Residual Learning based CNN for Breast Cancer Histopathological Image Classification

Mahesh Gour [1*], Sweta Jain [2], T. Sunil Kumar [3]

[*] *corresponding author: maheshgour0704@gmail.com*
[1,2] *Maulana Azad National Institute of Technology, Bhopal, MP, India*
[3] *Norwegian University of Science and Technology, Trondheim, Norway*

Abstract

The biopsy is one of the most commonly used modality to identify breast cancer in women, where tissue is removed and studied by the pathologist under the microscope to look for abnormalities in tissue. This technique can be time-consuming, error-prone, and provides variable results depending on the expertise level of the pathologist. An automated and efficient approach not only aids in the diagnosis of breast cancer but also reduces human effort. In this paper, we develop an automated approach for the diagnosis of breast cancer tumors using histopathological images. In the proposed approach, we design a residual learning-based 152-layered convolutional Neural Network, named as ResHist for breast cancer histopathological image classification. ResHist model learns rich and discriminative features from the histopathological images and classifies histopathological images into benign and malignant classes. In addition, to enhance the performance of the developed model, we design a data augmentation technique, which is based on stain normalization, image patches generation, and affine transformation. The performance of the proposed approach is evaluated on publicly available BreaKHis dataset. The proposed ResHist model achieves an accuracy of 84.34% and an F1-score of 90.49% for the classification of histopathological images, and also, this approach achieves an accuracy of 92.52% and F1-score of 93.45% when data augmentation is employed. The proposed approach outperforms the existing methodologies in the classification of benign and malignant histopathological images. Furthermore, our experimental

results demonstrate the superiority of our approach over the pre-trained networks, namely AlexNet, VGG16, VGG19, GoogleNet, InceptionV3, ResNet50, and ResNet152, for the classification of histopathological images.
Keywords: Breast Cancer, Histopathological Image, Data Augmentation, Residual Learning, Convolutional Neural Network, Deep Features.

---

1. INTRODUCTION

Breast cancer is the second leading cause of cancer-related death in the women, across the globe [1]. International agency for research on cancer (IARC), reported that approximately 8.2 million deaths were caused by cancer in the year 2012, and about 27 million new cases of cancer are expected by 2030 [2]. In general, breast cancer is diagnosed by testing of medical images (e.g., breast mammograms, ultrasound, and MRI images) and biopsy. In biopsy examination, tissues are studied under the microscope by the pathologists. Generally, cancerous cells are examined visually, based on shape and size, malignancy degree, tissue distribution, etc. [3]. If cancerous cells are present, then the biopsy is the only guaranteed way for diagnosis. The visual examination of the cancerous cell is time consuming and requires an expert pathologist.

Recent advancements in machine learning and image processing have enabled the development of computer-aided diagnosis (CAD) systems for detecting and diagnosing breast cancer from the histopathological images faster with very high accuracy. The CAD system analyzes the histopathological images of the sample tissue, and finds the histopathological patterns corresponding to the cancerous and non-cancerous condition and classifies the histopathological images respectively into benign and malignant class. The major challenges associated with the classification of breast cancer histopathological images include the inherent complexity in histopathological images such as cell overlapping, subtle differences between images and uneven color distribution.

The objective of this study is to develop an accurate and reliable solution for breast cancer classification. In this study, we have systematically investigate

the deep learning based approaches for automatic diagnosis of breast cancer. Key highlights of this work are as following:

- We develop a deep residual convolutional neural network (ResHist) for breast cancer diagnosis from the histopathological images.
- We propose a data augmentation technique based on stain normalization, image patches generation and affine transformation.
- We also investigate the performances of deep features (extracted from ResHist) with various classifiers.

This paper is organized as follows: Section 2 presents the literature review. Section 3 describes the proposed ResHist based approach for classification of benign and malignant histopathological images. Section 4 describes the dataset and proposed data augmentation method. Section 5 presents the evaluation metrics and also details the experimental results. Finally, the conclusion is drawn in Section 6.

2. LITERATURE REVIEW

The computer-aided diagnosis of cancer has been a topic of research since the last 40 years, but it is still challenging owing to the inherent complexities of histopathological images [4]. Several methodologies have been proposed in the literature for the classification of histopathological images into benign and malignant classes. This section briefly reviews prominent research works related to the classification of histopathological images.

Kowal et al. [5] have employed various nuclei segmentation algorithms such as K-means, competitive neural networks, Gaussian mixture model (GMM) and fuzzy C-means (FCM) to extract the region of interest (ROI) in order to classify benign and malignant images. Authors reported accuracy rates from 96% to 100%. George et al. [6] proposed nuclei segmentation based breast cancer diagnosis system using machine learning techniques. They have detected nuclei using Hough transform with Ostu's thresholding and FCM clustering.

They have applied four classifiers such as multilayer perceptron, learning vector quantization, support vector machine (SVM) and probabilistic neural network (PNN) on the extracted features. Authors have reported classification accuracies ranging between 76% and 94% on the dataset consisting of 92 cytological images.

Spanhol et al. [7] have investigated the performances of image descriptors such as local binary pattern (LBP), completed local binary pattern (CLBP), gray-level co-occurrence matrices (GLCM), parameter-free threshold adjacency statistics (PFTAS), local phase quantization (LPQ) and oriented fast and rotated brief (ORB) for diagnosis of breast cancer. They have investigated the performance of local descriptors with four classifiers, namely 1-nearest neighbor (1-NN), random forest (RF), quadratic discriminant analysis (QDA) and SVM. Authors have employed the approach as mentioned above on a dataset containing 7909 histopathological images of 82 patients (BreaKHis dataset). The classification accuracies of this approach are ranging between 80% and 85%.

Spanhol et al. [8] have proposed deep learning based methodology using fine-tuned pre-trained LeNet architecture for breast cancer tumor classification. This methodology has archived a classification accuracy of 72%. Further, they used a more sophisticated variant of AlexNet architecture, which consists of five learnable layers (three convolutional and two fully connected). In addition, the authors reported classification accuracy rates ranging from 80.8% to 85.6% using AlexNet architecture.

Bayramoglu et al.[9] have proposed a magnification independent approach in which they have developed two convolutional neural network (CNN) models named as single task CNN and multi-task CNN. Single task CNN is used to estimate malignancy, and multi-task CNN is used to estimate malignancy as well as magnification level of the image. Single task CNN-based strategy achieved accuracies ranging from 82.1% to 83.0% and multi-task CNN-based strategy achieved accuracies ranging from 80.6% to 83.3%.

Spanhol et al. [10] used a pre-trained CaffeNet model for extracting deep features from the histopathological images. Features are extracted at the fully

Table 1: Summary of literature review.

Author(s)	**Methodology**	**Dataset**	**Accuracy (in %)**
Kowal et al. [5]	• Nuclei segmentation: using K-means, FCM, competitive-NN and GMM. • Classification: K-NN, Naïve bayes and decision tree	Cytological images	96-100%
George et al. [6]	• Nuclei segmentation: using Hough transform, Ostu's thresolding and FCM clustering • Classification: multilayer perceptron, PNN, learning vector quantization and SVM	Cytological images.	76-94%
Spanhol et al. [7]	• Feature Extraction: using LBP, CLBP, LPQ, GLCM, PFTAS and ORB • Classification: 1-NN, QDA, SVM, and RF	BreaKHis	80-85%
Spanhol et al. [8]	• Pre-trained CNN Models: LeNet and AlexNet CNN	BreaKHis	80.8-85.6%.
Bayramoglu et al. [9]	• Proposed two CNN models: - Single task CNN: for predicting the malignancy. -Multi-task CNN: for predicting malignancy and image magnification levels.	BreaKHis	80.6-83.3%.
Spanhol et al. [10]	•Extracted deep features (DeCAF): using pre-trained CaffeNet.	BreaKHis	83.6-84.8%.
Sudharshan et al. [11]	• Proposed MIL based approach: APR, KNN, DD, SVM, non-parametric algorithm and MIL-CNN	BreaKHis	83.4-92.1%.

connected (fc) layers fc6, fc7, and fc8 of CaffeNet. They achieved accuracies between 83.6% and 84.8%.

Recently, Sudharshan et al. [11] used multiple instance learning (MIL) for histopathological image analysis. In which they have studied the performances of various MIL techniques such as seminal Axis Parallel Rectangle algorithm (APR), K-nearest neighbor (KNN), algorithms based on diversity density

(DD), SVM, non-parametric algorithm and convolutional neural network for MIL. They reported binary classification accuracies between 83.4% and 92.1%. Summary of literature review has been presented in Table 1.

3. METHODOLOGY

Deep learning is a data-driven learning approach that learns the features, semantic rules and meaningful information directly from the data itself and has found to be very effective for various task including object detection [12], voice recognition [13], signal recognition [14], visual tracking [15] and image classification [16, 17, 18]. The CNN is a part of a deep learning family, and it is widely used in medical image analysis such as MRI image segmentation [19], brain tumor detection in MRI images [20], retinal lesion detection in fundus images [21] and nuclei segmentation in histopathological images [22]. The effectiveness of CNN has motivated us to explore CNN based methodology for breast cancer image classification. Figure 1 depicts the schematic diagram of the proposed methodology. In this work, we have developed an improved residual learning-based convolutional neural network named as ResHist (**Res**idual convolutional neural network for the **Hist**opathological image) for breast cancer histopathological image classification. A detailed description of the proposed approach is given in the following section:

3.1. Proposed Approach

ResHist is a non-linear feature representation learning model, in which features extraction process is replaced by the features learning. This model has a strong ability to learn the discriminative features, semantics, features space distance constraint of inter-class from the complex breast cancer histopathological image at multiple levels of abstraction. The description of proposed ResHist architecture is shown in Table 2. The proposed ResHist model is inspired by ResNet152 [17]. The proposed network consists of 152 layers, with 13 residual blocks, in which 46 layers are learnable (45 Convolutional layers and 1 Fully connected). The residual block which has skip connections to jump over the layer

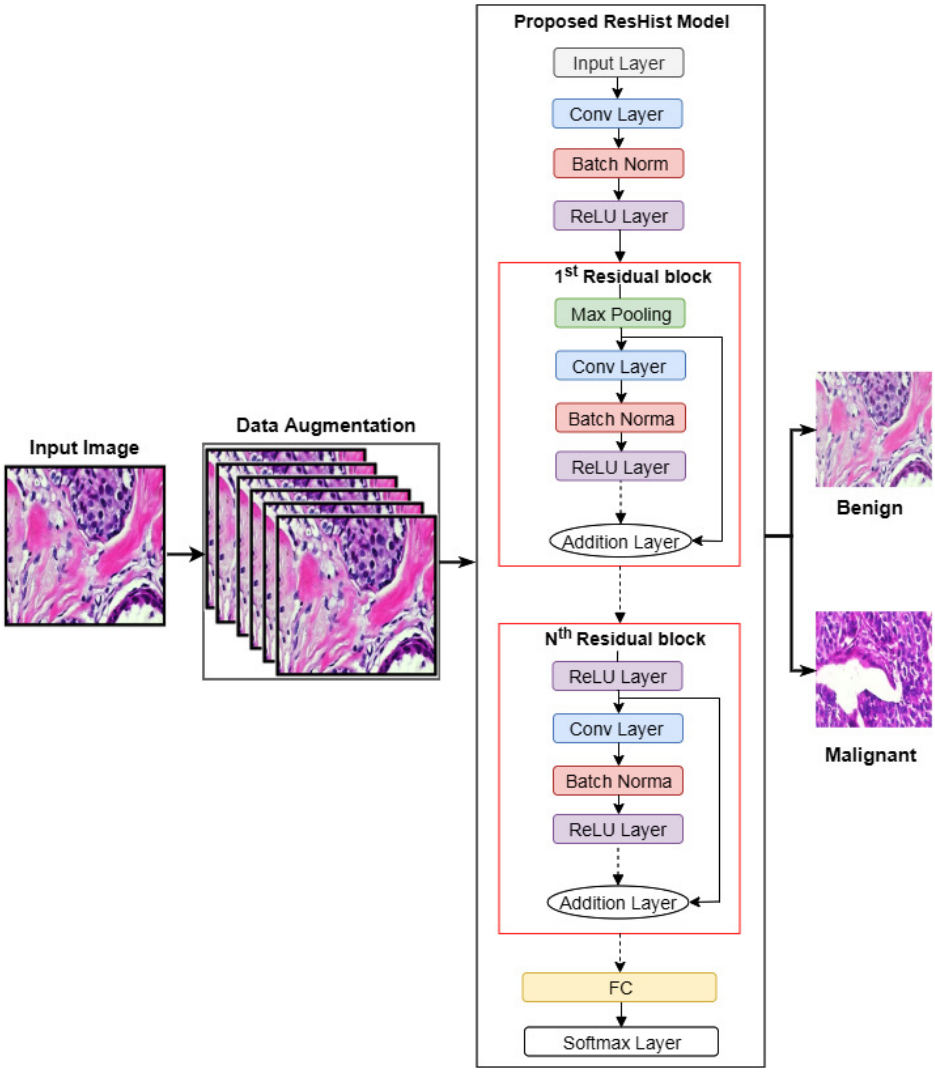


Figure 1: Block diagram of proposed method.

and it prevents the network from the vanishing gradient problem. We added residual blocks repetitively in the network because it increases representation power, leads to faster convergence and minimizes training errors as compared to stacked CNN [17].

In ResHist model, we have used the Input layer, which holds raw pixel values of input histopathological image. Followed by Convolutional layer have

Table 2: Layers of ResHist Model and parameters of the layers.

Layer Name	**Type**	**Parameter/ filter Size**
Input Layer	Input layer	[224 224 3]
Conv Layer	Convolutional Layer	[1 1], [3 3], [7 7], stride=[1 1], padding same
Batch Norm	Batch Normalization Layer	64, 128, 256, 512, 1024 channels
ReLU Layer	ReLU Layer	max(0, x)
Max Pooling	Max Pooling Layer	[3 3], stride = [2 2]
Avg Pooling	Average Pooling Layer	[7 7], stride = [7 7]
FC	Fully Connected Layer	2
Softmax Layer	Softmax Layer	Softmax function, Cross-entropy loss

applied, which is responsible for detecting and learning local features from the input image. Convolutional layer produces neurons output by computing the dot product between the filters of the Convolutional layer and connected local region of the Input layer.

We have used Batch normalization layer to increase the stability of the network and speed up the learning process. Batch normalization layer reduces the internal covariate shift and it also prevents the network from the overfitting by introducing the regularization effect in the network [23]. Following steps have been used in Batch normalization:

- Computes mini batch mean (μ_{bn}) and variance (σ_{bn}) of the activation layer using Eqn. (1) and Eqn. (2) respectively.

$$\mu_{bn} = \frac{1}{m}\sum_{i=1}^{m} x_j \tag{1}$$

$$\sigma_{bn} = \sqrt{\frac{1}{m}\sum_{j=1}^{m}(x_j - \mu_{bn})^2} \tag{2}$$

- Performs normalization on the input (x_j) of activation layer (in Eqn. (3)).

$$\overline{X}_j = \frac{(x_j - \mu_{\text{bn}})}{\sqrt{\sigma_{bn}^2 + \epsilon}} \tag{3}$$

- Scale and shift the normalized data ($\overline{X}_j$) using Eqn. (4), in order to produce the output of the Batch normalization layer.

$$Y_j = \gamma \overline{X}_j + \beta \tag{4}$$

Where m represents the size of the mini batch, ϵ is a constant, γ and β are learnable parameters which are initialized randomly and learned during training. After Batch normalization layer, we have employed ReLU activation function, as given in Eqn. (5), which provides a mapping between output to the set of inputs and introduces non-linearity in the network structure [24].

$$f(x) = \begin{cases} x & if\ x > 0 \\ 0 & otherwise \end{cases} \tag{5}$$

We have used Max pooling layer as 5^{th} layer and Average pooling layer as 143^{th} layered, to reduce the spatial dimension of feature maps. Pooling layer independently operates on each feature map and minimizes computation and amount of parameters in the network. We have employed a Softmax layer, which is responsible for generating the desired outputs. Hence layer 151 uses softmax function to predict the probability of class (benign or malignant), in which the input image belongs to. Softmax function is a logistic regression for binary classification that takes classes score $S = (s_1, s_2)$ as input and produces probability distributions $P = (p_1, p_2)$ corresponding to each class. Softmax function $f(S, \theta)$ is defined mathematically as in Eqn. (6).

$$P = f(S, \theta) = \frac{1}{\sum_{j=1}^{2} e^{\theta^T s_j}} \begin{pmatrix} e^{\theta^T s_1} \\ e^{\theta^T s_2} \end{pmatrix} \tag{6}$$

Where θ is the softmax classifier parameter. At the output layer, cross-entropy loss function as in Eqn. (7) have been applied to measure the performance of the network and to compute the classification error. During the training of

network, the classification error is minimized by using the ADAM optimization algorithm [25]. Let T $\in \{0, 1\}$ is target output and P is predicted probability. For N training sample of histopathological images, the loss function is defined as:

$$J(T, P) = -\sum_{i=1}^{N} J\left(t_i, p_i\right) = -\sum_{i=1}^{N}\sum_{c=1}^{2} t_{ic} \log\left(p_{ic}\right) \tag{7}$$

Where t_{ic} is 1, if and only if sample image i belongs to class c and p_{ic} is a predicted probability of sample i belonging to class c. L2 regularization term is also added to loss function as shown in Eqn. (8), which prevents network from the overfitting problem by penalizing network weight matrix of the nodes.

$$J(T, P) = -\sum_{i=1}^{N}\sum_{c=1}^{2} t_{ic} \log\left(p_{ic}\right) + \frac{\lambda}{2N} \sum \|w\|^2 \tag{8}$$

Where λ is a regularization parameter and it is initialized to λ=0.0005.

The weights of the first 40 convolutional layers of the ResHist are initialized using transfer learning from the pre-trained ResNet50 network and remaining layer's weights are initialized by the Gaussian distribution with a standard deviation of 0.01. The learning rate is set to 0.0001, and mini-batch size is set to 128. We experimentally found out that these hyperparameter values are the best suitable for our problem.

In the proposed approach, histopathological images are first augmented then ResHist model is trained end-to-end on the augmented dataset in a supervised learning manner. During training, ResHist model learns the discriminative features and mapping rule that maps the input histopathological images to output labels. In the test phase, for a given unlabelled test image, the trained ResHist model produces the probability of the presence of cancerous tumor in the histopathological image and classifies the image into benign or malignant class.

4. DATASET

In this section, we briefly explain the BreaKHis dataset. Further, this section details the proposed data augmentation technique.

4.1. BreaKHis Dataset

In order to evaluate the performance of the ResHist model, publicly available BreaKHis dataset [7] has been employed. This dataset includes 7909 breast cancer histopathological images of 82 patients. Out of 7909 images, 2480 images belong to a benign class and 5429 images belong to a malignant class. These images are acquired from the biopsy slides of breast tissue in four magnification factors. Figure 2 shows the sample images of benign and malignant classes in 40X, 100X, 200X and 400X magnification factors from the BreaKHis dataset. The images in dataset are the colored (RGB) images of size 700×460 pixels. Table 3 gives the details of the distribution of images in the benign and malignant classes corresponding to each magnification factor.

4.2. Data Augmentation

It is well known that the performance of deep learning models heavily depends on the size of the training dataset available. Therefore data augmentation

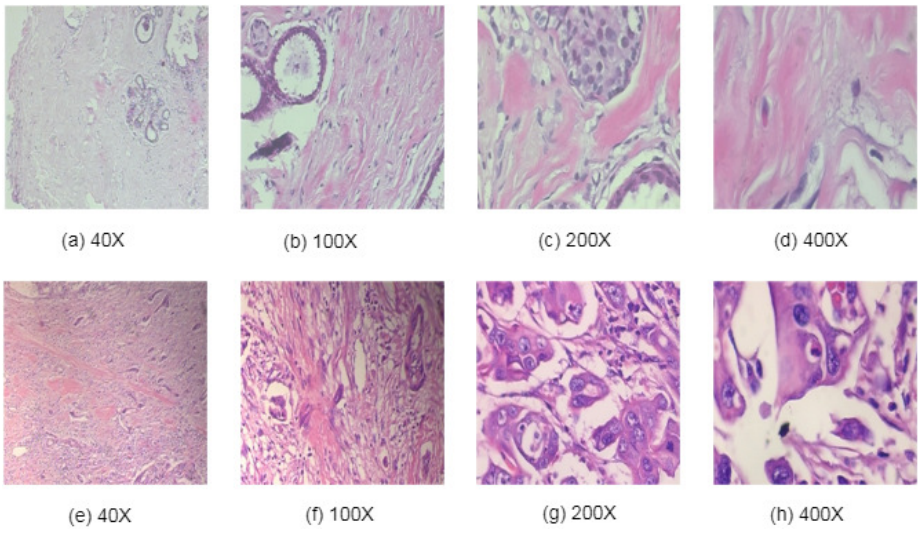


Figure 2: Samples of histopathological images of breast cancer from the BreaKHis database; the top row shows the benign tumor, and the bottom row shows the malignant tumor. From left to right (a, e), (b, f), (c, g) and (d, h) are images in 40X, 100X, 200X, and 400X magnifications factors, respectively.

Table 3: Images distribution in benign and malignant classes corresponding to magnification factor.

Class	Magnification Factors				Number of
	40X	100X	200X	400X	Patient
Benign	625	644	623	588	24
Malignant	1,370	1,437	1,390	1,232	58
Total	1,995	2,081	2,013	1,820	82

techniques play an important role in increasing dataset cardinality [26, 27]. Data augmentation techniques improve the performance of the network by overcoming the problem of the network overfitting. In this work, we have proposed a data augmentation method, in which we use affine transformation, stain normalization [28] and image patches generation algorithm for data augmentation. This data augmentation approach increases the dataset size ten times. The data augmentation algorithm is given as follows:

Algorithm : Data augmentation of histopathological images.

Input: H&E stain breast cancer histopathological image I_k.

Output: Augmented images ($I_{k1}....I_{k10}$) of the image I_k.

Step 1: Take histopathological image I_k from the training set.

Step 2: Apply affine transformations on image I_k:

(i) perform image rotation operation

$$I_{k1} = Rotation(I_k, 90^0)$$

$$I_{k2} = Rotation(I_k, 180^0)$$

(ii) perform flip operation

$$I_{k3} = Flip(I_k)$$

Step 3: Perform stain normalization operations on histopathological image I_k using target image I_t .

$$I_{k4} = StainNorm(I_k, I_t)$$

Step 4: Apply patch generation technique on image I_k.

$$[I_{k5}, ..., I_{k10}] = PatchGen(I_k)$$

**Where Rotation(), PatchGen(), StainNorm() are represents the methods of image rotation, patch generation and stain normalization respectively.*

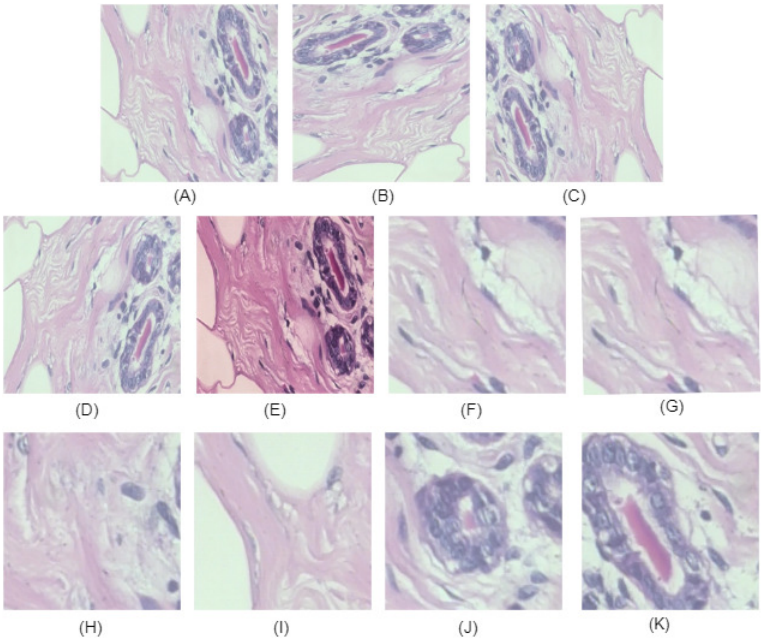


Figure 3: Histopathological image data augmentation. Where images represent in: (A) is original image; (B,C) are the rotated images at 90^0 and 180^0 respectively; (D) is the flipped image; (E) is the stain normalized image; and (F, G, H, I, J, K) are the image pages.

Figure 3 shows augmented images produced while applying the proposed data augmentation method. The proposed data augmentation approach can be divided into three steps. In the first step, affine transformations have been applied on the histopathological image (I_k), where the image has been rotated at the angle of 90^0 and 180^0 and the image pixels have been flipped vertically. This step produces three images (I_{k1}, I_{k2}, I_{k3}). In the second step, stain normalization has been applied to the image (I_k) to obtain stain normalized image (I_{k4}). In the stain normalization process, the color of an input image is mapped

to the target image using linear transformation in a perceptual colorspace. In the final step, a sliding window strategy has been applied on the image (I_k) to generate six image patches ($I_{k5}, I_{k6}, I_{k7}, I_{k8}, I_{k9}, I_{k10}$).

5. EXPERIMENTS AND DISCUSSION

The experimental setup used for the evaluation of the proposed method is the same as in that of [7]. Where the dataset has been randomly divided into the training set and test set in the ratio of 0.7 and 0.3 at the patient level (patients used to build the test set have not been used for training set) for each magnification factor. We have further divided the training set into a train set and validation set in the ratio of 0.8 and 0.2 and applied data augmentation approach on a train set. The augmented train set is used for network training and validation set is used for selecting the best hyperparameter for the model. The test set is used for performance evaluation of the proposed method. The results reported in this paper are the average of 5 trials. All the experiments are performed on an NVIDIA Quadro K5200 GPU with CUDA 10.0 using cuDNN.

5.1. Evaluation Metrics

To assess the performance of proposed method we have used precision, recall, F1-score, accuracy and confusion matrix as evaluation metrics. There are two types of recognition accuracies as in [7, 8, 9, 10]. (i) Accuracy at patient level: let N_{ntp} be total number of images of the patient P, N_{pc} be number of images of the patient P that are the correctly classified and N_p be the number of patient. It is defined as (in Eq. (9) and Eqn. (10)):

$$Patient\ Score = \frac{N_{pc}}{N_{ntp}} \tag{9}$$

$$Patient\ Recognition\ Accuracy = \frac{\sum Patient\ Score}{N_p} \tag{10}$$

(ii) Accuracy at image level: let N_{all} be total number of the histopathological images of the test set and N_r number of images that are correctly classified.

Image level accuracy is defined in the Eqn. (11).

$$Image \;\; Recognition \;\; Accuracy = \frac{N_r}{N_{all}} \tag{11}$$

Precision, recall and F1-score are defined as (in Eqn. (12), Eqn. (13) and Eqn. (14) respectively) :

$$Precision = \frac{TP}{(TP + FP)} \tag{12}$$

$$Recall = \frac{TP}{(TP + FN)} \tag{13}$$

Where TP represents true positive, which refers the correctly classified malignant class images. FN represents false negative, which refers to the incorrectly classified benign class images. Similarly, FP represents false positive, which refers to the incorrectly classified malignant class images. F1-score is mainly used for measuring test accuracy and it is computed by calculating the harmonic mean of the precision and recall.

$$F1 - Score = \frac{2TP}{(2TP + FP + FN)} \tag{14}$$

5.2. Experimental Results

This section presents the experimental results of the proposed approach evaluated on original BreaKHis dataset and dataset with augmentation. Section 5.2.1 presents the performance of ResHist model, where Section 5.2.2, shows the performance of deep features with different classifiers. Additionally, we evaluate the performance of the pre-trained networks on histopathological images in Section 5.2.3.

5.2.1. ResHist Model Results

ResHist model is a task-specific network for breast cancer images classification. Two sets of experiments have been performed for network training. In the first set of experiments, we have trained ResHist model on the original BreaKHis dataset. In the second set of experiments, we have performed data augmentation on the training set of BreaKHis dataset. Thus obtained augmented data

Table 4: Mean recognition accuracy of the proposed ResHist model at the image level and patient level, with respective standard deviation.

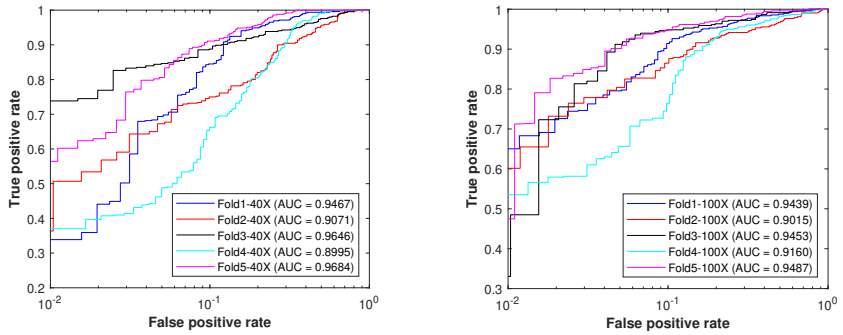
Accuracy @ in %	**Method (s)**	**Magnification Factors**			
		40X	**100X**	**200X**	**400X**
Image Level	ResHist	82.12±4.38	82.98±6.15	80.85±8.03	81.83±5.17
	ResHist-Aug	87.40±3.00	87.26±3.54	**91.15±2.30**	86.27±2.18
Patient Level	ResHist	83.24±4.18	84.34±5.72	81.65±10.88	83.00±4.20
	ResHist-Aug	87.47±3.22	88.15±2.97	**92.52±2.84**	87.78±2.46

"ResHist-Aug": ResHist model is trained on augmented BreaKHis dataset

have been used to train ResHist model. The results obtained in these sets of the experiments are provided in Table 4. From the experimental results, it can be noticed that there is a significant improvement in the performance of ResHist when data augmentation is employed. The best accuracy of 92.52% has been obtained at the patient level for 200X magnification factor. On the other hand, in Table 5, we have presented the assessment of ResHist model based on evaluation metrics like specificity, precision, recall, and F1-score. Thus, we have observed that the overall ResHist-Aug (with Data augmentation) model performed better for all magnification factors. We further analyze ResHist-Aug performance by drawing receiver operating characteristic (ROC) curves for various experimental

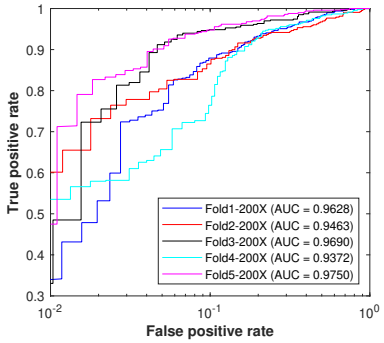
Table 5: Specificity, recall, precision and F1-score computed from the ResHist models.

Magnification Factors	**Method (s)**	**Specificity (in %)**	**Sensitivity / Recall (in %)**	**Precision (in %)**	**F1-Score (in %)**
40X	ResHist	87.30±0.05	86.39±0.03	95.07±0.02	90.49±0.02
	ResHist-Aug	85.97±0.08	87.99±0.03	94.16±0.03	90.94±0.02
100X	ResHist	80.89±0.09	86.98±0.03	91.59±0.04	89.20±0.03
	ResHist-Aug	83.31±0.10	89.54±0.03	91.87±0.06	90.58±0.03
200X	ResHist	75.02±0.14	90.80±0.03	85.25±0.13	87.57±0.08
	ResHist-Aug	85.62±0.05	93.69±0.02	93.23±0.02	**93.45±0.02**
400X	ResHist	78.86±0.10	88.53±0.03	88.59±0.07	88.38±0.03
	ResHist- Aug	81.62±0.07	88.91±0.03	90.83±0.05	89.75±0.02

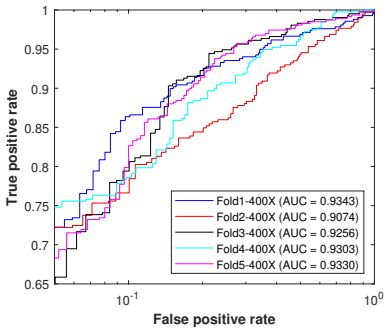


(a) 40X

(b) 100X



(c) 200X



(d) 400X

Figure 4: ROC curves of ResHist for various trials on BreaKHis dataset with data augmentation.

trial corresponding to their respective magnification factors (see Figure 4).

Based on the ROC curves, the developed model is observed to be stable for varied experimental trials and magnification factors. Confusion matrices for the same have been shown in Table 6. In confusion matrices, we can see that the proposed model produces very less false positive, which confirms that the ResHist is reliable for clinical uses.

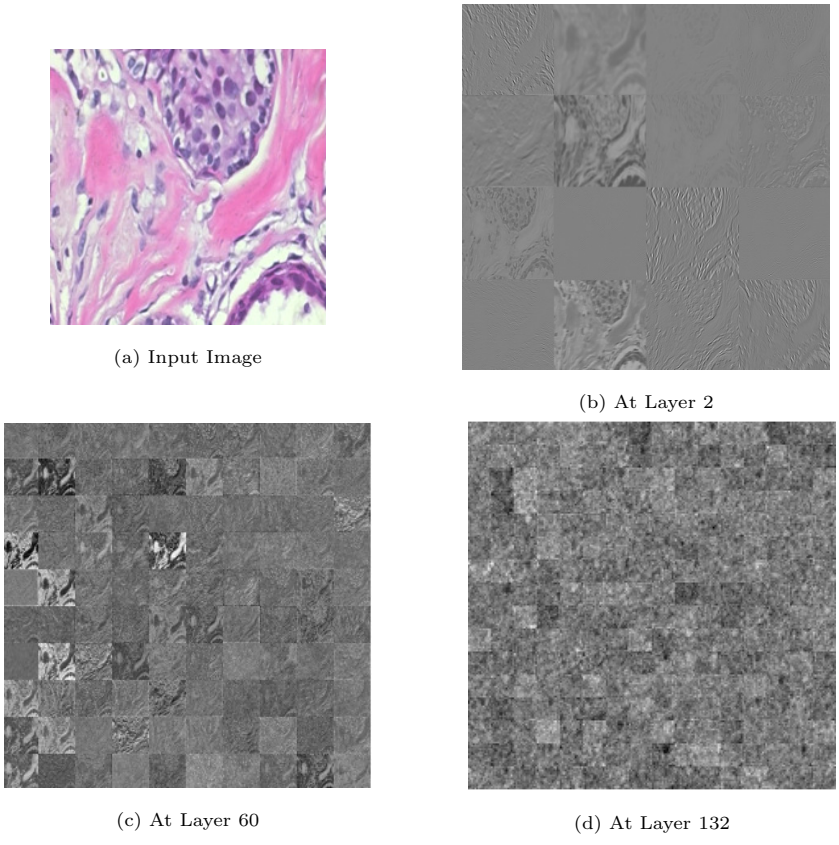
5.2.2. ResHist Features and Classifiers Results

ResHist model learned low-level to high-level deep features from the pathology images at the multiple levels of abstraction. To investigate the discriminative ability of deep features, we have used ResHist model as feature descriptor

Table 6: Confusion Matrices produced by ResHist Model.

	40X		**100X**		**200X**		**400X**	
	B	**M**	**B**	**M**	**B**	**M**	**B**	**M**
B	203 (26.2%)	67 (8.6%)	206 (25.1%)	70 (8.5%)	241 (31.3%)	32 (4.2%)	126 (21.3%)	53 (9.0%)
M	8 (1.0%)	497 (64.1%)	13 (1.6%)	531 (64.8%)	20 (2.6%)	477 (61.9%)	18 (3.0%)	395 (66.7%)

"B": Benign, "M": Malignant



(a) Input Image

(b) At Layer 2

(c) At Layer 60

(d) At Layer 132

Figure 5: Features maps of breast cancer histopathological image at different layers: b) Layer 2 output; c) Layer 60 output; d) Layer 132 output and a) input image.

for the histopathological image. Therefore, deep features are extracted from the input breast cancer histopathological images, and the extracted feature vectors

are used as input to the different classifiers. In this approach, four classifiers such as KNN [29], RF [30], QDA and SVM [31] have applied on the deep features. The pictorial representation of the features map at the different layers of ResHist model is shown in figure 5.

The performance of deep features with various classifiers are shown in Table 7. The performance of our model thus remains unaffected across diverse classifiers. The classifier's performance being quite similar and SVM giving the little better results among the classifiers (see underlined in Table 7).

Table 7: Performance of ResHist features with different classifiers.

Magnification Factors	**Method (s)**	**Image Level Accuracy (in %)**	**Patient Level Accuracy (in %)**
40X	ResHist+ KNN	86.37±5.34	85.59±4.82
	ResHist+QDA	86.86±4.49	86.61±3.72
	ResHist+RF	86.25±5.48	86.23±5.30
	ResHist+SVM	86.38±5.34	86.28±5.10
100X	ResHist+KNN	87.05±3.39	87.72±2.95
	ResHist+QDA	86.55±3.52	87.34±2.92
	ResHist+RF	87.03±3.17	87.72±2.89
	ResHist+SVM	87.28±3.49	87.91±3.05
200X	ResHist+KNN	**90.40±2.65**	**91.29±3.32**
	ResHist+QDA	**90.98±1.80**	**92.12±2.66**
	ResHist+RF	**90.26±2.41**	**91.58±3.11**
	ResHist+SVM	**91.35 ±2.19**	**92.46±3.10**
400X	ResHist+KNN	85.73±2.48	87.42±3.18
	ResHist+QDA	85.84±2.47	87.55±2.74
	ResHist+RF	86.23±2.27	87.88±2.61
	ResHist+SVM	86.29±2.26	87.90±2.76

Best results over the magnification factors are shown in bold. For each magnification factor, best results over classifiers is shown in underline.

5.2.3. Pre-trained Networks Results

The pre-trained networks that are trained on the massive labeled dataset of different applications that can be used for the new application using the transfer learning. In the transfer learning, weights of the pre-trained network are used and fine-tuned according to the new application.

In this paper, we have tested the performance of pre-trained CNN(s) namely AlexNet[16], GoogleNet [32] and VGG16 [33] on original BreaKHis dataset as well as augmented BreaKHis dataset. To fine-tuned these networks, last three layers (fully-connected layer, softmax layer, and output layer) of the network have been replaced with the new fully-connected layer (with two neurons), softmax layer and output layer. During fine-tuning, we have used 0.001 learning rate for newly added layers and 0.00001 learning rate for the remaining layers. Pre-trained networks require that input size should be the same as the

Table 8: Performance of pre-trained networks on BreaKHis (with and without data augmentation.

Accuracy @ in %	**Method (s)**	**Magnification Factors**			
		40X	**100X**	**200X**	**400X**
Image Level	AlexNet	80.58±2.50	80.59±2.98	84.095.29	79.26±7.13
	AlexNet-Aug	81.57±3.53	**83.64±2.97**	86.30±3.85	81.04±4.74
	GoogleNet	80.32±2.60	79.93±3.11	83.05±3.19	78.39±4.50
	GoogleNet-Aug	82.87±1.86	81.98±5.44	**87.03±1.59**	79.68±1.44
	VGG16	79.15±4.03	77.70±1.98	83.53±6.44	75.92±9.78
	VGG16-Aug	**86.03±2.85**	81.36±5.74	86.88±2.82	**81.25±4.17**
Patient Level	AlexNet	80.59±3.61	81.80±2.64	86.62±4.34	81.10±6.43
	AlexNet-Aug	83.07±4.93	**85.47±2.32**	**88.40±3.77**	81.83±3.74
	GoogleNet	80.95±3.43	80.40±3.16	83.38±4.23	79.77±5.62
	GoogleNet-Aug	83.68±2.29	82.34±4.00	86.71±2.54	80.14±2.54
	VGG16	81.33±3.57	78.43±3.81	83.43±6.88	79.01±9.30
	VGG16-Aug	**86.62±3.67**	81.27±5.31	88.18±2.94	**82.87±3.94**

Best results over the magnification factors are shown in underline. For each magnification factor, best results over the pre-trained networks are shown in the bold.

dimension of the input layer. Hence histopathological images have been re-sized to $227 \times 227 \times 3$ for AlexNet and $224 \times 224 \times 3$ for GoogleNet and VGG16 before applying them as input to the network. The experimental results are presented in Table 8. From the experimental results, it is observed that the all pre-trained networks achieved the best performance at the 200X magnification factor (best results are underlined) and worst at the 400X magnification factor. While comparing performances of the pre-trained networks, it is observed that GoogleNet-Aug achieved the overall best accuracy of 87.03% at image level and AlexNet-Aug achieved overall best accuracy of 88.40% at the patient level.

5.3. Comparison with Existing Methods

The performance comparison of ResHist model with the existing methods is shown in Table 9. Our approach has clearly outperformed the methodologies in [7], [8] and [9]. The proposed approach shows the superiority over the methods in [10] and [11]; however, the methodology in [10] has achieved better accuracy at image level for 40X magnification factor. Similarly, methodology in [11] has achieved better accuracy at the patient level for 40X and 100X magnification factors. Also, it can be observed from Table 9 that the proposed approach has yielded less standard deviation as compared to the existing methods. This suggests that the proposed approach is stable as compared to existing works.

Some of the salient features of the proposed model could be summarized as:

- ResHist classifies the whole slide images, thereby preserves the global information of histopathological images.
- Proposed method is fully automatic, with no image pre-processing requirement.
- Proposed data augmentation technique improves the performance of ResHist model significantly from 84.34% to 92.52% (8.18% margin).

Although the proposed approach has many advances over the hand-crafted local feature descriptors, it required immense computational power and time while training ResHist model as compare to feature descriptors.

Table 9: Performance comparison with existing methods in terms of accuracy (in%).

Author(s)	Magnification Factors							
	40X		100X		200X		400X	
	Accuracy @ IL	Accuracy @ PL	Accuracy @ IL	Accuracy @ PL	Accuracy @ IL	Accuracy @ PL	Accuracy @ IL	Accuracy @ PL
Spanhol et al. [7]	-	83.8±4.1	-	82.1±4.9	-	85.1±3.1	-	82.3±3.8
Spanhol et al. [8]	85.6±4.8	90.0±6.7	83.5±3.9	88.4±4.8	83.1±1.9	84.6±4.2	80.8±3.0	86.1±6.2
Bayramoglu et al. [9]	-	83.0±3.0	-	83.1±3.5	-	84.6±2.7	-	82.1±4.4
Spanhol et al. [10]	**89.6±6.5**	90.0±6.7	85.0±4.8	88.4±4.8	84.2±1.7	86.3±3.5	81.6±3.7	86.1±6.2
Sudharshan et al. [11]	87.8±5.6	**92.1±5.9**	85.6±4.3	**89.1±5.2**	80.8±2.8	87.2±4.3	82.9±4.1	82.7±3.0
Proposed	87.4±3.0	87.4±3.3	**87.2±3.5**	88.1±2.9	**91.1±2.3**	**92.5±2.8**	**86.2±2.1**	**87.7±2.4**

"Accuracy @ PL": Accuracy at Patient Level, "Accuracy @ IL": Accuracy at Image Level, "-": Unknown

**Bold shows the best results for the respective magnification factors*

6. CONCLUSION

In this paper, we have developed the ResHist model for automated diagnosis of breast cancer using histopathological images. The ResHist model achieves a classification accuracy of 92.52% and F1-score of 93.45% on the publicly available histopathological image (BreaKHis) dataset. Our approach outperforms the existing approaches for the classification of histopathological images. In addition, we have also investigated the performance of different classifiers on the deep features extracted from ResHist model. Specifically, we achieve a classification accuracy of 92.46% when deep features are fed to the SVM classifier. The results of our approach show the effectiveness of the ResHist model for classification of benign and malignant histopathological images. More importantly, the ResHist model outperforms the pre-trained networks, including AlexNet, GoogleNet, and VGG16 for classification of histopathological images.

Although the ResHist is very effective for breast cancer diagnosis, this approach needs to be validated on the larger dataset before being used for clinical

purpose. We intend to take up this exercise as a part of our future work. Further, we also like to explore the ResHist for diagnosis of lung cancer, colon cancer, and prostate cancer.

References

References

[1] P. Boyle, B. Levin, World Cancer Report 2008. Lyon: IARC (2008).
URL http://www.iarc.fr/en/publications/pdfs-online/wcr/2008/wcr_2008.pdf

[2] World Health Organization (Accessed: June. 15, 2019).
URL http://www.who.int/en/

[3] L. He, L. R. Long, S. Antani, G. R. Thoma, Histology image analysis for carcinoma detection and grading, Computer methods and programs in biomedicine 107 (3) (2012) 538–556.

[4] B. Stenkvist, S. Westman-Naeser, J. Holmquist, B. Nordin, E. Bengtsson, J. Vegelius, O. Eriksson, C. H. Fox, Computerized nuclear morphometry as an objective method for characterizing human cancer cell populations, Cancer research 38 (12) (1978) 4688–4697.

[5] M. Kowal, P. Filipczuk, A. Obuchowicz, J. Korbicz, R. Monczak, Computer-aided diagnosis of breast cancer based on fine needle biopsy microscopic images, Computers in biology and medicine 43 (10) (2013) 1563–1572.

[6] Y. M. George, H. H. Zayed, M. I. Roushdy, B. M. Elbagoury, Remote computer-aided breast cancer detection and diagnosis system based on cytological images, IEEE Systems Journal 8 (3) (2014) 949–964.

[7] F. A. Spanhol, L. S. Oliveira, C. Petitjean, L. Heutte, A dataset for breast cancer histopathological image classification, IEEE Transactions on Biomedical Engineering 63 (7) (2016) 1455–1462.

[8] F. A. Spanhol, L. S. Oliveira, C. Petitjean, L. Heutte, Breast cancer histopathological image classification using convolutional neural networks, in: Neural Networks (IJCNN), 2016 International Joint Conference on, IEEE, 2016, pp. 2560–2567.

[9] N. Bayramoglu, J. Kannala, J. Heikkilä, Deep learning for magnification independent breast cancer histopathology image classification, in: Pattern Recognition (ICPR), 2016 23rd International Conference on, IEEE, 2016, pp. 2440–2445.

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

[11] P. Sudharshan, C. Petitjean, F. Spanhol, L. E. Oliveira, L. Heutte, P. Honeine, Multiple instance learning for histopathological breast cancer image classification, Expert Systems with Applications 117 (2019) 103–111.

[12] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks, in: Advances in neural information processing systems, 2015, pp. 91–99.

[13] O. Abdel-Hamid, A.-r. Mohamed, H. Jiang, L. Deng, G. Penn, D. Yu, Convolutional neural networks for speech recognition, IEEE/ACM Transactions on audio, speech, and language processing 22 (10) (2014) 1533–1545.

[14] J. P. Dominguez-Morales, Q. Liu, R. James, D. Gutierrez-Galan, A. Jimenez-Fernandez, S. Davidson, S. Furber, Deep spiking neural network model for time-variant signals classification: a real-time speech recognition approach, in: 2018 International Joint Conference on Neural Networks (IJCNN), 2018, pp. 1–8. doi:10.1109/IJCNN.2018.8489381.

[15] P. Li, D. Wang, L. Wang, H. Lu, Deep visual tracking: Review and experimental comparison, Pattern Recognition 76 (2018) 323–338.

[16] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in neural information processing systems, 2012, pp. 1097–1105.

[17] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.

[18] H. Lu, Y. Li, M. Chen, H. Kim, S. Serikawa, Brain intelligence: go beyond artificial intelligence, Mobile Networks and Applications 23 (2) (2018) 368–375.

[19] J. Kleesiek, G. Urban, A. Hubert, D. Schwarz, K. Maier-Hein, M. Bendszus, A. Biller, Deep mri brain extraction: a 3d convolutional neural network for skull stripping, NeuroImage 129 (2016) 460–469.

[20] M. Talo, U. B. Baloglu, Ö. Yıldırım, U. R. Acharya, Application of deep transfer learning for automated brain abnormality classification using mr images, Cognitive Systems Research 54 (2019) 176–188.

[21] C. Lam, C. Yu, L. Huang, D. Rubin, Retinal lesion detection with deep learning using image patches, Investigative ophthalmology & visual science 59 (1) (2018) 590–596.

[22] P. Naylor, M. Laé, F. Reyal, T. Walter, Nuclei segmentation in histopathology images using deep neural networks, in: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), 2017, pp. 933–936. doi:10.1109/ISBI.2017.7950669.

[23] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, arXiv preprint arXiv:1502.03167 (2015).

[24] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016, http://www.deeplearningbook.org.

[25] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).

[26] A. A. A. Setio, F. Ciompi, G. Litjens, P. Gerke, C. Jacobs, S. J. Van Riel, M. M. W. Wille, M. Naqibullah, C. I. Sánchez, z. van Ginneken, Pulmonary nodule detection in ct images: false positive reduction using multi-view convolutional networks, IEEE transactions on medical imaging 35 (5) (2016) 1160–1169.

[27] S. C. Wong, A. Gatt, V. Stamatescu, M. D. McDonnell, Understanding data augmentation for classification: when to warp?, arXiv preprint arXiv:1609.08764 (2016).

[28] D. Magee, D. Treanor, D. Crellin, M. Shires, K. Smith, K. Mohee, P. Quirke, Colour normalisation in digital histopathology images, in: Proc Optical Tissue Image analysis in Microscopy, Histopathology and Endoscopy (MICCAI Workshop), London,U.K, 2009, pp. 100–110.

[29] L. E. Peterson, K-nearest neighbor, Scholarpedia 4 (2) (2009) 1883.

[30] L. Breiman, Random forests, Machine learning 45 (1) (2001) 5–32.

[31] C. Cortes, V. Vapnik, Suport-vector networks, Mach. Learning 20 (1995) 273–297.

[32] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1–9.

[33] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556 (2014).