, T. Tazin, M. Zunaid Hussain, M. Mostakim, T. Rehman, S. Singh, V. Gupta, and O. Alomeir, "Breast Tumor Detection Using Robust and Efficient Machine Learning and Convolutional Neural Network Approaches," *Computational Intelligence and Neuroscience*, vol. 2022, p. 6333573, 2022. [Online]. Available: https://doi.org/10.1155/2022/6333573

[22] A. Cruz-Roa, A. Basavanhally, F. Gonźalez et al., "Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks," in *SPIE Medical Imaging*, vol. 9041, 03/2014 2014, pp. 904103–904103–15. [Online]. Available: http://dx.doi.org/10.1117/12.2043872

[23] P. Mooney, "Breast Histopathology Images," https://www.kaggle.com/paultimothymooney/breast-histopathology-images, March 2020.

[24] A. Cruz-Roa, H. Gilmore, A. Basavanhally, M. Feldman, S. Ganesan, N. Shih, J. Tomaszewski, A. Madabhushi, and F. Gon

[25] A. Cruz-Roa, H. Gilmore, A. Basavanhally, M. Feldman, S. Ganesan, N. Shih, J. Tomaszewski, A. Madabhushi, and F. Gon

[26] S. Dabeer, M. M. Khan, and S. Islam, "Cancer diagnosis in histopathological image: CNN based approach," *Informatics in Medicine Unlocked*, vol. 16, p. 100231, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2352914819301133

[27] A. H. Ornek and M. Ceylan, "Comparison of Traditional Transformations for Data Augmentation in Deep Learning of Medical Thermography," in *2019 42nd International Conference on Telecommunications and Signal Processing (TSP)*, pp. 191–194, July 2019.

[28] M. Lin, Q. Chen, and S. Yan, "Network In Network," *CoRR*, vol. abs/1312.4400, 2013. [Online]. Available: http://arxiv.org/abs/1312.4400

[29] G. E. Dahl, T. N. Sainath, and G. E. Hinton,

"Improving deep neural networks for LVCSR using rectified linear units and Dropout," *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. Vancouver, BC, Canada, pp. 8609–8613, 2013.

[30] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," *CoRR*, vol. abs/1311.2901, 2013. [Online]. Available: http://arxiv.org/abs/1311.2901

[31] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," in *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, ser. ICML'15. JMLR.org, 2015, pp. 448–456. [Online]. Available: http://dl.acm.org/citation.cfm?id=3045118.3045167

[32] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014. [Online]. Available: http://jmlr.org/papers/v15/srivastava14a.html

[33] A. Araujo, W. Norris, and J. Sim, "Computing Receptive Fields of Convolutional Neural Networks," *Distill*, 2019, https://distill.pub/2019/computing-receptivefields.

[34] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *Journal of Machine Learning Research*, vol. 12, no. Jul, pp. 2121–2159, 2011.

[35] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," CoRR, vol. abs/1412.6980, 2014. [Online]. Available: http://arxiv.org/abs/1412.6980

[36] I. Sutskever, J. Martens, G. Dahl, and F. Hinton, "On the importance of initialization and momentum in deep learning," in Proceedings of the 30th International Conference on Machine Learning, ser. Proceedings of Machine Learning Research, S. Dasgupta and D. McAllester, Eds., vol. 28, no. 3. Atlanta, Georgia, USA: PMLR, 17–19 Jun 2013, pp. 1139–1147. [Online]. Available: https://proceedings.mlr.press/v28/sutskever13.html

[37] J. Han, M. Kamber, and J. Pei, *Data mining: Concepts and Techniques*, 3rd ed. MA, USA: Morgan Kaufmann, 2011.

[38] T. Saito and M. Rehmsmeier, "The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets," *PLOS ONE*, vol. 10, no. 3, pp. 1–21, 03 2015. [Online]. Available: https://doi.org/10.1371/journal.pone.0118432

[39] G. Lemaitre, F. Nogueira, and C. K. Aridas, "Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning," *Journal of Machine Learning Research*, vol. 18, no. 17, pp. 1– 5, 2017. [Online]. Available: http://jmlr.org/papers/v18/16-365.html

[40] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *CoRR*, vol. abs/1409.1556, 2014. [Online]. Available: http://arxiv.org/abs/1409.1556

[41] K. Damkliang, T. Wongsirichot, and P. Thongsuksai, "TISSUE CLASSIFICATION FOR COLORECTAL CANCER UTILIZING TECHNIQUES OF DEEP LEARNING AND MACHINE LEARNING," *Biomedical Engineering: Applications, Basis and Communications*, vol. 33, no. 03, p. 2150022, 2021. [Online]. Available: https://doi.org/10.4015/S1016237221500228

[42] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," *CoRR*, vol. abs/1610.02357, 2016. [Online]. Available: http://arxiv.org/abs/1610.02357

[43] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *CoRR*, vol. abs/1512.03385, 2015. [Online]. Available: http://arxiv.org/abs/1512.03385

[44] L. J. Herrera, F. Ahmad, A. Farooq, and M. U. Ghani, "Deep ensemble model for classification of novel coronavirus in chest x-ray images," *Computational Intelligence and Neuroscience*, vol. 2021, p. 8890226, 2021. [Online]. Available: https://doi.org/10.1155/2021/8890226

[45] Keras, "Keras Applications," https://keras.io/applications/, 2022, [Online; accessed 14-December-2022].

[46] F. Chollet et al., "Keras," https://keras.io, 2020, [Online; accessed 14-December-2022]. [Online]. Available: https://keras.io/

[47] F. Pedregosa, G. Varoquaux, A. Gramfort et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, no. Oct, pp. 2825–2830, 2011.

[48] S. van der Walt, J. L. Schonberger, J. NunezIglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, T. Yu, and the scikit-image contributors, "scikit-image: image processing in Python," *PeerJ*, vol. 2, p. e453, 6 2014. [Online]. Available: https://doi.org/10.7717/peerj.453

[49] (2022, 12) The Center of Computational Imaging and Personalized Diagnostics (CCIPD). [Online]. Available: https://engineering.case.edu/centers/ccipd/

[50] P&D Laboratory – Pathological Anatomy and Cytopathology, "Breast Cancer Histopathological Database (BreakHis)," https:

//www.kaggle.com/ambarish/breakhis, January 2021.

Kasikrit Damkliang received a BS in Computer Science in 2005, an MEng in Computer Engineering in 2009, and a PhD in Computer Engineering in 2019 from Prince of Songkla University (PSU) in Thailand. He is currently an Assistant Professor at the Division of Computational Science in the Faculty of Science at PSU. His research interests include medical image analysis, deep learning and machine learning, bioinformatics, web service, cloud computing, and workflow technology.

Thakerng Wongsirichot is currently an Assistant Professor at the Division of Computational Science, Faulty of Science, Prince of Songkla University (PSU), Thailand. He received his PhD in Computer Science from King Mongkut's Institute of Technology Ladkrabang, and Master of Information Systems and B. Com (E-Commerce and Business Information Systems) from University of Wollongong, Australia. His research interests include deep learning, machine learning, and health informatics.

Chawisa Khongrak received a BS in Information and Communication Technology in 2020 from Prince of Songkla University. Currently she is a Software Tester at Kasikorn Business-Technology Group Company. Her research interests include medical image analysis and AI.

Piyathida Suwannarat received a BS in Information and Communication Technology in 2020 from Prince of Songkla University. She is currently a Software Engineer at Freewill Solution Co., Ltd. Her research interests include medical image analysis and AI.