ISSN 2278-3091
Volume 10, No.1, January - February 2021
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse171012021.pdf
https://doi.org/10.30534/ijatcse/2021/171012021

# Classification of breast cancer types based on deep learning approach

**Osama Mohamed [1], Hala Nagy[2], Hamdi Mahmoud[3]**
[1]Faculty of computer and artificial intelligence, Beni-Suef University, Egypt, osama.mohamed@fcis.bsu.edu.eg
[2]Faculty of computer and artificial intelligence, Helwan University, Egypt, hala.nagy@fci.helwan.edu.eg
[3]Faculty of computer and artificial intelligence, Beni-Suef University, Egypt, Dr_hamdimahmoud@yahoo.co

## ABSTRACT

Breast cancer is one of the most serious diseases that affect women, so it must be discovered in the early stages to avoid complications such as redness of the skin, pain in the armpits or breast, and discharge from a nipple, possibly containing blood. Recently, the CAD system that is based on the classification of microscopic image play a vital rule to limit cancer disease and reduce cases. Microscopic image is the currently recommended image system used to detect cancer. A computer-aided diagnosis system will help radiologists to accurately detection of cancerous cells and achieve the best result. This paper proposes a deep learning technique that exploits CAD system features and microscopic images to fight breast cancer. The proposed technique builds a classification model based on the DenseNet-161 deep learning method. The proposed model classifies the microscopic images of breast cancer into benign with four types and malignant with four types. Our proposed technique is experimentally tested and the result confirmed that a proposed technique outperforms baseline techniques.

**Key words:** Deep learning, Transfer learning, one fit cycle policy, Cyclical Learning Rates (CLRs), Breast cancer, Machine learning, Image classification

## 1. INTRODUCTION

Automated detection of many types of breast cancer from a histopathological images is valuable for a clinical pathologists. According to reports raised in 2018, there are new cases of breast cancer which are around 2,088,849 over the world, this number represents 11.6% of all total cases of cancer types reported in this year, besides the number of death reported is 626,679 almost 6.6% from all total cases [1, 2]. Additionally, results in [3] report that incidence rates are reported as 19.3 per 100,000 women in Eastern Africa, to 89.7 per 100,000 women in Western Europe. It is expected that the number of new cases continues to grow until this number is increased to be 27 million in 2030. For the next 40 years, the diagnosis of breast cancer is based on X-ray, MRI (Magnetic Resonance Image) and ultrasound, etc. [4].

Pathologists diagnosis of the tissue slices by microscope examination [6]. Pathologists with no experience are likely to make diagnostic errors. The use of Computer-Aided Diagnosis (CAD) improves the diagnostic efficiency of histopathological image classification [6, 7]. A biopsy is known to be the only way to diagnose if the affected area is cancerous [5].

In detecting breast cancer through a classification of microscopic images, there are many challenges, the first of which is the histological images of breast cancer, complex, many colors, which are difficult to separate, to distinguish between cancer and other types, and the second is feature extraction is a big problem. Common methods for extracting image attributes are scale-invariant feature transform (SIFT) [12] and gray-level co-occurrence matrix (GLCM) [13]. Deep learning techniques have a large and distinct ability to extract image features and can quickly learn and is a solution [8, 9] to the problem of extracting image features, so it has been applied and achieved successful results in many areas such as the medical fields, the pharmaceutical industry, and other fields related to the field of computer vision [10, 11]. It is known that it is impossible to train deep learning from scratch by using a small set of data, so pre-trained models were used to save time that researchers obtained with training in big data called ImageNet. This is what is known as transfer learning, by re-training the fully connected layer of the model [16]. It is reported that deep learning achieves amazing success in many fields but it takes a long time is not necessary to train the model with the suboptimal hyperparameters so setting hyper-parameter is a big challenge [15] which face any researcher who uses deep learning and takes many years of experience to tune these parameters [14]. The solution is one cycle policy is a method that is used to reduce the time of training, improve performance, and tune all hyperparameters of deep learning models such as learning-rate, Weight-decay [14]. Figure 6 show that one cycle policy achieves training result higher than the result from the standard learning rate.

**1.1 *Contributions.*** The main contribution of this work can be summarized as follows:

1- We introduce a technique that classifies microscopic images into benign with four classes (Adenosis, Fibroadenoma, Tabular Adenoma, Phyllodes Tumors) and malignant with four classes (Ductal Carcinoma, Lobular Carcinoma, Mucinous Carcinoma, Papillary Carcinoma)

2- We develop a new method that reduce the number of epochs and iterations which allows the model to work on large-scale data.
3- We conduct several experimental evaluations for the proposed model over real-life datasets. Results confirmed that our model outperforms the state of the art classification model. Additionally, classification accuracy achieved is around 97.7.
This paper is organized as follows: Related work is described in section 2. The background is described in section 3. The proposed model is described in section 4. Experimental results and analysis is described in section 5.

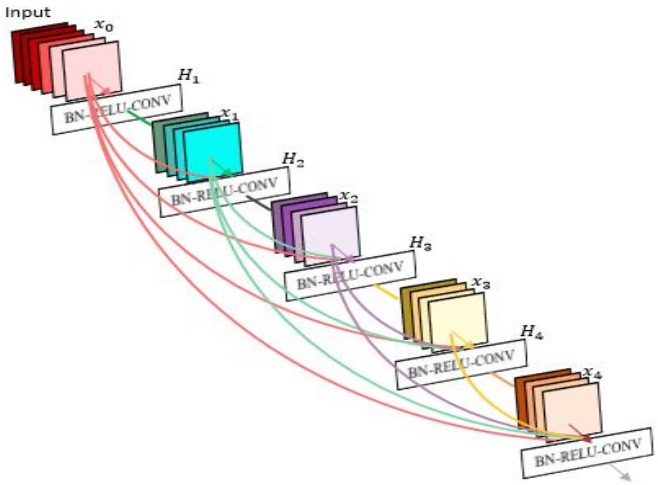


**Figure 1:** A 5-layer dense block with a growth rate of k = 4. Each layer takes all preceding feature-maps as input.

## 2. RELATED WORK

This section surveys the state of the art for all previous studies that occurred in the diagnosis of breast cancer disease using machine learning algorithms [9, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, and 31]. Many studies have been conducted in detecting breast cancer types. Researchers have proposed and developed many models to solve this problem of classification. Spanhol, Fabio, Luiz Oliveira, Caroline, and Laurent used four classification algorithms, they are SVM (Support Vector Machine), KNN (K-Nearest Neighbour), RF (Random Forest) and QDA (quadratic lin-ear analysis) to classify tumor to malignant or benign and achieve accuracy between 80 and 85 [21]. Spanhol, Fabio, Luiz Oliveira, Caroline, and Laurent used CNN (Convolution Neural Network) to classify Histopathological images to malignant or benign and got accuracy between 80 and 89.6, AlexNet is used as feature extractor [22]. Chan, Alan, and Jack used fractal dimension as feature extractor and SVM as a classification algorithm and achieve the highest accuracy on 40X magnification factor are 97.7 [23]. Spanhol, Fabio, Luiz, Paulo, Caroline, and Laurent used Logistic regression as classifier and CNN as feature extractor to work on a binary class classification to malignant or benign and accuracy achieved between 81 and 85 as image-level [24]. Han, Zhongyi, Benzheng, Yuanjie, Yilong, Kejian, and Shuo used GoogleNet as a classifier and achieved accuracy between 95 and 97 on binary classification and achieved accuracy on multiclass classification between 92.9 and 93.9 [25]. Song, Yang, Ju, Hang, and Weidong used SVM as a classifier and VGG as a feature extractor and achieve accuracy between 82.9 and 87 as image-level [26]. Kahya, Mohammed, Waleed, and Zakariya applied experiments on binary classification and achieved results between 93.62 and 94.97 using Adaptive Sparse Support Vector Machine [27]. Das, Kausik, Sri, Abhijit, Jyotirmoy, and Debdoot used GoogleNet to achieve accuracy 93.49 and 94.82 on binary class classification [28]. Zhi, Weiming, Henry, Zhenghao, Seid, Zhicheng, and Yuk used VGG to classify histopathological images to malignant or benign using VGG and achieve accuracy between 84.58 and 91.45 [29]. Benhammou, Yassir, Siham, Boujemâa, and Francisco applied the experimental results on binary class classification and achieved accuracy between 80.3 and 86.5 [30]. Gandomkar, Ziba, Patrick, and Claudia has worked on a multiclass classification problem and achieve accuracy between 94.6 and 95.6 by using ResNet-152 [31]. Jiang, Yun, Li, Hai, and Xiao used small SE-ResNet to classify a histopathological images into benign and malignant subtypes applied on an augmented dataset and achieve the higher accuracy in 40X magnification factor is 94.43 on multi-classification and achieve the higher accuracy in 40X magnification factor is 98.87 in binary classification. The dataset used is BreakHis. This author used 300 epochs than 200 because it achieves the best results. The author started with a learning rate 0.1 and are decreased by factor 5 after each of 60,120 and 160 epochs. The previous author uses of SE-ResNet to reduce the number of training parameters of the model. These results were achieved due to the cell overlap and uneven color distribution [17]. Bardou, Dalal, Kun, and Sayed used kernel principal component analysis (KPCA) model ensemble as a classification model and applied binary and multi-classification and achieve the higher accuracy in 40X magnification factor is 88.23. The previous author made a comparison between CNN and feature-based classification on the BreakHis dataset. This paper is used data augmentation techniques such as rotation by 90, 180, and 270 and applying horizontal flip to the whole training images. The author is set the learning rate to 0.001.On the CNN method, the previous author achieves the best results in binary and multi-class classification with CNN and prove that handcrafted features gave lower results than CNN features [18]. Xie, Juanying, Qi, Yinghuan, Lv, Liping, Fuzhen, Junping, Xiaoyang, and Shengquan divided data set into 70:30 training and testing, applied his experiments on raw and augmented data with Inception_V3 (INV3) and Inception_ResNet_V2 (IRV2) with transfer learning, not from scratch and achieve the higher result on raw data in binary classification is 97.90 in 40X magnification factor and multiclass classification in 40X is 92.07. After applying augmentation, 40X magnification factor accuracy is 99.79 in binary classification and accuracy of 97.63 in 40X magnification factor in multiclass classification. In the previous paper, it displayed their result in terms of ACC_IL, ACC_PL, Kappa, and Micro F1 [9]. Kassani, Sara, Peyman,

Michal, Kevin, and Ralph has applied their experimental results on the BreakHis dataset. The author used a method of an ensemble of three models of deep learning are VGGNet, MobileNetV2, and DenseNet201. This research focuses on binary classification. The author resizes all images to 244x244 using bicubic interpolation and the batch size was set to 32 and all models trained for 1000 epochs. The author applied data augmentation by using horizontal flip, vertical flip, contrast enhancement, zoom range with 0.2, shear range with 0.2 rotational range with 0.2, and fill mode with nearest. This paper achieves an accuracy of 98.13% on the BreakHis dataset, 94.64% on the PatchCamelyon dataset, 95.00% on the ICIAR dataset, and 83.10% on the Bioimaging dataset by using Adam optimizer. Bioimaging achieves the worst accuracy [19]. Zhu, Chuang, Fangzhou, Ying, Huihui, Yao, and Jun used a novel approach HybridCNN applied to Bach dataset and BreakHis dataset. The author achieves the best accuracy on the BreakHis dataset [20]. My work differentiates itself.

## 3. BACKGROUND

This section shows the types of DenseNets and their layers and why it used in this research paper. DenseNets are the subsequent stage to increase the depth of deep Convolution Network. When CNN's go deeper the problems are arises. This happened because of the big path for information from input to the output layer. DenseNet-161 is a simple connectivity pattern because it connects all layers directly with each other to be sure that information flow is maximum between layers in the network. Feedforward nature is maintained by obtaining each layer additional inputs from the preceding layers. Figure 1 shows the architecture. The feature is combined by concatenation. DenseNet is not the same as traditional architecture because it introduces $\frac{L(L+1)}{2}$ connections in an L-layer network instead of L. Handling problem of vanish gradient, Reusing feature, lacking parameter's number and propagating features are the most important feature of DenseNets. Figure 2 presents architectures of DenseNet for ImageNet.

| Layers | Output Size | DenseNet-121($k = 32$) | DenseNet-169($k = 32$) | DenseNet-201($k = 32$) | DenseNet-161($k = 48$) |
|---|---|---|---|---|---|
| Convolution | 112 × 112 | 7 × 7 conv, stride 2 | | | |
| Pooling | 56 × 56 | 3 × 3 max pool, stride 2 | | | |
| Dense Block (1) | 56 × 56 | [1 × 1 conv, 3 × 3 conv] × 6 | [1 × 1 conv, 3 × 3 conv] × 6 | [1 × 1 conv, 3 × 3 conv] × 6 | [1 × 1 conv, 3 × 3 conv] × 6 |
| Transition Layer (1) | 56 × 56 | 1 × 1 conv | | | |
| | 28 × 28 | 2 × 2 average pool, stride 2 | | | |
| Dense Block (2) | 28 × 28 | [1 × 1 conv, 3 × 3 conv] × 12 | [1 × 1 conv, 3 × 3 conv] × 12 | [1 × 1 conv, 3 × 3 conv] × 12 | [1 × 1 conv, 3 × 3 conv] × 12 |
| Transition Layer (2) | 28 × 28 | 1 × 1 conv | | | |
| | 14 × 14 | 2 × 2 average pool, stride 2 | | | |
| Dense Block (3) | 14 × 14 | [1 × 1 conv, 3 × 3 conv] × 24 | [1 × 1 conv, 3 × 3 conv] × 32 | [1 × 1 conv, 3 × 3 conv] × 48 | [1 × 1 conv, 3 × 3 conv] × 36 |
| Transition Layer (3) | 14 × 14 | 1 × 1 conv | | | |
| | 7 × 7 | 2 × 2 average pool, stride 2 | | | |
| Dense Block (4) | 7 × 7 | [1 × 1 conv, 3 × 3 conv] × 16 | [1 × 1 conv, 3 × 3 conv] × 32 | [1 × 1 conv, 3 × 3 conv] × 32 | [1 × 1 conv, 3 × 3 conv] × 24 |
| Classification Layer | 1 × 1 | 7 × 7 global average pool | | | |
| | | 1000D fully-connected, softmax | | | |

**Figure 2:** DenseNet architectures for ImageNet

The proposed method includes types of CNN (Convolution Neural Network), which is DenseNet-161. Hung, Liu and Maaten [32] developed DenseNet which had the best classification results on the available datasets such as ImageNet. DenseNet isn't used to direct connections among the hidden layers but it used a dense connections to build a model. Its construction was based on linking each to a subsequent layer. In any layer, any important features learned are involved within the network. Due to the previous feature, deep network training became more efficient, the performance of the model increased. The number of parameters has become less than CNN because feature maps are sent directly to all subsequent layers. The DenseNet has a very important feature which is that it reduces overfitting in the model because of using of dense connections.

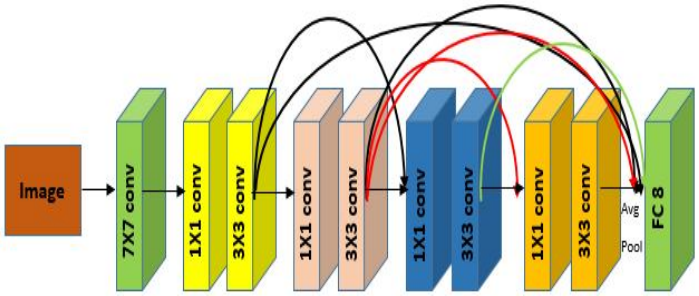


**Figure 3:** Schematic representation of DenseNet-161 architecture.

## 4. PROPOSED MODEL

This section describes the system architecture of our proposed model. The first sub-section introduces the main components of the proposed system. Besides, the second subsection introduces the improvements in the proposed model to ensure run-time optimization.

### 4.1 SYSTEM OVERVIEW

The proposed architecture topology is given in figure 3. Training deep learning models with very large numbers of parameters take much time. More and more data and powerful GPU are required to train these models from scratch. Transfer learning is used to overcome the previous problem. By using transfer learning you are saving time. In transfer learning, models have been previously trained on a dataset which represents features of the dataset it was trained on. Learned features are often transferable to different data. For example, a model trained in Dataset for animal images that include learned features such as edges and lines that would be transferable to your dataset. In transfer learning, the feature extractor is done by a full connected layer after removing it from the model used. Well-known architecture, DenseNet-161 is used based on performance in vision tasks and used in real-time applications. The difference between our work and the state of the art is to use a method of training is called one fit cycle policy that facilitates the process of training and no need to tune any hyper-parameter of the used network.

Our proposed model is applied to raw data and augmented data is displayed in figure 4. The detailed description of our proposed technique is composed of three major steps that can be described as follows.

### STEP 1: PREPROCESSING THE INPUT SAMPLES

The input histopathological image sample is processed by resizing images to 244x244 and normalize training images only. The concept of image normalization is changing the range of pixel intensity values. Image normalization is used to change over the pixel's range values into a range that is more familiar with the sense. The preprocessing step is a common step on raw and augmented data, but other differences that will be clarified. The other steps on augmented data are Data augmentation, applying DenseNet-161. The other steps on raw data are weighted cross-entropy, applying DenseNet-161.

### STEP 2: DATA AUGMENTATION

Data augmentation is used because of the restricted size of dataset samples [19]. This limited size of dataset samples leads to the occurrence of overfitting in CNN [37]. To solve this issue, data augmentation is used. The aim of data augmentation is to create more training samples from the available data [38]. There are many forms of data augmentation such as zooming, rotating, shear, and horizontal and vertical flip are applied to generate more training data.

### STEP 3: WEIGHTED CROSS-ENTROPY

Weighted cross-entropy is used to balance the imbalanced dataset's classes. In this paper, the dataset used is an imbalanced so balanced cross-entropy loss is used to overcome this problem. There is a built-in function is called Cross Entropyloss is used to handle imbalanced datasets in python.

### STEP 4: A PREDICTION MODEL

It is critical to note that this prediction model is built upon DenseNet-161 model notations. Figure 4 describes the step of applying DenseNet-161 as the final step to obtain the multi-class classification result. The architecture of DenseNet-161 is shown in figure 2. In this paper, DenseNet-161 with ImageNet is used but the last layer "fully connected" layer is changed according to our problem from 1000 classes to 8 classes. DenseNet-161 is applied by using a method called one fit cycle policy.

### 4.2 ONE FIT CYCLE POLICY

This section shows the deep learning model used and its performance. It is known training of Deep Neural Network (DNN) is a difficult optimization problem. Tuning of hyper-parameter such as Learning Rate (LR) is very important. Small Learning Rate is led to very slowly training, while a Large Learning Rate hinders the convergence. Low LR is good but it takes a long time to train perfectly. When training speed is increased, LR is increased until LR gets too large and diverges. To obtain the exact LR, you need to make many experiments and be patient. A new method is discovered by Leslie N. Smith for setting LR, is named Cyclical Learning Rates (CLRs), this method is employed in this proposed technique for training. CLR made LR values between minimum and maximum range instead of fixed, increasing, and decreasing LR. CLR cycle has two steps, one of them being an increase in LR and the other one being a decrease in LR. CLR eliminates the need to find the optimal LR but the optimal rate between minimum and maximum range. Figure 5 shows that classification accuracy while training CIFAR-10. The red curve is CLR. From figure 5, the CLR achieves the same accuracy of the original learning rate but in iteration less than the original learning rate method [39].

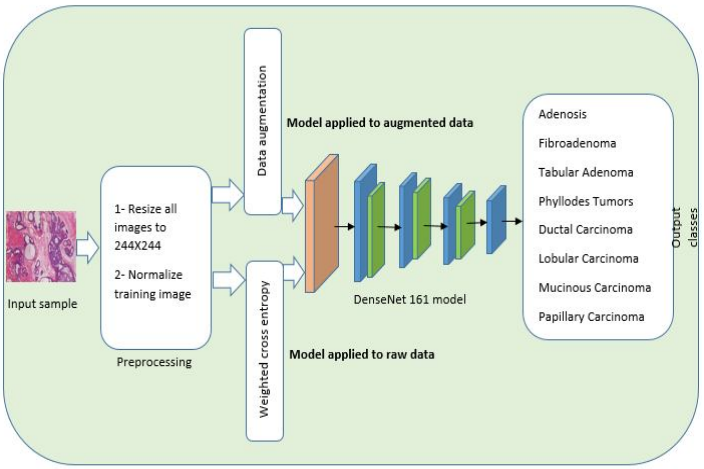


**Figure 4:** Proposed model

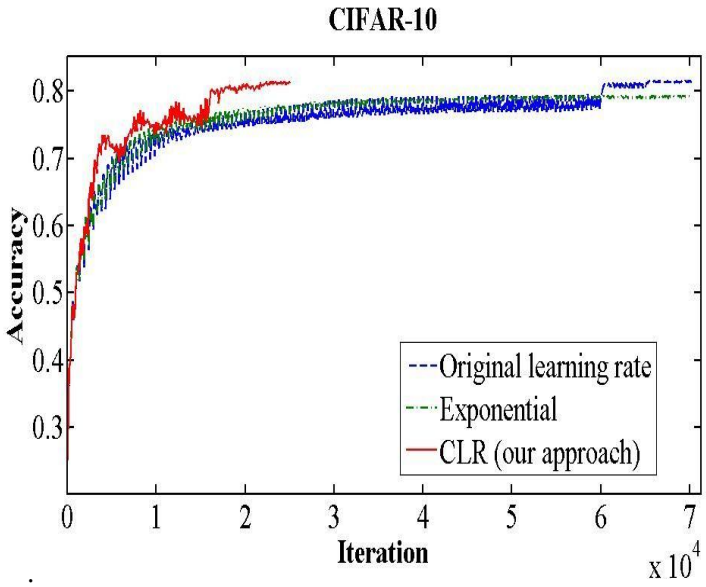


**Figure 5:** An illustrative example of the CLR method.

In Leslie N. Smith's research, super-convergence is a method that uses CLR, but with one cycle which contains two learning rate steps. A total number of iteration must be larger than the size of a cycle. After completing the cycle, LR is decreased much further for the remaining iterations. Leslie N. Smith named this method to one fit cycle policy. In super-convergence, LR is started from a small value and is increased to a very large value then returns to a value lower than its initial one. The impact of LR many values is a training accuracy curve. In super-convergence, the training accuracy is moved fast as LR is increased, is become oscillated as LR is very large and then jumps again to an extreme point of accuracy as to be shown in figure 6 [14, 40].

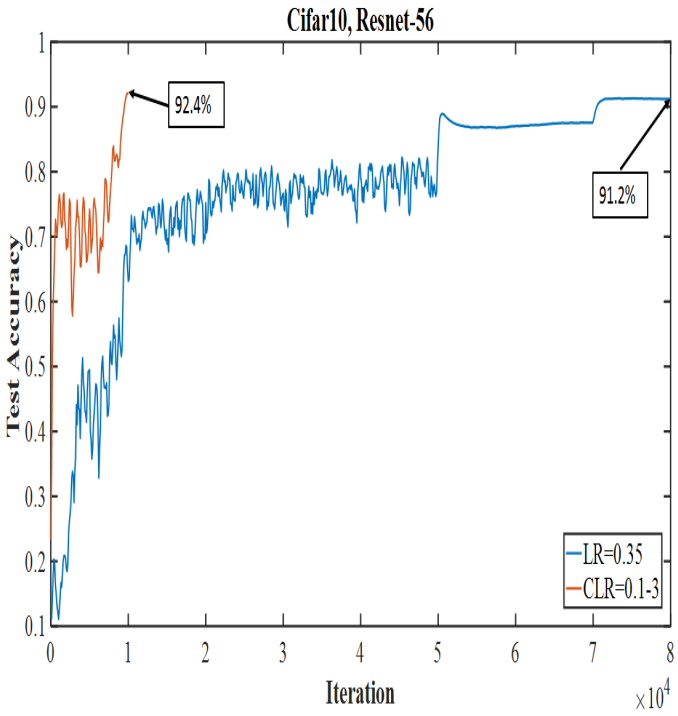


**Figure 6:** An illustrative example of the super-convergence training accuracy curve

## 5. Experimental Results and Analysis

### 5.1 Experiments Settings

Experiments are performed on Google colab [41]. The dataset used in our research is named BreakHis. It is collected from a clinical study in 2014 from Brazil [33] which is composed of 7,909 histopathological images collected from 82 patients. These picture sizes are 700 x 460 with 3 RGB channels. It is divided into 5429 images of malignant breast cancer and 22480 of benign breast cancer [34]. It consists of eight types of malignant and benign classes [35]. Malignant classes are Ductal Carcinoma (DC), Lobular Carcinoma (LC), Mucinous Carcinoma (MC), and Papillary Carcinoma (PC). Benign classes are Adenosis (A), Fibroadenoma (F), Phyllodes Tumors (PT), and Tabular Adenoma (TA). It contained four magnification factors: 40X, 100X, 200X and 400X [36]. Table 1 shows the detailed description of the BreakHis dataset. Table 2 shows image and patient distribution over the types of tumor and sub-types. Figure 7 shows a sample of BreakHis 40x.

**Table 1:** Structure of the BreakHis dataset.

| Classes | Subtypes | Magnification factors | | | | Total |
|---|---|---|---|---|---|---|
| | | 40X | 100X | 200X | 400X | |
| **Benign** | (A) | 114 | 113 | 111 | 106 | 444 |
| | (F) | 253 | 260 | 264 | 237 | 1,014 |
| | (TA) | 109 | 121 | 108 | 115 | 453 |
| | (PT) | 149 | 150 | 140 | 130 | 569 |
| **Malignant** | (DC) | 864 | 903 | 896 | 788 | 3,451 |
| | (LC) | 156 | 170 | 163 | 137 | 626 |
| | (MC) | 205 | 222 | 196 | 169 | 792 |
| | (PC) | 145 | 142 | 135 | 138 | 560 |
| **Total** | | 1,995 | 2,081 | 2,013 | 1,820 | 7,909 |

**Table 2:** Image and patient distribution in Main-types and sub-types of breast cancer.

| Main-types | Benign | | | | Total Benign | Malignant | | | | Total Malignant | Total Both |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Sub-types | A | F | TA | PT | | DC | LC | MC | PC | | |
| Number of images | 444 | 1014 | 569 | 453 | 2480 | 3451 | 626 | 792 | 560 | 5429 | 7,909 |
| Number of patients | 4 | 10 | 7 | 3 | 24 | 38 | 5 | 9 | 6 | 58 | 82 |

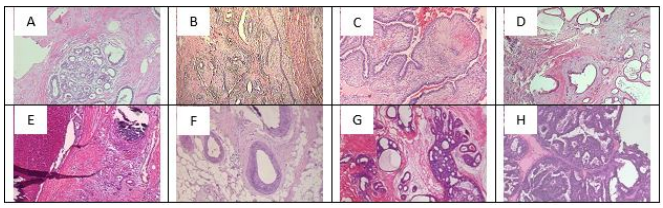


**Figure 7:** (A) Adenosis, (B) Fibroadenoma, (C) Phyllodes Tumor, (D) Tubular Adenoma, (E) Ductal Carcinoma, (F) Lobular Carcinoma, (G) Mucinous Carcinoma, (H) Papillary Carcinoma.

### 5.2 Evaluation Criteria

The evaluation criteria used to evaluate the performance of the classification model, including image-level test accuracy (ACC_IL), patient-level test accuracy (ACC_PL), Macro-F1, Micro-F1, and Kappa criteria. Macro-F1 takes the average of the precision and recall of each class. ACC_IL is defined by the ratio of Ncor (the number of the correctly classified images in a testing set), to Nall (Total number of images in a testing set). Equation (1) defines image-level test accuracy (ACC_IL). ACC_PL is defined by the summation of patient

score divided by a total number of patients in the testing set. The patient score is a ratio of a correctly classified image of patient P (Ncor) divided by all images associated with patient P in the testing set (NP). Equation (2) defines patient-level test accuracy (ACC_PL). Micro-F1 is defined by the sum up the individual true positives, false positives, and false negatives for all classes. Precision is calculated as the sum of true positives across all classes divided by the sum of true positives and false positives across all classes. A recall is calculated as the sum of true positives across all classes divided by the sum of true positives and false negatives across all classes. Equation (3) defines Micro-F1. Kappa measure, based on confusion matrix calculation, can handle a problem such as imbalanced datasets and multiclass problems. Equation (4) defines Kappa coefficient, where $p_0$ is the image level test accuracy (ACC_IL) is defined in (1), and $p_e$ is the summation of the product of the number of the image in each type of cancer and the predicted number of images in each type of cancer to the square of the total number of images in the testing set. Precision means the percentage of your results that are relevant. On the other hand, recall refers to the percentage of total relevant results correctly classified by your algorithm.

$$\text{ACC\_IL} = \frac{Ncor}{N_{all}} \quad (1)$$

$$\text{ACC\_PL} = \frac{\sum \text{Patient Score}}{\text{Total Number of Patients}}, \text{ Patient Score} = \frac{N_{cor}}{N_p} \quad (2)$$

$$\text{F1-score} = \frac{2X(\text{precision} \times \text{recall})}{(\text{precision} + \text{recall})},$$

$$\textit{Precision} = \frac{\text{Sum c in C TruePositives\_c}}{\text{Sum c in C (TruePositives\_c + FalsePositives\_c)}},$$

$$\text{Recall} = \frac{\text{Sum c in C TruePositives\_c}}{\text{Sum c in C (TruePositives\_c + FalseNegatives\_c)}} \quad (3)$$

$$\text{Kappa} = \frac{p_0 - p_e}{1 - p_e}, \; p_0 = \frac{Ncor}{N_{all}}, \; p_e = \frac{\sum NtrueXNpre}{N_{all}XN_{all}} \quad (4)$$

### 5.3 Classification Results

The next subsection discusses the classification result on histopathological images of the BreakHis dataset based on the default one fit cycle policy approach. The experimental result is applied to a raw dataset and augmented dataset. Moreover, the results of our research experiments are compared with the results of other researchers.

### 5.4 Experiments on the Raw Dataset (Multi-class problem)

All research papers only focus on binary classification but multi-classification is the most significant problem. We used Dense-Net161 to perform classification of a histopathological image into benign subclasses (Adenosis, Fibroadenoma, Phyllodes Tumor, and Tubular Adenoma) and malignant subclasses (Ductal Carcinoma, Lobular Carcinoma, Mucinous Carcinoma, and Papillary Carcinoma) by using a pre-trained model in terms of ACC_IL, ACC_PL, F1, AUC, and Kappa. Our experimental result of the multi-classification problem on raw data shown in table 3 according to ACC_IL, ACC_PL, Macro-F1, Micro-F1, and Kappa. The dataset is divided into 65% training set, 15% validation set, and 20% testing set. We run the raw data on 90 epochs. All classification results are given in table 3 and the obtained accuracy for the raw data is 93.98, 91.85, 90.32, and 89.65 for the magnification factors 40X, 100X, 200X, and 400X. The loss curves are shown in figure 8 and the confusion matrices are shown in figure 9.

**Table 3:** The result of each evaluation on raw data with respect to different magnification factors.

| Network | Criteria | Magnification factors | | | |
|---|---|---|---|---|---|
| | | **40X** | **100X** | **200X** | **400X** |
| Dense-Net 161 | ACC_IL | 93.98 | 91.85 | 90.32 | 89.65 |
| | ACC_PL | 95.85 | 93.92 | 93.07 | 91.16 |
| | Macro_F1 | 94.09 | 91.04 | 90.77 | 88.90 |
| | Micro_F1 | 93.98 | 91.85 | 90.32 | 89.65 |
| | Kappa | 92.16 | 89.26 | 86.95 | 86.36 |

**Table 5:** Precision, Recall and F1-score for raw data.

| Subtypes | Precision | | | | Recall | | | | F1-score | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 40X | 100X | 200X | 400X | 40X | 100X | 200X | 400X | 40X | 100X | 200X | 400X |
| (DC) | 0.95 | 0.96 | 0.90 | 0.93 | 0.92 | 0.93 | 0.92 | 0.90 | 0.94 | 0.95 | 0.91 | 0.91 |
| (LC) | 0.79 | 0.68 | 0.64 | 0.72 | 0.79 | 0.78 | 0.59 | 0.84 | 0.79 | 0.73 | 0.61 | 0.78 |
| (MC) | 0.94 | 0.96 | 0.97 | 1.00 | 1.00 | 0.85 | 0.90 | 0.85 | 0.97 | 0.90 | 0.94 | 0.92 |
| (PC) | 0.95 | 0.88 | 0.96 | 0.92 | 1.00 | 0.95 | 1.00 | 0.86 | 0.97 | 0.91 | 0.98 | 0.89 |
| (PT) | 1.00 | 1.00 | 1.00 | 0.94 | 1.00 | 0.88 | 0.86 | 0.85 | 1.00 | 0.94 | 0.92 | 0.89 |
| (TA) | 0.87 | 0.87 | 1.00 | 0.91 | 0.91 | 1.00 | 0.96 | 0.97 | 0.89 | 0.93 | 0.98 | 0.94 |
| (A) | 1.00 | 0.96 | 0.95 | 0.89 | 1.00 | 1.00 | 1.00 | 0.94 | 1.00 | 0.98 | 0.98 | 0.91 |
| (F) | 0.98 | 0.93 | 0.90 | 0.80 | 0.96 | 0.97 | 0.98 | 0.95 | 0.97 | 0.95 | 0.94 | 0.87 |

The experimental results in table 3 show that all evaluation metric on 40X magnification factor (which is indicated by black underline) is better than the other magnification factors. The reason for 40X achieves the best accuracy because it contains a more significant features of breast cancer. From table 3 kappa values show that our model classification results are so perfect. 400X unlike all other research achieve good classification result.

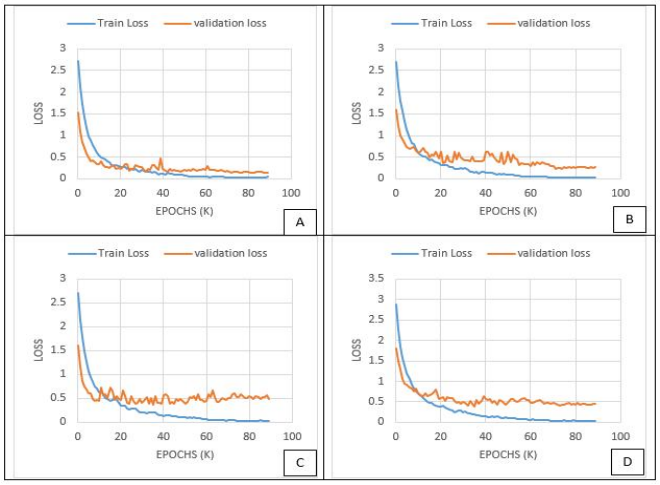


**Figure 8:** A. is the loss curve for 40X, B is the loss curve for 100X, B is the loss curve for 200X, and C is the loss curve for 400X. These loss curves for raw data with 4 magnification factors.

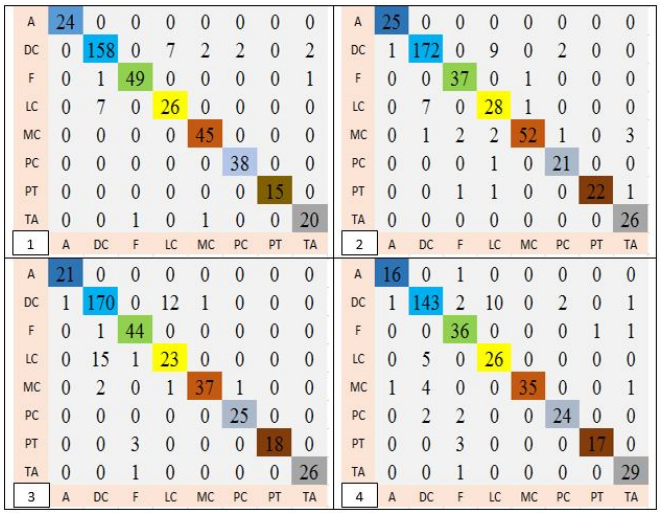


**Figure 9:** The confusion matrix of raw data for all magnification factors 40X, 100X, 200X, and 400X shown in orders 1, 2, 3, and 4.

### 5.5 Experiments on the Augmented Dataset (Multi-class problem)

According to table 3 and figure 8 the multi-classification is not the best classification result because the number of classes are large and there isn't sufficient data to train the deep learning model and the imbalanced dataset distribution of classes. It is noticeable that class Ductal Carcinoma is classified as Lobular Carcinoma. It is noticeable that a lower number of benign breast cancer classified as malignant breast cancer and vice versa. To improve the accuracy of classification, we increase the size of the dataset to overcome the problem of an imbalanced datasets. The expanding process of a dataset is done based on the DC subclass. This expanding done by expanding the other remaining classes by oversampling and apply transforms such as max zoom with 1.25, flip vertical, max warp with 0.4, and the probability of applying these transformations 0.5. By applying the previous transformation the new distribution of the dataset is shown in table 6.

**Table 6:** The new distribution of image after oversampling applied in all magnification factors.

| Magnification Factor | Benign | | | | Malignant | | | | Total |
|---|---|---|---|---|---|---|---|---|---|
| | A | F | PT | TA | DC | LC | MC | PC | |
| 40X | 798 | 759 | 763 | 894 | 864 | 936 | 1,025 | 870 | 6,909 |
| 100X | 791 | 780 | 847 | 900 | 903 | 1,020 | 1,110 | 852 | 7,203 |
| 200X | 777 | 792 | 756 | 840 | 896 | 978 | 980 | 810 | 6,829 |
| 400X | 742 | 711 | 805 | 780 | 788 | 822 | 845 | 828 | 6,321 |
| Total | 3,180 | 3,042 | 3,171 | 3,414 | 3,451 | 3,756 | 3,960 | 3,360 | 27,262 |

In this experiment, we divide the dataset into 65% training set, 15% validation set, and 20% testing set as we did on the raw data but we change the of epochs to 55. We used the deep learning model Dense-net 161. We use the pre-trained model and change the last layer with the number of classes in the BreakHis dataset. We plot the loss curve in each magnification factor to observe the change in the loss curve in each training and validation phase. Figure 10 shows the loss curve in all magnification factor. Table 7 shows the experimental result on the augmented data.

**Table 7:** The result of each evaluation on augmented data with respect to different magnification factors.

| **Network** | **Criteria** | **Magnification factors** | | | |
|---|---|---|---|---|---|
| | | **40X** | **100X** | **200X** | **400X** |
| Dense-Net 161 | ACC_IL | 97.70 | 95.70 | 93.73 | 92.92 |
| | ACC_PL | 98.68 | 96.65 | 94.75 | 93.72 |
| | Macro_F1 | 98.03 | 95.88 | 93.43 | 92.96 |
| | Micro_F1 | 97.61 | 95.70 | 93.73 | 92.92 |
| | Kappa | 97.24 | 95.04 | 92.81 | 91.75 |

**Table 8:** Precision, Recall and F1-score for augmented data.

| Subtypes | Precision | | | | Recall | | | | F1-score | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 40X | 100X | 200X | 400X | 40X | 100X | 200X | 400X | 40X | 100X | 200X | 400X |
| Ductal carcinoma | 0.87 | 0.97 | 0.87 | 0.80 | 0.94 | 0.90 | 0.88 | 0.94 | 0.90 | 0.93 | 0.87 | 0.86 |
| Lobular Carcinoma | 0.97 | 0.85 | 0.81 | 0.95 | 0.92 | 0.96 | 0.81 | 0.75 | 0.94 | 0.91 | 0.81 | 0.84 |
| Mucinous Carcinoma | 1.00 | 0.97 | 0.99 | 1.00 | 1.00 | 0.98 | 0.92 | 0.94 | 1.00 | 0.97 | 0.96 | 0.97 |
| Papillary Carcinoma | 0.99 | 0.97 | 0.99 | 0.99 | 1.00 | 1.00 | 1.00 | 0.96 | 1.00 | 0.98 | 1.00 | 0.97 |
| Phyllodes Tumor | 1.00 | 1.00 | 0.97 | 0.90 | 1.00 | 0.92 | 0.96 | 0.93 | 1.00 | 0.96 | 0.96 | 0.91 |
| Tubular Adenoma | 1.00 | 1.00 | 1.00 | 0.94 | 1.00 | 0.94 | 0.97 | 1.00 | 1.00 | 0.97 | 0.98 | 0.97 |
| Adenosis | 1.00 | 0.94 | 0.91 | 1.00 | 1.00 | 1.00 | 1.00 | 0.95 | 1.00 | 0.97 | 0.95 | 0.97 |
| Fibroadenoma | 1.00 | 0.95 | 0.95 | 0.90 | 1.00 | 1.00 | 0.94 | 0.98 | 1.00 | 0.98 | 0.94 | 0.93 |

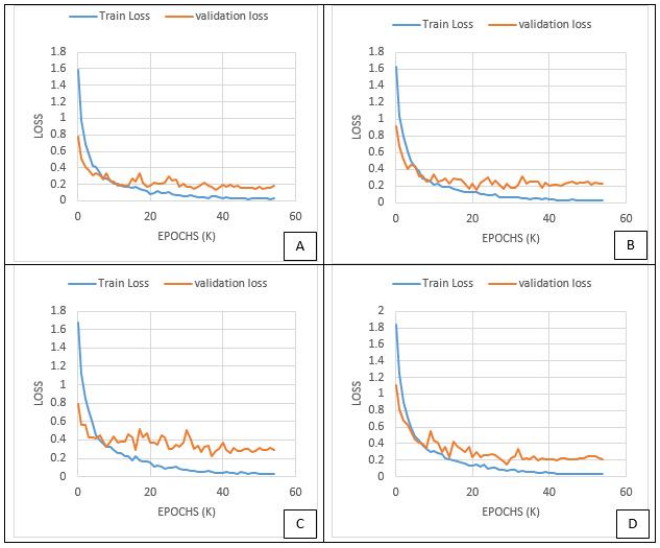


**Figure 10:** A. is the loss curve for 40X, B is the loss curve for 100X, B is the loss curve for 200X, and C is the loss curve for 400X. These loss curves for augmented data with 4 magnification factors.

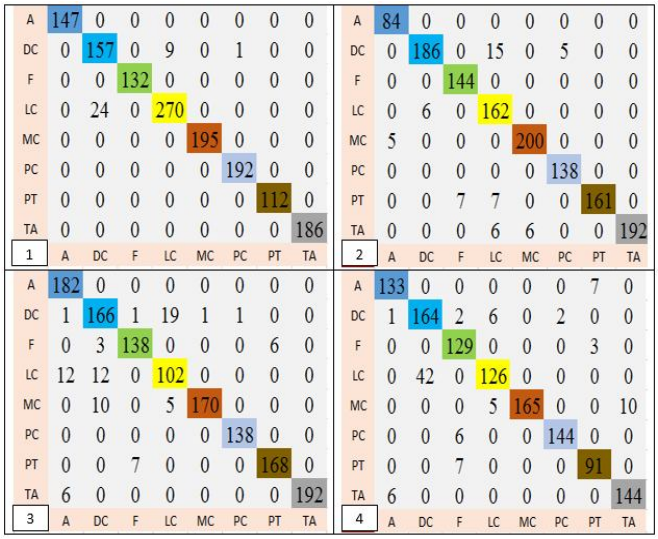


**Figure 11:** The confusion matrix of augmented data for all magnification factors 40X, 100X, 200X, and 400X shown in orders 1, 2, 3, and 4.

The experimental results on augmented data in table 7 show that all evaluation metric on 40X magnification factor (which is indicated by black underline) is better than the other magnification factors. Refer to table 7 and table 3, the classification result on augmented data achieves a better classification result than raw data so it is reliable. From the result of augmented data, we can say that it improves the accuracy and reliability of the diagnosis system. Our augmentation method improves accuracy between 3-4% accuracy in different magnificent factors. According to figure 8, 9, 10, and 11, the loss curves and confusion metrics show that augmented data loss is less than a loss in raw data, and the confusion matrix result on augmented data is better than raw data.

### 5.6 EXPERIMENTAL SUMMARY

This section compares our experimental results with the other experimented carried by other research papers on raw data and augmented data. ACC_IL, ACC_PL are the two terms of comparison. The comparison of raw data is displayed in table 9 and the comparison of augmented data is displayed in table 10.

Our experimental result applied on raw and augmented data, in table 8 and table 9, tells us the evaluation criteria specified (ACC_IL, ACC_PL) achieve the best results by applying a method in training called one fit cycle policy and with a small number of batches and the fewest number of epochs. If we show that in CSDCNN uses 128 batch size and 300 epoch and not achieving the best accuracy and waste of time in the training of the network. On the contrary, we use 32 batch sizes and 90 epochs on raw data and 55 epochs on augmented data, this helped us to reduce the time of training and achieve better accuracy than the other researches.

**Table 9:** Comparison of raw data

| Criteria | Methods | Magnification factor | | | |
|---|---|---|---|---|---|
| | | 40X | 100X | 200X | 400X |
| ACC_IL | CSDCNN+Raw | 89.40 | 90.80 | 88.60 | 87.60 |
| | INV3+Raw | 90.28 | 85.35 | 83.99 | 82.08 |
| | IRV2+Raw | 92.07 | 88.06 | 87.62 | 84.50 |
| | HybridCNN+Raw | 85.70 | 84.20 | 84.09 | 80.10 |
| | Dense-net161+Raw | 93.98 | 91.85 | 90.32 | 89.65 |
| ACC_PL | CSDCNN+Raw | 88.30 | 89.80 | 87.60 | 87.00 |
| | INV3+Raw | 90.44 | 89.05 | 80.63 | 81.08 |
| | IRV2+Raw | 89.11 | 88.45 | 86.07 | 71.42 |
| | HybridCNN+Raw | 85.20 | 83.50 | 84.10 | 79.30 |
| | Dense-net161+Raw | 95.85 | 93.92 | 93.07 | 91.16 |

**Table 10:** Comparison of augmented data

| Criteria | Methods | Magnification factor | | | |
|---|---|---|---|---|---|
| | | 40X | 100X | 200X | 400X |
| ACC_IL | KPCA ensemble+Aug | 88.2 | 84.6 | 83.3 | 83.98 |
| | AlexNet + Aug | 86.4 | 75.8 | 72.6 | 84.6 |
| | CSDCNN+Aug | 92.8 | 93.9 | 93.7 | 92.9 |
| | Dense-net161+Aug | 97.7 | 95.7 | 93.7 | 92.92 |
| ACC_PL | KPCA ensemble+Aug | - | - | - | - |
| | AlexNet + Aug | 74.6 | 73.8 | 76.4 | 79.2 |
| | CSDCNN+Aug | 94.1 | 93.2 | 94.7 | 93.5 |
| | Dense-net161+Aug | 98.7 | 96.65 | 94.75 | 93.52 |

## 6. CONCLUSION AND FUTURE WORK

This paper proposed a method for breast cancer type classification. DenseNet-161 is deployed in this work for classification purposes. Besides, this proposed method utilizes one fit cycle policy to reduce the number of cycles used to train CNN and improve the accuracy of the proposed method. The experiment was conducted on the BreakHis dataset and the result is evaluated using different performance metrics. The proposed model is tested against the following metrics: image level, patient level, Kappa, Macro-f1, and Micro-f1. Results show the best accuracy achievement in raw data on 40X magnification factor is 93.98 and 40X in augmented data is 97.7 as image-level in multiclass classification problem. As a result, the experiments summary concludes that this work outperforms the state-of-art studies in terms of image level, patient level, Kappa, Macro-f1, and Micro-f1. In the future, we will try to study color distribution in histopathological images and apply augmentation in another way to improve testing accuracy.

## REFERENCES


1. Alom, Md Zahangir, Chris Yakopcic, Mst Shamima Nasrin, Tarek M. Taha, and Vijayan K. Asari. **Breast cancer classification from histopathological images with inception recurrent residual convolutional neural network.** *Journal of digital imaging 32*, no. 4: pp. 605-617, 2019.

2. Bray, Freddie, Jacques Ferlay, Isabelle Soerjomataram, Rebecca L. Siegel, Lindsey A. Torre, and Ahmedin Jemal. **Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.** *CA: a cancer journal for clinicians 68*, no. 6: pp. 394-424, 2018.

3. Al Rahhal, M. M. **Breast cancer classification in histopathological images using convolutional neural network.** *Breast Cancer,* 9(3), 2018.

4. Bardou, Dalal, Kun Zhang, and Sayed Mohammad Ahmad. **Classification of breast cancer based on histology images using convolutional neural networks.** *IEEE Access* 6: pp. 24680-24693, 2018.

5. Horvat, Joao V., Delia M. Keating, Halio Rodrigues-Duarte, Elizabeth A. Morris, and Victoria L. Mango. **Calcifications at digital breast tomosynthesis: imaging features and biopsy techniques.** *RadioGraphics* 39, no. 2: pp. 307-318, 2019.

6. Das, Kausik, Sailesh Conjeti, Abhijit Guha Roy, Jyotirmoy Chatterjee, and Debdoot Sheet. **Multiple instance learning of deep convolutional neural networks for breast histopathology whole slide classification.** *In 2018 IEEE 15th International Symposium on Biomedical Imaging* (ISBI 2018), pp. 578-581. IEEE, 2018.

7. Aswathy, M. A., and M. Jagannath. **Detection of breast cancer on digital histopathology images: Present status and future possibilities.** *Informatics in Medicine Unlocked 8*: pp. 74-79, 2017.

8. He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. **Delving deep into rectifiers: Surpassing human-level performance on imagenet classification.** *In Proceedings of the IEEE international conference on computer vision*, pp. 1026-1034. 2015.

9. Xie, Juanying, Qi Hou, Yinghuan Shi, Lv Peng, Liping Jing, Fuzhen Zhuang, Junping Zhang, Xiaoyang Tang, and Shengquan Xu. **The automatic identification of butterfly species.** *arXiv preprint arXiv*:1803.06626 (2018).

10. Gulshan, Varun, Lily Peng, Marc Coram, Martin C. Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan et al. **Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs.** *Jama 316*, no. 22: pp. 2402-2410, 2016.

11. Esteva, Andre, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau, and Sebastian Thrun. **Dermatologist-level classification of skin cancer with deep neural networks.** *nature 542*, no. 7639: pp. 115-118, 2017.

12. Lowe, David G. **Object recognition from local scale-invariant features.** *In Proceedings of the seventh IEEE international conference on computer vision*, vol. 2, pp. 1150-1157. *Ieee*, 1999.

13. Haralick, Robert M., Karthikeyan Shanmugam, and Its' Hak Dinstein. **Textural features for image classification.** *IEEE Transactions on systems, man, and cybernetics 6:* pp. 610-621, 1973.

14. Smith, Leslie N. **A disciplined approach to neural network hyper-parameters: Part 1--learning rate, batch size, momentum, and weight decay.** *arXiv preprint arXiv:*1803.09820 (2018).

15. Jastrzębski, Stanisław, Devansh Arpit, Nicolas Ballas, Vikas Verma, Tong Che, and Yoshua Bengio. **Residual connections encourage iterative inference.** *arXiv preprint arXiv:*1710.04773 (2017).

16. Pan, Sinno Jialin, and Qiang Yang. **A survey on transfer learning.** *IEEE Transactions on knowledge and data engineering 22*, no. 10: pp. 1345-1359, 2009.
17. Jiang, Yun, Li Chen, Hai Zhang, and Xiao Xiao. **Breast cancer histopathological image classification using convolutional neural networks with small SE-ResNet module.** *PloS one 14*, no. 3: e0214587, 2019.
18. Bardou, Dalal, Kun Zhang, and Sayed Mohammad Ahmad. **Classification of breast cancer based on histology images using convolutional neural networks.** *IEEE Access 6:* pp. 24680-24693, 2018.
19. Kassani, Sara Hosseinzadeh, Peyman Hosseinzadeh Kassani, Michal J. Wesolowski, Kevin A. Schneider, and Ralph Deters. **Classification of histopathological biopsy images using ensemble of deep learning networks.** *arXiv preprint arXiv*:1909.11870 (2019).
20. Zhu, Chuang, Fangzhou Song, Ying Wang, Huihui Dong, Yao Guo, and Jun Liu. **Breast cancer histopathology image classification through assembling multiple compact CNNs.** *BMC medical informatics and decision making 19*, no. 1: 198, 2019.
21. Spanhol, Fabio A., Luiz S. Oliveira, Caroline Petitjean, and Laurent Heutte. **A dataset for breast cancer histopathological image classification.** *IEEE Transactions on Biomedical Engineering 63*, no. 7: pp. 1455-1462, 2015.
22. Spanhol, Fabio Alexandre, Luiz S. Oliveira, Caroline Petitjean, and Laurent Heutte. **Breast cancer histopathological image classification using convolutional neural networks.** *In 2016 international joint conference on neural networks (IJCNN)*, pp. 2560-2567. IEEE, 2016.
23. Chan, Alan, and Jack A. Tuszynski. **Automatic prediction of tumour malignancy in breast cancer with fractal dimension.** *Royal Society open science 3*, no. 12: 160558, 2016.
24. Spanhol, Fabio A., Luiz S. Oliveira, Paulo R. Cavalin, Caroline Petitjean, and Laurent Heutte. **Deep features for breast cancer histopathological image classification.** *In 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 1868-1873. IEEE, 2017.
25. Han, Zhongyi, Benzheng Wei, Yuanjie Zheng, Yilong Yin, Kejian Li, and Shuo Li. **Breast cancer multi-classification from histopathological images with structured deep learning model.** *Scientific reports 7*, no. 1: pp. 1-10, 2017.
26. Song, Yang, Ju Jia Zou, Hang Chang, and Weidong Cai. **Adapting fisher vectors for histopathology image classification.** *In 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pp. 600-603. IEEE, 2017.
27. Kahya, Mohammed Abdulrazaq, Waleed Al-Hayani, and Zakariya Yahya Algamal. **Classification of breast cancer histopathology images based on adaptive sparse support vector machine.** *Journal of Applied Mathematics and Bioinformatics 7*, no. 1: 49, 2017.
28. Das, Kausik, Sri Phani Krishna Karri, Abhijit Guha Roy, Jyotirmoy Chatterjee, and Debdoot Sheet. **Classifying histopathology whole-slides using fusion of decisions from deep convolutional network on a collection of random multi-views at multi-magnification.** *In 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pp. 1024-1027. IEEE, 2017.
29. Zhi, Weiming, Henry Wing Fung Yueng, Zhenghao Chen, Seid Miad Zandavi, Zhicheng Lu, and Yuk Ying Chung. **Using transfer learning with convolutional neural networks to diagnose breast cancer from histopathological images.** *In International Conference on Neural Information Processing*, pp. 669-676. Springer, Cham, 2017.
30. Benhammou, Yassir, Siham Tabik, Boujemâa Achchab, and Francisco Herrera. **A first study exploring the performance of the state-of-the art CNN model in the problem of breast cancer.** *In Proceedings of the International Conference on Learning and Optimization Algorithms: Theory and Applications*, pp. 1-6. 2018.
31. Gandomkar, Ziba, Patrick C. Brennan, and Claudia Mello-Thoms. **MuDeRN: Multi-category classification of breast histopathological image using deep residual networks.** *Artificial intelligence in medicine 88*, pp. 14-24, 2018.
32. Huang, Gao, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger. **Densely connected convolutional networks.** *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700-4708. 2017.
33. Spanhol, Fabio Alexandre, Luiz S. Oliveira, Caroline Petitjean, and Laurent Heutte. **Breast cancer histopathological image classification using convolutional neural networks.** *In 2016 international joint conference on neural networks (IJCNN)*, pp. 2560-2567. IEEE, 2016.
34. Jiang, Yun, Li Chen, Hai Zhang, and Xiao Xiao. **Breast cancer histopathological image classification using convolutional neural networks with small SE-ResNet module.** *PloS one 14*, no. 3 (2019): e0214587.
35. Bardou, Dalal, Kun Zhang, and Sayed Mohammad Ahmad. **Classification of breast cancer based on histology images using convolutional neural networks.** *IEEE Access 6*, pp. 24680-24693, 2018.
36. Benhammou, Yassir, Boujemâa Achchab, Francisco Herrera, and Siham Tabik. **BreakHis based breast cancer automatic diagnosis using deep learning: Taxonomy, survey and insights.** *Neurocomputing 375*, pp. 9-24, 2020.
37. Li, Hua, Shasha Zhuang, Deng-ao Li, Jumin Zhao, and Yanyun Ma. **Benign and malignant classification of mammogram images based on deep learning.** *Biomedical Signal Processing and Control 51*, pp. 347-354, 2019.
38. Kassani, Sara Hosseinzadeh, and Peyman Hosseinzadeh Kassani. **A comparative study of deep learning**

**architectures on melanoma detection.** *Tissue and Cell 58*, pp. 76-83, 2019.

39. Smith, Leslie N**. Cyclical learning rates for training neural networks.** *In 2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 464-472. IEEE, 2017.
40. Smith, Leslie N., and Nicholay Topin. **Super-convergence: Very fast training of neural networks using large learning rates.** *In Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications*, vol. 11006, p. 1100612. International Society for Optics and Photonics, 2019.
41. https://colab.research.google.com/, accessed on (10.December -2020).