Correspondence to:

M. S. Miah and M. Alibakhshikenari,
sipon@ict.iu.ac.bd;
mohammad.alibakhshikenari@uc3m.es

Breast Cancer Detection Based on Simplified Deep Learning Technique With Histopathological Image Using BreaKHis Database

Tania Afroz Toma[1], Shivazi Biswas[1,2], Md Sipon Miah[1,3,4] , Mohammad Alibakhshikenari[3] , Bal S. Virdee[5] , Sandra Fernando[5], Md Habibur Rahman[6], Syed Mansoor Ali[7], Farhad Arpanaei[8], Mohammad Amzad Hossain[9], Md Mahbubur Rahman[1], Ming-bo Niu[4] , Naser Ojaroudi Parchin[10], and Patrizia Livreri[11]

[1]Department of Information and Communication Technology, Islamic University, Kushtia, Bangladesh, [2]Machine Learning Engineer, Cefalo Bangladesh Limited, Dhaka, Bangladesh, [3]Department of Signal Theory and Communications, Universidad Carlos III de Madrid, Leganes, Spain, [4]Internet of Vehicle to Road Research Institute, Chang'an University, Shaanxi, China, [5]Centre for Communications Technology, School of Computing & Digital Media, London Metropolitan University, London, UK, [6]Department of Computer Science and Engineering, Islamic University, Kushtia, Bangladesh, [7]Department of Physics and Astronomy, College of Science, King Saud University, Riyadh, Saudi Arabia, [8]Department of Telematic Engineering, Universidad Carlos III de Madrid, Leganés, Spain, [9]Department of Information and Communication Engineering, Noakhali Science and Technology University, Noakhali, Bangladesh, [10]School of Engineering and the Built Environment, Edinburgh Napier University, Edinburgh, UK, [11]Department of Engineering, University of Palermo, Palermo, Italy

Abstract Presented here are the results of an investigation conducted to determine the effectiveness of deep learning (DL)-based systems utilizing the power of transfer learning for detecting breast cancer in histopathological images. It is shown that DL models that are not specifically developed for breast cancer detection can be trained using transfer learning to effectively detect breast cancer in histopathological images. The outcome of the analysis enables the selection of the best DL architecture for detecting cancer with high accuracy. This should facilitate pathologists to achieve early diagnoses of breast cancer and administer appropriate treatment to the patient. The experimental work here used the BreaKHis database consisting of 7909 histopathological pictures from 82 clinical breast cancer patients. The strategy presented for DL training uses various image processing techniques for extracting various feature patterns. This is followed by applying transfer learning techniques in the deep convolutional networks like ResNet, ResNeXt, SENet, Dual Path Net, DenseNet, NASNet, and Wide ResNet. Comparison with recent literature shows that ResNext-50, ResNext-101, DPN131, DenseNet-169 and NASNet-A provide an accuracy of 99.8%, 99.5%, 99.675%, 99.725%, and 99.4%, respectively, and outperform previous studies.

1. Introduction

Breast cancer is the most common and deadly disease among women worldwide. The statistics, released in December 2020 (WHO, 2021) by the International Agency for Research on Cancer, states that breast cancer is the most prevalent disease diagnosed amongst women in the world. In fact, breast cancer has become even more predominant than lung cancer. According to Global Cancer Statistics, in 2020 about 2.3 million new cases of breast cancer resulted in 685,000 deaths. Cancer begins when the cells in the breast tissue start to grow and multiply in an uncontrolled manner and spread into surrounding healthy tissues. Breast cancer cells invade healthy breast tissue first, then spread to distant body parts. Cancer cells spread to other parts of the body via the lymph nodes.

Medical images can be obtained using a variety of techniques, including MRI scans, mammography, ultrasound, thermography, computed tomography scans, and histopathology. Among these methods, the histopathology test is the gold standard for the clinical diagnosis of cancers (Gurcan et al., 2009; Lakhani et al., 2012). Histopathology involves taking a biopsy which is colored by a staining agent, typically hematoxylin and eosin (Veta et al., 2014). The stained tissue is finely sliced by a pathologist who then uses a high-resolution microscope to examine the cell structure and the presence of cancer cells (Spanhol et al., 2016). This manual procedure must be repeated to ensure reliability, which is highly time-consuming. Moreover, histological images obtained by different pathologists may also vary.

Radio Science

Resources: Sandra Fernando, Md Habibur Rahman, Mohammad Amzad Hossain
Software: Tania Afroz Toma, Shivazi Biswas
Supervision: Md Sipon Miah, Md Mahbubur Rahman
Validation: Shivazi Biswas, Md Sipon Miah, Mohammad Alibakhshikenari, Bal S. Virdee, Sandra Fernando, Md Habibur Rahman, Syed Mansoor Ali, Farhad Arpanaei, Mohammad Amzad Hossain, Md Mahbubur Rahman, Ming-bo Niu, Naser Ojaroudi Parchin, Patrizia Livreri
Visualization: Md Sipon Miah, Bal S. Virdee, Sandra Fernando, Md Mahbubur Rahman, Ming-bo Niu, Patrizia Livreri
Writing – original draft: Tania Afroz Toma, Shivazi Biswas
Writing – review & editing: Md Sipon Miah, Mohammad Alibakhshikenari, Bal S. Virdee, Sandra Fernando, Md Habibur Rahman, Syed Mansoor Ali, Farhad Arpanaei, Mohammad Amzad Hossain, Md Mahbubur Rahman, Ming-bo Niu, Naser Ojaroudi Parchin, Patrizia Livreri

Digital histopathological images can be collected from microscopic examination using a camera. This allows the development of automatic computer-aided diagnosis (CAD) of breast cancer detection using these images. CAD can play a significant role in assisting pathologists by improving the accuracy of the cancer diagnosis. Many researchers are exploring opportunities to improve the performance of CAD systems (Aswathy & Jagannath, 2017; Dromain et al., 2013; Spanhol et al., 2016; Veta et al., 2014). Due to a large amount of inter-intra class variations in histopathological images and the limitations of feature extraction methods, the conventional computerized diagnostic approaches are not able to produce very reliable classification results.

Several studies aiming to resolve the above-mentioned issues are reported in LeCun et al. (2015). These studies use an advanced computational model based on multiple layers of the neural network, that is, deep learning (DL). DL is inspired by the structure and function of the human brain's neural networks. While it is not an exact mimicry of the brain, it does draw some inspiration from the brain's information processing capabilities. It is a sub-field of machine learning that utilizes multiple layers of neural network architectures to perform several imaging tasks. DL models (Dimitriou et al., 2019; Litjens et al., 2017) have a high success rate in computer vision, object detection, classification in medical image processing, etc. due to their ability to learn complicated tasks and extract different features automatically from images.

DL models have the advantage of not requiring prior knowledge about the data. DL just requires the input data to be in a proper format and the network parameters relevant to the problem at hand be specified. Because DL-based cutting-edge technology provides excellent performance many researchers are focusing their attention of using DL on the area of breast cancer detection. Many studies have been published by various scholars where different algorithms have been proposed to detect and classify breast cancer. However, the performance reported using these algorithms is limited (Gurcan et al., 2009; Spanhol et al., 2016; Veta et al., 2014). Recently, research has been performed on the classification issue of identifying breast cancer using machine learning and DL techniques (Aswathy & Jagannath, 2017). This study shows there is still considerable room to improve the quality of the automatic detection of DL methods.

In this paper a strategy is proposed for accurately detecting breast cancer from histopathological images. This was achieved by: (a) exploiting the advantages offered by transfer learning convolutional neural network (CNN) models, that is, using knowledge learned from one task to adapt to a different but related task, and leveraging the generalization power of the pre-trained model and improve its performance on the new task by fine-tuning these pre-trained models on a specific data set; (b) performing the investigation using BreaKHis data set that consisted of 2,480 benign images and 5,429 malignant images in four magnification factors (40×, 100×, 200×, and 400×); and (c) preprocessing the images prior to application to the models. Using the proposed strategy, the DL training models are shown to outperform the existing breast cancer detection algorithms reported in the literature regardless of the image magnification factor.

The rest of the paper is organized as follows: Section 2 describes the related work. Section 3 on data and methods describes the several DL-based approaches for automatically identifying breast cancer using histopathological images. The experimental results are presented in Section 4. Sights from this work are discussed in Section 5, and this is followed by the conclusion and future work in Section 6.

2. Related Work

Over the last decade, a lot of research has been done to detect breast cancer from histological images. Much of this work has focused on utilizing CAD to identify the two basic forms of breast cancer (benign and malignant). Various machine learning techniques have been used to diagnose breast cancer. Several research papers outlined below on detection of breast cancer use DL methods and CNN for classification.

The categorization of carcinoma and non-carcinoma histopathological images of breast tissue by Hameed et al. (2020) used four distinct DL models: fully trained VGG-16, fully trained VGG-19, fine-tuned VGG-16, and fine-tuned VGG19. Among these, an ensemble of fine-tuned VGG-16 and VGG-19 models provided an average accuracy of 95.29% and F1 score of 95.29%. The group of fine-tuned models show an overall sensitivity of 97.73% and an F1 score of 95.29% for the carcinoma class.

K. Gupta and Chawla (2020) suggested a two-phase model where four pretrained DL models (VGG-16, VGG-19, ResNet-50, and Xception) are utilized to extract features on the BreaKHis data set. Here two common machine

learning classifiers, namely support vector machine and logistic regression, are used for the classification. They achieved maximum accuracy of 92.5% for the ResNet-50 model. Their approach provided a better performance on 100× and 40× magnification as compared to the others.

Likewise, in a study by Lim et al. (2018), two CNN models, that is, VGG-16 and Inception-V3, utilize a transfer learning technique. They applied several data augmentation techniques like rotation, flip, shift, and zooming on the BreaKHis database. The two types of learning used by each model are custom layer and whole layer learning. With the Inception-V3 model, 98% accuracy was achieved, and with the VGG-16 model, 97% accuracy was achieved. It was also shown that the whole layer learning method gives better performance than the custom layer.

Jiang et al. (2019) proposed a new model of a neural network (BHCNet), which contains a small SE-ResNet module requiring fewer parameters. They designed the Gauss error scheduler, that resolves the problem of fine-tuning the stochastic gradient descent (SGD) algorithm's learning rate parameter. The model was trained from scratch and provided satisfactory results. For binary classification, the model had an average accuracy score between 98.87% and 99.34%. For multi-classification, the performance was between 90.66% and 93.81%.

Haija and Adebanjo (2020) authors take advantage of the power of transfer learning by using a CNN model (ResNet-50) that was pre-trained on ImageNet. BreaKHis data set was used in this study where 75% of the data was used for training and 25% used for testing. Their model outperforms other training on the same data sets with an accuracy rate of 99%.

Multi-class breast cancer classification using the DL-based CNN model was conducted in Nawaz et al. (2018). In this experiment, the BreaKHis database was used, and the classification result was reported as patient-wise and image-wise classification. By fine-tuning the DenseNet CNN model, the accuracy of image classification was 95.4%, and for patient classification the accuracy was 96.48%.

A combined network called improved inception-residual CNN (IRRCNN) was suggested by Alom et al. (2019) for classifying breast cancer histopathological images. The IRRCNN is a combination of powerful deep CNN (DCNN) networks, namely the residual network (ResNet), the inception network (Inception-V4), and the recurrent convolutional neural network. Compared to any of these individual models, a combined model is shown to provide superior results. The model was tested on two separate public data sets, that is, the BreaKHis data set and the breast cancer classification challenge, 2015. The testing accuracy of the model for binary and multi-class classification reported is 99.05% and 98.59%, respectively.

In (Kassani et al., 2019), author proposed a DL-based algorithm where a DCNN descriptor and pooling operation are used. In their experiment, they processed the data using popular stain normalization techniques proposed by Macenko et al. (2009) and Reinhard et al. (2001). They trained five pre-trained DL networks, that is, Inception-V3, InceptionResNetV2, Xception, VGG-16, and VGG19. The best performance was achieved by Xception with an average classification accuracy of 92.50%.

Several fine-tuned pre-trained deep neural networks were used by Motlagh et al. (2018) on two public databases, namely BreaKHis and Tissue Micro Array (TMA) databases. Pre-processing techniques, such as normalization and color distortion, were employed. Different pre-trained models, like Inception (V1, V2, V3, and V4) and ResNet (V1 50, V1 101, and V1 152) were investigated. For binary and multi-class classification the ResNet V1 152 model was able to achieve an accuracy of 98.7% and 96.4%, respectively. Among the ResNet models, ResNet-152 provides the highest accuracy, which indicates that the deep layer network is important for classification.

Deep convolutional neural networks, that is, Inception V3 and Inception ResNet V2, trained with transfer learning techniques were used by the authors in J. Xie et al. (2019) as a method for examining the histological images of breast cancer. To do a clustering analysis on the images, they developed a new autoencoder network that converted the features acquired by Inception ResNet V2 to low dimensional space. It was found that clustering resulting from the autoencoder network outperforms the Inception ResNet V2 network.

According to the studies described above, it is obvious that DL models are very useful in the field of image-based detection applications. In certain cases, it has become an integral part of the routine for pathologists and doctors in clinical practices. Despite its significant success in medical imaging, automated breast cancer diagnosis based on histological images is still a long way from being a practical application. This is because a massive amount of label data is required, which is not yet available in this application domain. The main reason is because the annotation of a data set is time-consuming and costly. In the present paper, we compare several advanced DL

Table 1
Distribution of Images in BreaKHis Data Set

Types	Sub-types	40×	100×	200×	400×
Benign	A	114	113	111	106
	F	253	260	264	237
	PT	109	121	108	115
	TA	149	150	140	130
Malignant	DC	864	903	896	788
	LC	156	170	163	137
	MC	205	222	196	169
	PC	145	142	135	138
Total		1995	2081	2013	1820

methods utilizing the power of transfer learning for detecting breast cancer in histopathological images.

## 3. Data and Methods

3.1. Data Sets

The work presented in this paper was conducted on a pathology data set, that is, the breast cancer histopathological image classification (BreaKHis) (Spanhol et al., 2016). This data set has been gathered from the result of a survey by P D Lab, Brazil over the period between January 2014 and December 2014. This data set consists of a total of 7909 histopathology images from 82 clinical breast cancer patients. The data set has 2480 benign images and 5429 malignant images of different magnification factors (40×, 100×, 200×, and 400×). Additionally, benign images have four different subsets: adenosis (A), fibroadenoma (F), phyllodes tumor, and tubular adenoma. The malignant images have four different subsets: ductal carcinoma, lobular carcinoma, mucinous carcinoma and papillary carcinoma. All images are in RGB mode with a resolution of 700×460 pixels. Table 1 shows the distribution of the images in the BreaKHis data set (Spanhol et al., 2016).

3.2. Data Pre-Processing

The standardization of the images in the data set using DL requires the data set to be noiseless, clean, and understandable. Hence it was necessary to pre-process the breast cancer histopathology images in the data set before using DL to improve the overall data quality and removing redundant attributes from the images. The various stages in the pre-processing used were:

- Image rotation where the images were randomly rotated from −45 to +45°. Uniform distribution was used to generate random angles between the specified range with equal probability. The image rotation step produces a single rotated image for each original image. Photographs of the biological tissue are taken from a microscope. These histopathological images are not always perfectly aligned and may be skewed at an angle. The angle range between −45 and 45° was chosen as extremes of image misalignment. This choice in angle helped the model to become more robust in relation to the changes in the direction of objects.
- Random cropping of some portion of the image to a specified size. This was done using the uniform distribution of PyTorch's "RandomResizedCrop" transform library. A single cropped image is generated for each single input image. The images were resized to 224×224 pixels for every model in our experiment except for InceptionV3, which requires the fixed size input of 299×299 pixels. This is to ascertain a meaningful object in the image that may be oriented in a different position.
- Flipping the images vertically and horizontally to make the model more resilient. This step produces a single flipped image for each input image.
- Conversion of all images to a tensor data structure and application of parallel processing at the training and inferencing phase.
- Resizing the images before using the data set in the model to align the size with the requirement of model. It should be noted that during the image resizing process, all images were converted to the required fixed size and then applied to the model for training. In the data transformation the total size of the data was unchanged to enhance the model's ability to learn meaningful features from the available data and improve its performance on unseen samples. The consequence of this was significantly reduced complexity and therefore less time required for training. The data set was transformed using the PyTorch library (Paszke et al., 2019), specifically version 2.0.1, which is compatible with Compute Unified Device Architecture (CUDA) 11.8.
- Normalization and standardization were applied to scale down the input values in the data set to a nominal range between zero to one. This was important because the images may have a different range of features that may affect the results of the model (Yuan & Suh, 2018). The input images were normalized separately to scale down the features between 0 and 1. In the present study, the mean values were [0.485, 0.456, 0.406] and the standard deviation values were [0.229, 0.224, 0.225]. Normalization helps the network to keep the weights close to zero, which makes the network more stable for back propagation. Examples of the original and the processed images are shown in Figure 1.
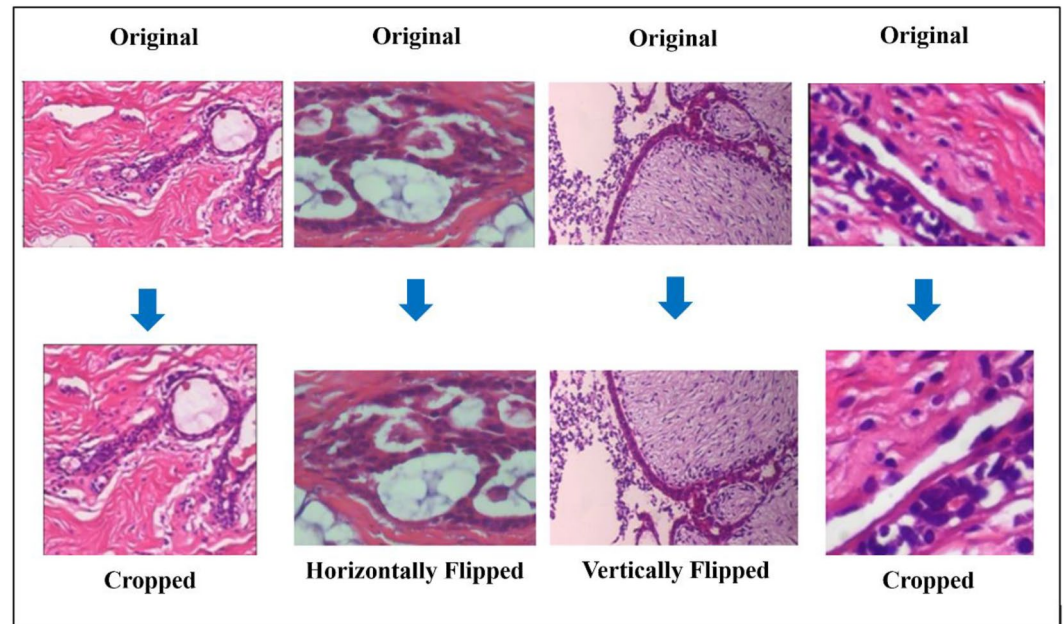
Figure 1. Original images and the processed images.

3.3. Deep Convolutional Neural Network

Conventional neural network (CNN) is a type of DL architecture that has become very popular in various computer vision tasks such as image classification, face recognition, and action recognition. It is designed to extract the most significant features from any visual imagery. Basically, CNN has three main characteristics that give it an advantage in comparison to other learning algorithms. These characteristics are (a) sparse interactions, (b) parameter sharing, and (c) equivariant representations. In medical science, most researchers are using the DL-based CNN model because of its outstanding performance in image classification.

Generally, CNN architecture is composed of three layers: (a) convolution layers, (b) pooling layers, and (c) fully connected layers. The convolution layers and pooling layers are essential components for feature extraction and the fully connected layers are essential for the final classifications. A standard CNN architecture consists mainly of a convolution layer and a pooling layer repeatedly, with one or more fully connected layers. The basic architecture of CNN is depicted in Figure 2.

Convolution layer is a building block of CNN that performs several linear and nonlinear operations. Convolution is a linear operation where the kernel (filter) is applied to the input tensor for feature extraction. The kernel is a set of weights in a matrix used to extract a particular feature from an image. A direct element-wise product is calculated between the kernel matrix elements and input tensor elements and then summed up for obtaining a feature map. Moreover, multiple kernels are applied repeatedly to extract different types of features from the input tensor. For example, if the input image is denoted by I and the kernel by K, then the convolution operation is given by

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(m, n) K(i - m, j - n) \tag{1}$$

where i denotes the indexes of rows of the result matrix, and j denotes the indexes of columns of the result matrix.

While implementing the convolution operation, the output dimensions are not the same as the input dimensions. The consequence of this is loss of some border data. To perfectly fit the input image entails adding or padding zeros around the edges of the image. Considering the padding and stride values, the dimensions of the output matrix can be calculated by

$$n_{out} = \left\lceil \frac{n_{in} + 2p - f}{s} + 1 \right\rceil \tag{2}$$
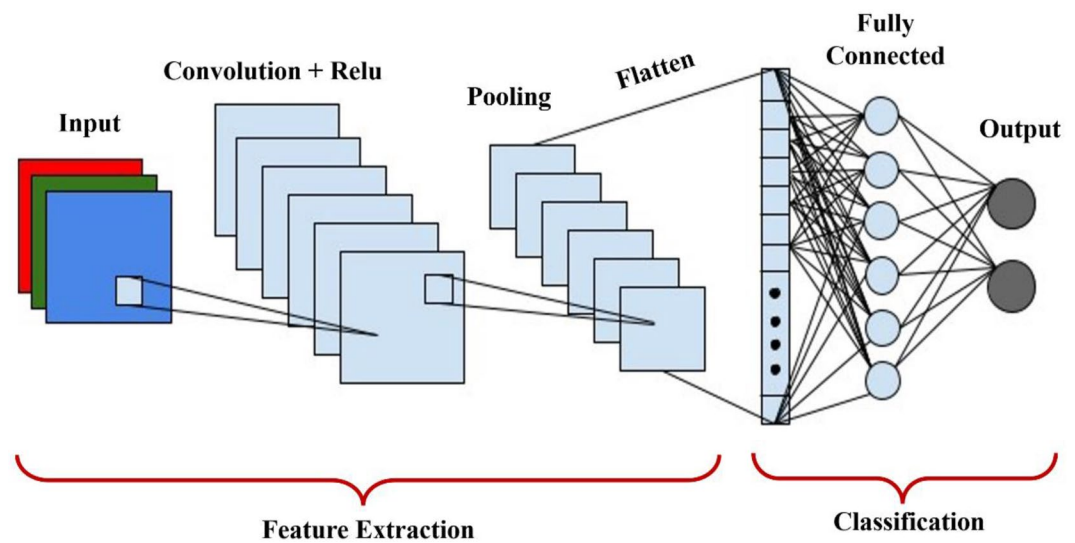
Figure 2. The basic architecture of the convolutional neural network model.

where the padding is denoted by p, while the stride and the kernel are denoted by s and f, respectively.

A pooling layer is another important block of CNN. After the convolution operation, each image size is decreased, and the features map size is increased. The main idea of using the pooling operation is to reduce the dimension of the features map and down-sample the volume of the network. Furthermore, stride and padding also affect the network because they help the model to change or preserve the size of the image.

After successfully performing the convolution and pooling operations, the model will have learnt several important features from the input. The output is then flattened and fed to a fully connected layer for final classification. After completing the forward phase, which is based on predicted output and actual output, the loss function is calculated. Using the gradients, the back propagation is applied for updating the weights. The generic formula used for updating the values of parameter is given by

$$X_{k+1} = X_k - [\in \times \nabla f(X_k)] \qquad (3)$$

where X_{k+1} denotes the new parameter, X_k denotes the old parameter, ϵ denotes the learning rate, and ∇ denotes the Gradient operator. It was found that over a series of iterations, this model gains the ability to reduce the error rate and gradually increase its performance. In such a way, the model can be trained to perform the classifications more robustly.

Several different techniques exist in the practice for dealing with imbalanced data set. The naivest class of techniques is sampling: changing the data presented to the model by under-sampling common classes, oversampling (duplicating) rare classes, or both. It should be noted that under-sampling shrinks the data size, therefore less time is necessary for learning. The disadvantage of this is that this is likely to lead to losing useful information. On the other hand, over-sampling is used to add minority examples to the data set to achieve a balance, in which the existing minority examples are replicated.

Experimental work conducted here compares the various pre-trained deep neural networks. It was found that these networks are capable of automatically learning selected features from histopathological images. From this investigation we were able to identify the model that provides the most satisfactory results in the detection of breast cancer.

In the current investigation, we used the ResNet-18, ResNet-34, ResNet-50, ResNet-101, and ResNet-152 models. The models were trained on the 1.2 million images on the data set by ImageNet. To enhance the accuracy of these models, the transfer learning technique in Hosna et al. (2022) was utilized. Also, the breast cancer classifier was adjusted to fit with these networks. The ResNet base model had 1,000 outputs in the last layer but based on the need for this work, the output was changed to two classes, benign and malignant. Adam was used as an optimizer

(Kingma & Ba, 2014) and the learning rate was set to 0.00001 because from numerous experiment trials we conducted it was found that this learning rate was the best.

In deep neural networks, it is observed that increasing the layers of the network can also increase the probability of the error rate. The ResNet model, proposed by He et al. (2016), resolves this problem by addressing a new residual block without degrading accuracy. Residual blocks have a skip connection which helps the model learn the identity function. There are many variants of the ResNet model.

In 2017, (Xie et al., 2017) introduced another new model named ResNeXt. The idea behind this model was to introduce the "splittransform-merge" strategy which uncovered another measurement, called "cardinality." Pre-trained models ResNeXt-50 (32 × 4d), ResNeXt-101 (32 × 4d), and ResNeXt-101 (64 × 4d) were investigated in this work. It was necessary to resize the input images to 224 × 224. Only the classifier layer was trained while keeping all other layers frozen using transfer learning. These fully connected layer models were fine-tuned using perceptrons of 2,048. It was found that the optimizer and learning rate were the same as for the ResNet architecture.

In this investigation, we have used the pre-trained network SENet-154, which is a squeeze and excitation network first introduced by Hu et al. (2018). This network mainly consists of three parts, that is, (a) squeeze module, (b) excitation module, and (c) scale module. The pre-trained model (Alex, 2019) was imported by first installing the cnn_finetune package using the pip manager in Python. The make_model package could then be imported from cnn_finetune. By using fine-tuning, this model was retrained for all layers with the input size of 224 × 224 × 3 and the batch size of 16.

Dense convolution network was designed by a research group (Huang et al., 2017). This model is like the ResNet model, except that ResNet uses summed (+) technique that adds the output of the previous layer (identity) with the next layer whereas DenseNet uses concatenation (.) instead of summation. In this work we investigated the different variants of pre-trained DenseNet networks such as DenseNet-121, DenseNet-161, DenseNet169, and DenseNet-201. In the work here, the classifier with the last linear layer of 1920 perceptions was modified.

Dualpath network architecture was proposed by Chen et al. (2017). It uses both ResNet and DenseNet blocks to create a simple but high-efficient system. The combination of these two strong modules results in a new macro-block, as residual blocks allow feature reuse and DenseNet blocks allows exploration of new features. DpNet-131 is a convolution neural network which we pre-trained on ImageNet and with 79.25 M parameters for 1,000 classes. By applying the fine-tune with two classes in the last linear layer, the number of parameters reduces to 76.57 M. The learning rate was found to be 0.00001. Adam optimizer was set to this model for retraining.

NASNet by Zoph and Le (2016) was inspired by the automated neural architecture search method. The idea behind this model is to use a recurrent neural network (RNN) to generate a network and train it by using reinforcement learning methods. In this model, a new ScheduledDropPath regularization method is utilized which allows the maximum expected performance to be obtained. Moreover, by taking the advantage of controller recurrent neural network and CNN, this model can find variable-length architecture space. In this study, a pretrained NASNet model was fine-tuned and then retrained with the last linear layer of 4,032 perceptrons. Adam optimizer was utilized in this study and 0.5 dropout was applied. Like the SENet model, this model was downloaded from the cnn_finetune package. The input size of the images for this model was set to 331 × 331 × 3.

Despite the cutting-edge performance of deep residual networks on thousands of layers, several layers must be duplicated to improve by a fraction of a percent, which slows down the model efficiency for training. To resolve these challenges, a novel architecture was designed by Zagoruyko and Komodakis (2016), where the depth of residual networks is decreased, and their width is increased. This model is made up of ResNet by adding more feature planes to the convolution layers to widen and increase the size of the filters in the convolution layers. The cutting-edge wide ResNet-50-2 was also investigated. This model provides an outstanding performance in medical imaging. In the present study, this model was fine-tuned with fully connected layers of 2,048 perceptrons.

3.4. Transfer Learning

Transfer learning is a technique in DL where previously acquired knowledge is used to solve related problems. Deep CNN models that have been trained on ImageNet were found to be more capable of achieving satisfactory results. Nowadays, transfer learning is widely used and provides satisfactory results in computer vision tasks. It speeds the model training and enhances the model performance though on limited data. Moreover, transfer learning can solve complex real-world problems and helps to build more robust models which can perform a variety of tasks. The steps followed in our experiment are outlined in Figure 3.
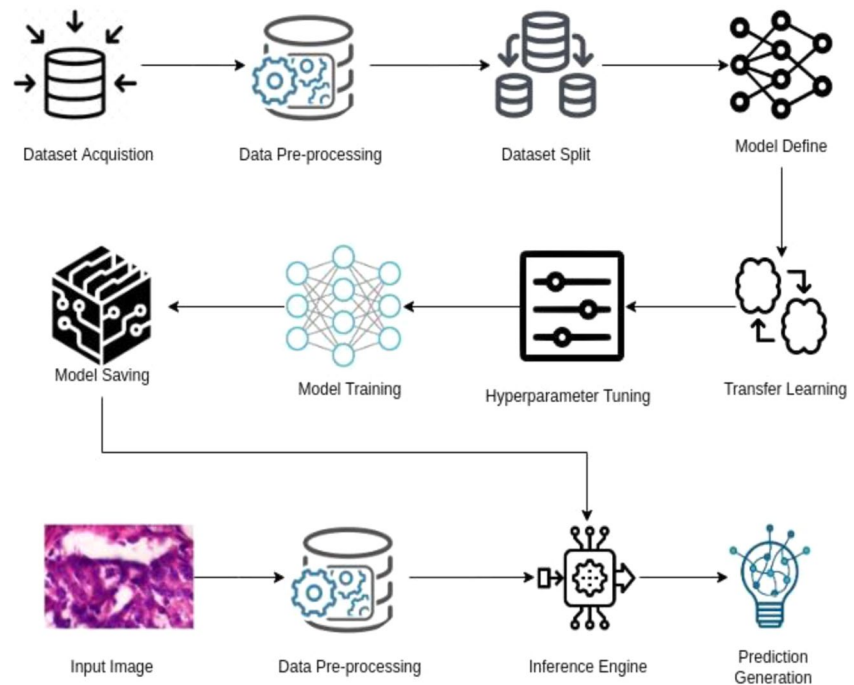
Figure 3. Block diagram of the following methodology.

3.5. Environmental Setup

As DL models needed to handle a huge amount of data it was necessary to use Graphics Processing Unit (GPU). In our experiment, we used colaboratory (known as Colab) cloud service that provides free access to GPU and Tensor Processing Units (TPU) computing. Google colab provided the same runtime as the Jupyter Notebooks, which is mainly configured for DL tasks. Pytorch library, an opensource framework based on the Torch library (Paszke et al., 2019), was used to facilitate training and testing of the DL models.

3.6. Model Training

Before training, the selected models were prepared according to the above-mentioned requirements. All models used were trained for 100 epochs with batches of different sizes and 70% training data. The remaining 20% data was used for validation and 10% for testing. Data pre-processing techniques was used in every model to improve the systems performance. The number of total images was the same after image pre-processing. Adam's method, a variant of stochastic gradient descent, was chosen as an optimizer for the backpropagation of these models. The learning rate was generally set to 0.00001. Dropout method was used for regulating the results in this study. Using google colab, models like ResNeXt-50 ($32 \times 4d$), DenseNet, and wide ResNet-50_2 took about 6 hr to be trained. Other models, such as SENet-154, DpNet-131, NASNet large, and ResNeXt-101 ($64 \times 4d$) took about 20 hr to be trained for completing 100 epochs.

Due to the number of the parameters and input image size we faced some difficulty to retrain SENet-154, DPN-131, ResNeXt50 ($32 \times 4d$), and ResNeXt-101 ($64 \times 4d$) with the batch size of 32. Therefore, the batch size was decreased to 16 instead of 32 to avoid GPU memory errors while training. For the same reason the batch size was set to eight in the NASNet-A (large) model. Finally, during the experiments, the parameters were modified sometimes to avoid computational errors.

3.7. Model Testing

After completing the training phase, each model was saved for future inference. Inference is very significant in DL because it helps bridge the gap between the data the models were trained on and ambiguous information in the real world. Here, a trained model was used to predict the testing samples and a forward pass. It should be noted that image transformation was not applied in the inference phase. Only the image size was changed according to
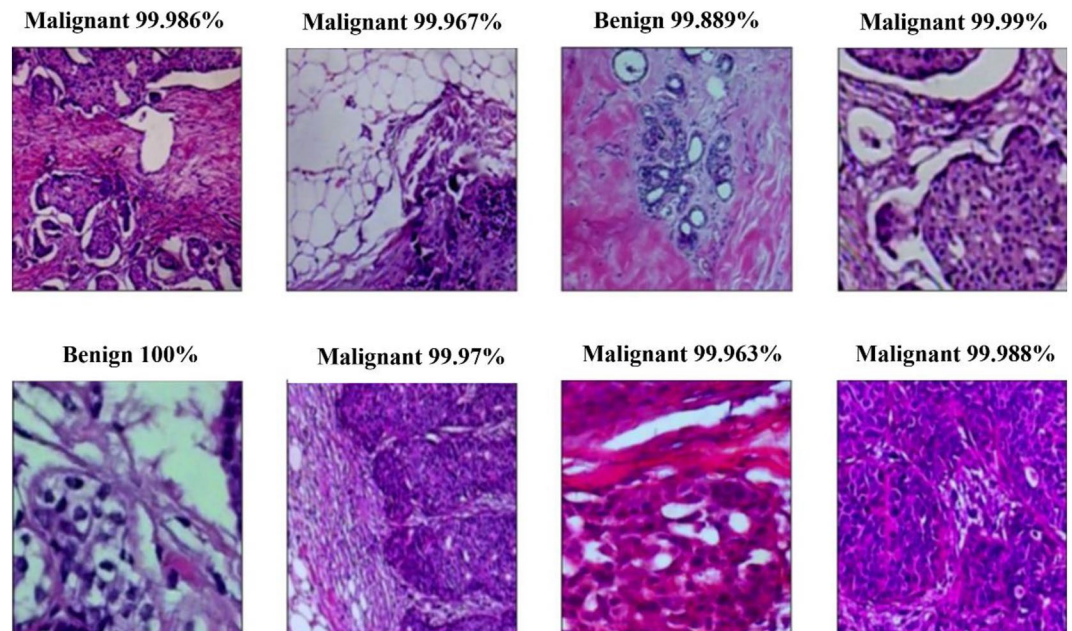
Figure 4. Sample prediction from inference model.

the specific model. For single image our inference generates predictions within 150 milliseconds utilizing K80 GPU. A sample prediction from the inference model is shown in Figure 4.

4. Experimental Results

In this paper, we provide the findings from tests using several deep neural networks to detect breast cancer in histological images. All these pre-trained networks were retrained and tested on the publicly available BreaKHis data set which contains a total of 7,909 images divided into two classes of 2,480 benign images and 5,429 malignant images. The data set was randomly split into three parts: 70% for training, 20% for validation, and 10% for testing. We followed the same pipeline for every model to train and test. To facilitate the comparison of different models, we structured the results from these models given in Table 2 with their accuracy at different magnification magnitudes of the BreaKHis database.

In our experiment, a pre-trained ResNeXt-50 (32 × 4d) network achieved the highest average accuracy score among all examined models. This model was retrained with 70% training data for 100 epoch and for magnification magnitude of 400×. This model obtained the highest accuracy score of 100% in the inference phase which is impressive. The accuracy and loss curves of the ResNeXt-50 (32 × 4d) model for all magnification magnitudes are shown in Figures 5 and 6. Similarly, the accuracy and loss curves of the NAS-Net-N (large) model for all magnification magnitudes, which outperformed all previously reported investigations, are shown in Figures 7 and 8.

The second highest accuracy of 99.75% was obtained with Dense-Net161. ResNet-34, DenseNet-169, and SENet-201 models achieved an average accuracy of 99.725%. In fact, these results indicate a cutting-edge performance in breast cancer histopathological image classification. Another high average accuracy of 99.675% was achieved by DPN131 and Wide ResNet-50_2. The DPN-131 achieved a 100% accuracy score for magnification magnitude of 400×, and the wide ResNet-50_2 achieved accuracy of 99.8% for 100× and 200×. Table 2 shows the remaining variants of the ResNet network provide satisfactory results. ResNet-18, ResNet-50, ResNet-101, and ResNet-152 achieved an accuracy level of 99.5%, 99.425%, 99.52%, and 99.5%, respectively. Variant of the DenseNet architecture, DenseNet-121, achieved accuracy of 99.65%. SENet154 network provided an accuracy of 99.55%, which is 0.85% greater than ResNet V1 152 by Motlagh et al. (2018). ResNeXt-101 (32 × 4d) provided 99.525% accuracy for the binary classifications. The test loss achieved by this model was 0.0010 in an epoch of 94 for magnification magnitude of 200×. Similarly, ResNeXt-101 (64 × 4d) had the same accuracy score as

Table 2
Comparative Analysis to Detect Breast Cancer Using BreaKHis Database

Papers	Model	40×	100×	200×	400×	Accuracy
S. Xie et al. (2019)	Inception-ResNet-V2	0.997	0.989	0.989	0.987	0.98970
Boumaraf et al. (2021)	ResNet-18	–	–	–	–	0.98800
Our examination	ResNet-18	0.998	0.994	0.997	0.991	0.99500
Jannesari et al. (2018)	ResNet-50	–	–	–	–	0.97800
Choudhary et al. (2021)	ResNet-50	–	–	–	–	0.92000
Our examination	ResNet-34	0.998	0.997	0.998	0.996	0.99725
Our examination	ResNet-50	0.997	0.998	0.993	0.989	0.99425
Jannesari et al. (2018)	ResNet-101	–	–	–	–	0.98400
Our examination	ResNet-101	0.997	0.994	0.992	0.998	0.99525
Gandomkar et al. (2018)	ResNet-152	0.985	0.979	0.983	0.976	0.98100
Our examination	ResNet-152	0.993	0.997	0.995	0.995	0.99500
Shahidi et al. (2020)	ResNeXt-101 (32 × 4d)	–	–	–	–	0.99490
Our examination	ResNeXt-50 (32 × 4d)	0.997	0.998	0.997	1	0.99800
Our examination	ResNeXt-101 (32 × 4d)	0.998	0.995	0.995	0.993	0.99525
Shahidi et al. (2020)	ResNeXt-101 (64 × 4d)	–	–	–	–	0.99360
Our examination	ResNeXt-101 (64 × 4d)	0.995	0.994	0.997	0.995	0.99525
Our examination	SENet-154	0.993	0.994	1	0.995	0.99550
Our examination	DPN-131	0.993	0.997	0.997	1	0.99675
Man et al. (2020)	DenseNet-121	–	–	–	–	0.88400
Our examination	DenseNet-121	0.997	0.998	0.998	0.993	0.99650
Our examination	DenseNet-161	0.998	0.998	0.998	0.996	0.99750
V. Gupta and Bhavsar (2018)	DenseNet-169	–	–	–	–	0.94120
Our examination	DenseNet-169	0.997	0.997	0.997	0.998	0.99725
Nawaz et al. (2018)	DenseNet-201	–	–	–	–	0.95400
Zheng et al. (2023)	DenseNet-201	–	–	–	–	0.98300
Our examination	DenseNet-201	0.998	0.997	0.998	0.996	0.99725
Shahidi et al. (2020)	NASNet-A (large)	–	–	–	–	0.99240
Our examination	NASNet-A (large)	0.997	0.994	0.995	0.989	0.99375
Our examination	Wide ResNet-50_2	0.995	0.998	0.998	0.996	0.99675

ResNeXt-101 (32 × 4d), while the test loss achieved by this model was 0.0009 in an epoch of 33 for magnification magnitude of 40×. In our comparative investigation, the NASNet-A (large) model was very accurate too, with an average accuracy of 99.375%, being 0.135% higher than the study reported by Shahidi et al. (2020) in two-class classification. The model gained a 99.7% accuracy score for magnification magnitude of 40×.

The overall performance of these experimental models was measured based on the confusion matrix also referred to as error matrix. This confusion matrix contains four elements, namely, true positive (TP), true negative (TN), false positive (FP), and false negative (FN). In our work, TP refers to images that were correctly predicted as cancerous, and TN refers to non-cancerous images that were correctly predicted. The FP refers to the fact that the model mistakenly classified the non-cancerous images as cancerous, whereas the FN indicates that the model mistakenly classified the cancerous images as non-cancerous. Based on these four basic metrics other metrics can be determined by which the model's performance can be evaluated more robustly. The other three metrics presented here are precision, recall, and F1-score. These three metrics and the confusion matrix of ResNeXt-50 (32 × 4d) and NASNet-N (large) for different magnification factors of 40× 100×, 200×, and 400× are shown in Figures 5–8. The findings of this experimental investigation suggest that ResNext-50 and DenseNet-169 convolutional networks can be used for breast cancer detection because these latest models provide very good
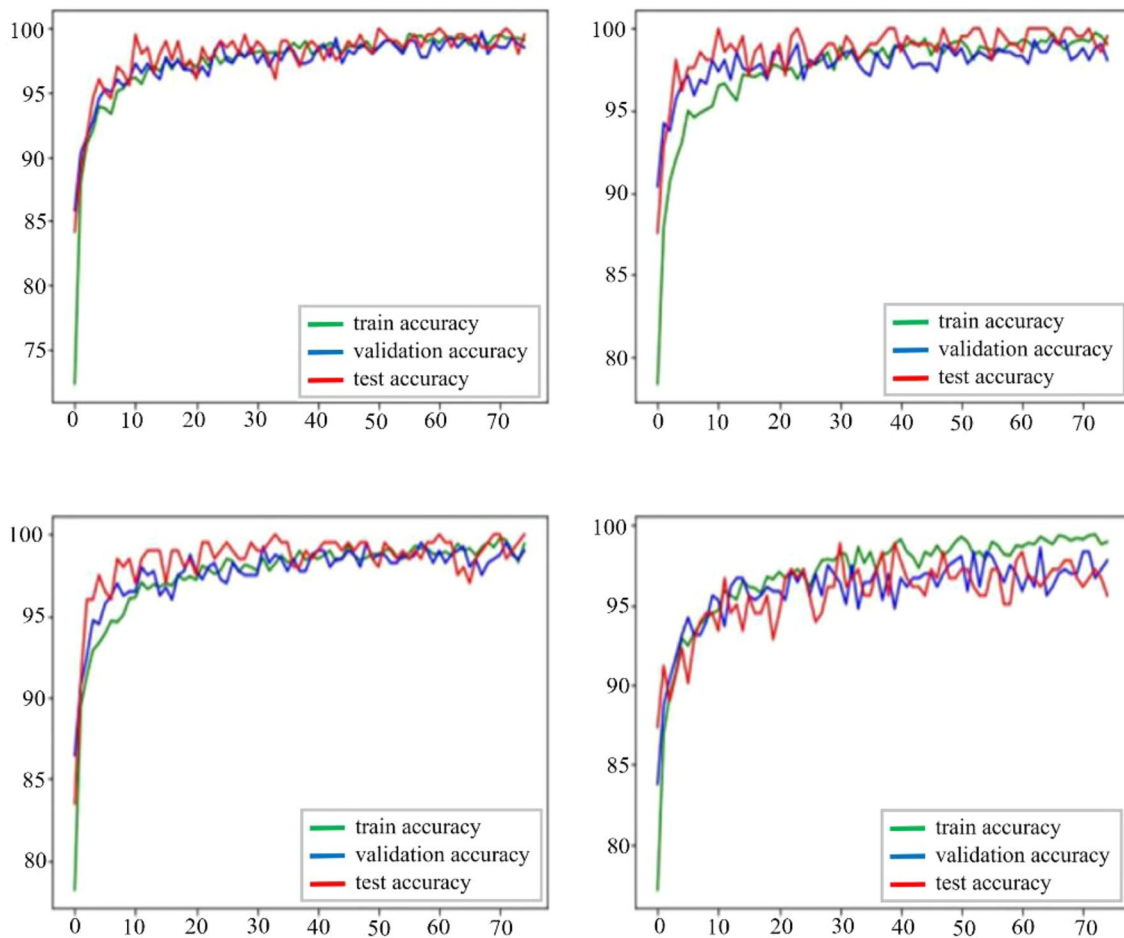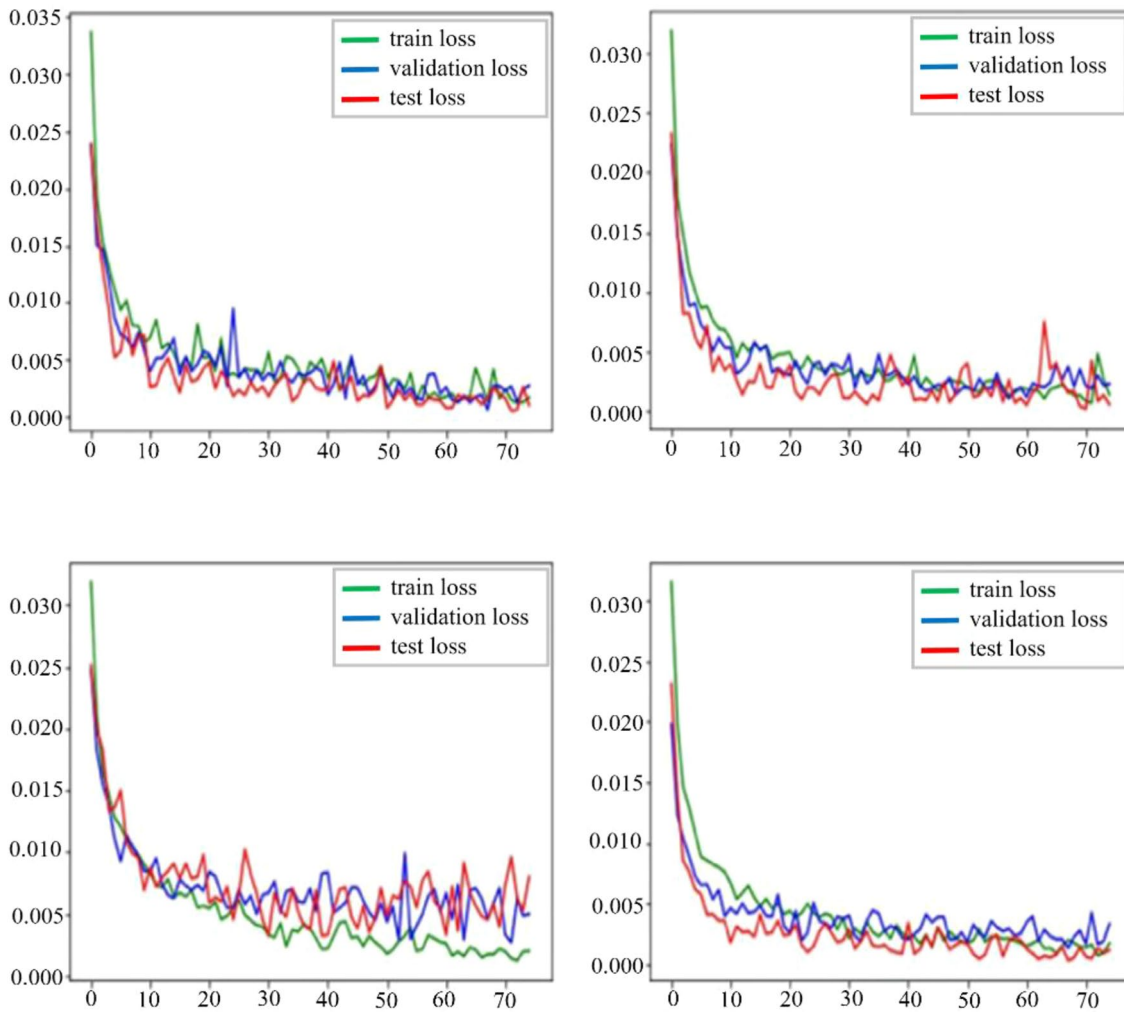
Figure 5. Accuracy curve of ResNeXt-50 (32 × 4d) model, (top left 40×, top right-100×, below left-200×, and below right-400×).

performance outcomes compared to the other models. The graphical plot of performance metrics distribution for all the proposed models is shown in Figure 9.

It can be observed in Figures 5 and 7 that our selected models performed very well on the test data set. With increasing number of epochs, the accuracy sharply increases and then saturates, which means that the model training is completed for the selected network. Moreover, from the training phase, an important conclusion is observed that the network did not have the characteristics of neither under-fitting nor over-fitting. The validation accuracy curves, and training accuracy curves are very similar.

5. Discussion

The goal of this research study was to establish the ability of DL models to extract discriminative features for breast cancer detection based on histopathological images. The study was focused on transfer learning, and the results demonstrate the effectiveness of the models examined. The accuracy scores obtained are comparable with other state-of-the-art studies of breast cancer histopathological image classification. The outcomes of the study are compared in Table 2 with previous studies reported in literature. It should be noted that even if the improvements by the proposed models may appear marginal, DL considers them significant. Because even a slight increase in accuracy has a significant impact on overall performance in the medical imaging industry. For a detailed analysis of the model performance, the other metrics such as precision, recall, and F1-score are provided in Figures 9 and 10. Among all the investigated models, ResNeXt-50 (32 × 4d) provided the highest accuracy score of 99.8%. The precision, recall and F1-score of the ResNeXt-50 (32 × 4d) model are 99.825%, 99.9%, and 99.875%, respectively. Models like ResNet-34 and DenseNet (161, 169, and 201) have provided comparable results on the test data set with accuracies of 99.725%, 99.75%, 99.725%, and 99.725%, respec-

Figure 6. Loss curve of ResNeXt-50 (32 × 4d) model, (top left-40×, top right-100×, below left-200×, and below right-400×).

tively. However, statistically there is not a significance difference with respect to other convolutional networks for detecting breast cancer. The remaining models investigated provide good results compared to the results reported in literature. Figure 11 shows the performance statistics for the various models in terms of precision, recall and F1 score.

In this work we have introduced several strategies for training the models, such as data pre-processing, transformation, parameter tuning, etc. These strategies allow the models to deal with different challenges. In summary, the advanced DL model selected here has exhibited the capability to extract many multi-level features from histopathological images which can be applied in the automatic detection of breast cancer.

To ensure that our results are not solely a result of bias on the model or data, we took the following steps to validate our approach: We randomly shuffled our data and carefully split the data set into training and testing sets. This process guarantees that the training and testing data are unique and independent of each other, reducing the chances of data leakage or overfitting. We used cross-validation to evaluate the performance of our model. This means that we trained and tested the model on different folds of the data, which helps to reduce the variance of the results. By employing these steps, we have demonstrated the generalization ability of our transfer learning model and confirm that its performance is indeed a result of the application of transfer learning to our specific task.
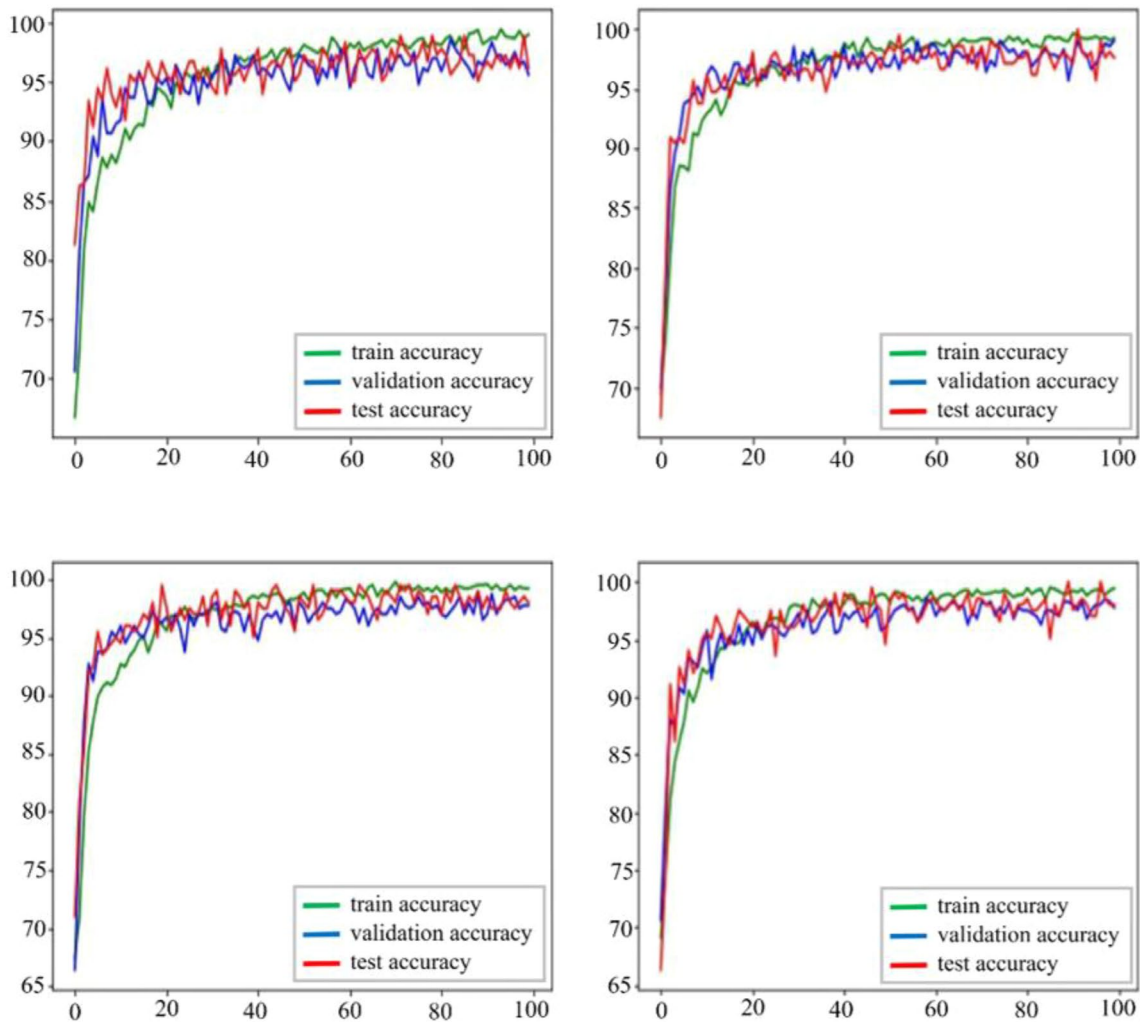
Figure 7. Accuracy curve of NASNet-N (large) model, (top left-40×, top right-100×, below left-200×, and below right-400×).

6. Conclusion and Future Work

In medical applications, the interest in DL models is growing fast as it outperforms traditional learning models in terms of speed and reliability. In our investigation, a series of experiments were conducted utilizing the most recent and precise CNN models that were trained on the ImageNet data set. We have fine-tuned these models and trained them using the BreaKHis database. Because of the lack of training images and the complexity of medical images, the transfer learning technique has been applied in this experiment. It is found that transfer learning is a very effective solution to meet these challenges. Experimental results of the models investigated indicate exceptional performance compared to other traditional machine learning models for detecting breast cancer in histopathological images. In this work, it is shown that the models provide different results for different degrees of image resolution. The results show DL models find it challenging to deal with low resolution and noisy images. Our work was conducted on 2D histopathological images, but it is known that the amount of information obtained by 2D images has their limitations. On the other hand, the 3D approach provides enriched images and enables the visualization of differences in size and volume of the mammary glands. Although it is a challenging task, efforts are currently being made to reconstruct 3D histology images by stacking 2D images (Mertzanidou et al., 2017). Future work will be focused on experimenting with various DL models applied to 3D histopathological images. Additionally, new strategies will be devised and adopted to improve the resolution of the histopathological images to improve the performance of DL models.
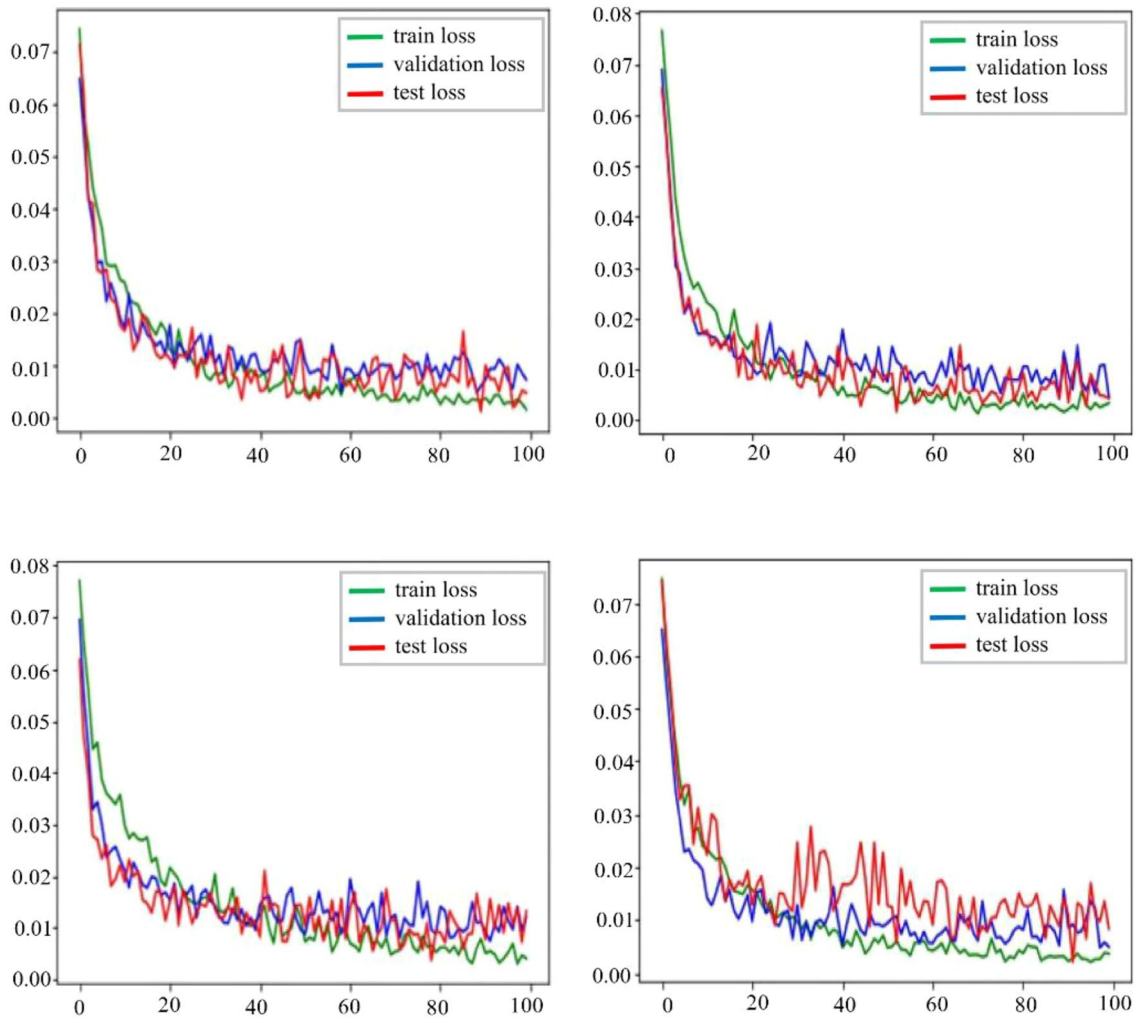
Figure 8. Loss curve of NASNet-N (large) model, (top left-40×, top right-100×, below left-200×, and below right-400×).
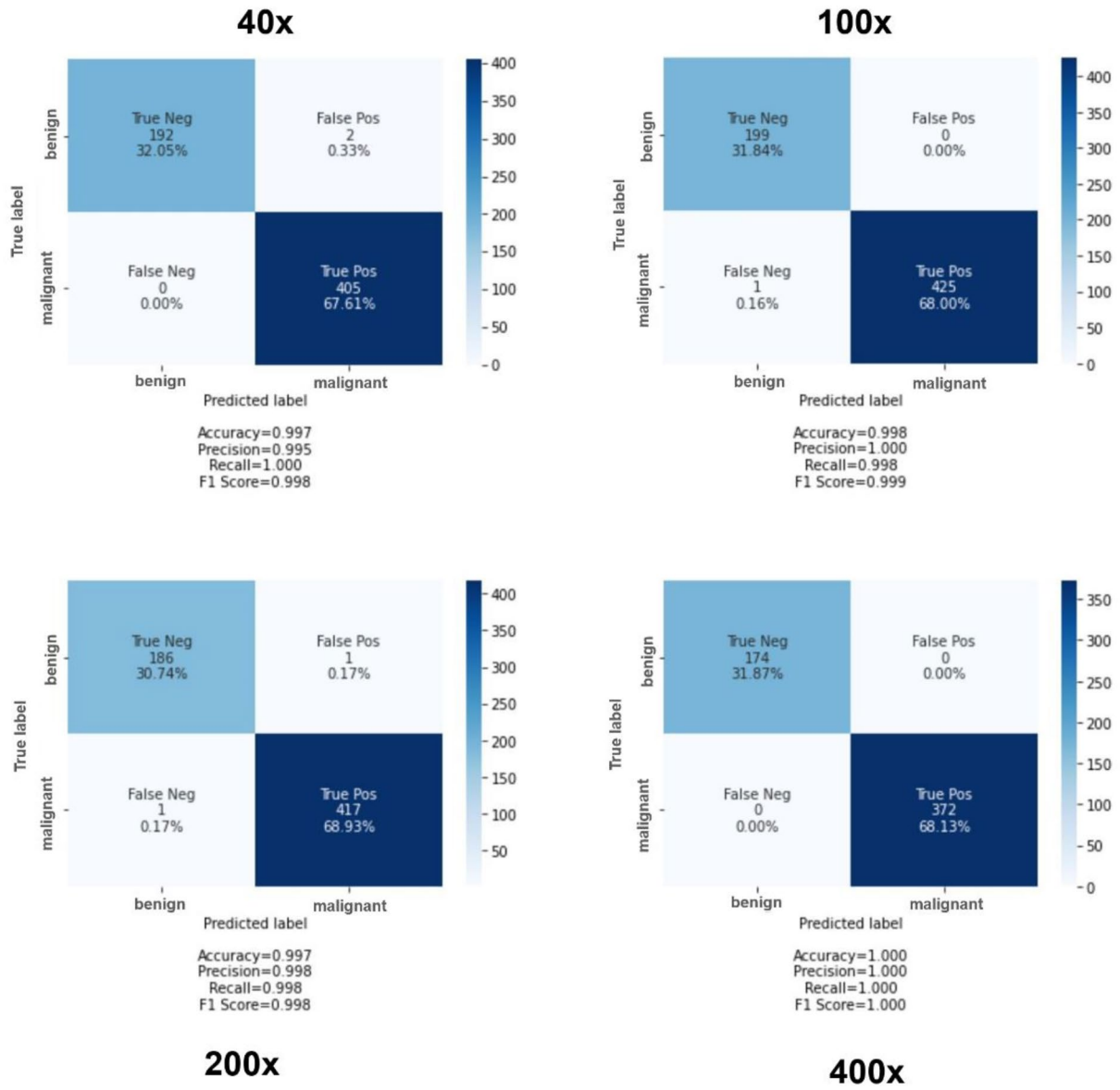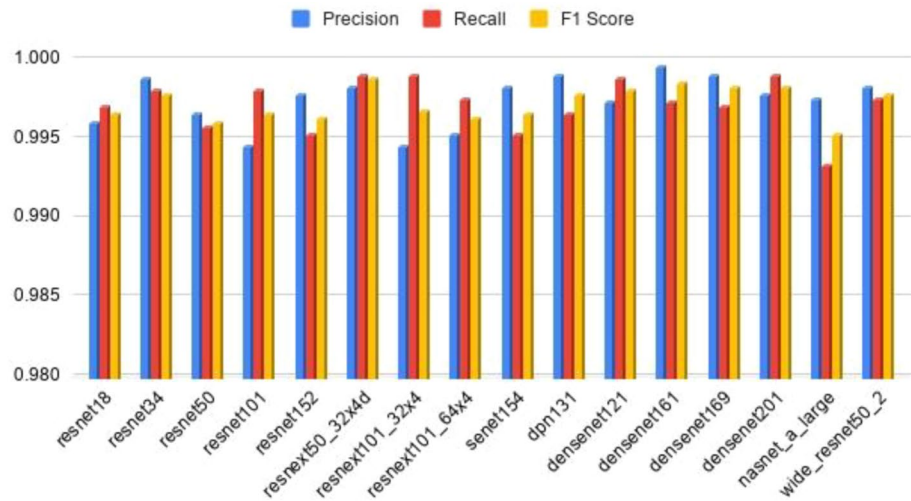
Figure 9. Confusion matrix of ResNeXt-50 (32 × 4d) model for all magnitudes.
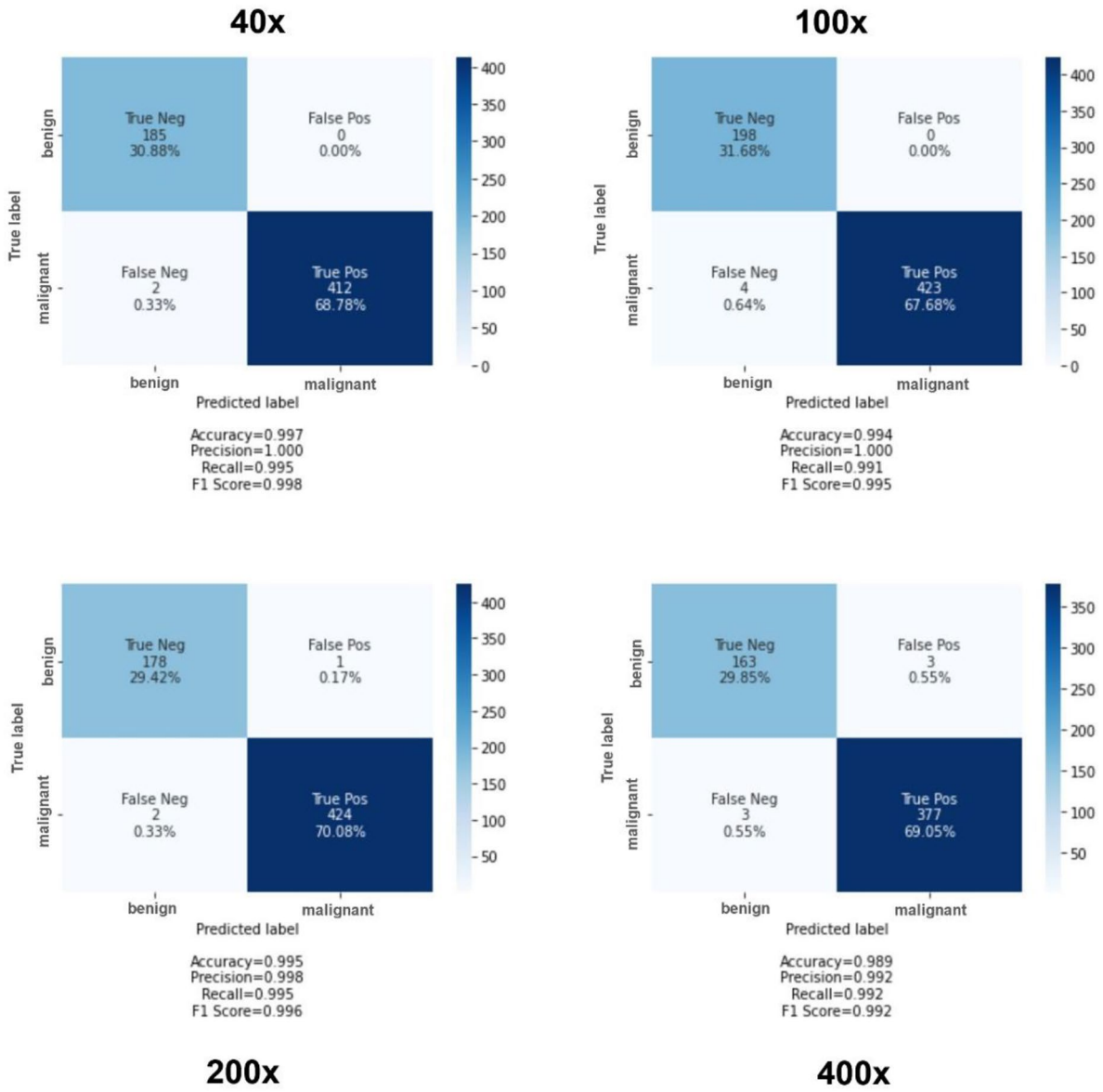
Figure 10. Confusion matrix of NASNet-N (large) model for all magnitudes.

Figure 11. Performance statistics for the various models.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

Data were not used, nor created for this research.

References

Alex, P. (2019). Cnn-finetune 0.6.0. Retrieved from https://pypi.org/project/cnn-finetune/

Al-Haija, Q. A., & Adebanjo, A. (2020). Breast cancer diagnosis in histopathological images using ResNet-50 convolutional neural network. In *IEEE international IOT, Electronics and Mechatronics Conference* (pp. 1–7). https://doi.org/10.1109/IEMTRONICS51293.2020.9216455

Alom, M. Z., Yakopcic, C., Nasrin, M. S., Taha, T. M., & Asari, V. K. (2019). Breast cancer classification from histopathological images with inception recurrent residual convolutional neural network. *Journal of Digital Imaging*, *32*(4), 605–617. https://doi.org/10.1007/s10278-019-00182-7

Aswathy, M. A., & Jagannath, M. (2017). Detection of breast cancer on digital histopathology images: Present status and future possibilities. *Informatics in Medicine Unlocked*, *8*, 74–79. https://doi.org/10.1016/j.imu.2016.11.001

Boumaraf, S., Liu, X., Zheng, Z., Ma, X., & Ferkous, C. (2021). A new transfer learning based approach to magnification dependent and independent classification of breast cancer in histopathological images. *Biomedical Signal Processing and Control*, *63*, 102192. https://doi.org/10.1016/j.bspc.2020.102192

Chen, Y., Li, J., Xiao, H., Jin, X., Yan, S., & Feng, J. (2017). Dual path networks. *Advances in Neural Information Processing Systems*, *30*.

Choudhary, T., Mishra, V., Goswami, A., & Sarangapani, J. (2021). A transfer learning with structured filter pruning approach for improved breast cancer classification on point-of-care devices. *Computers in Biology and Medicine*, *134*, 104432. https://doi.org/10.1016/j.compbiomed.2021.104432

Dimitriou, N., Arandjelović, O., & Caie, P. D. (2019). Deep learning for whole' slide image analysis: An overview. *Frontiers of Medicine*, *6*, 264. https://doi.org/10.3389/fmed.2019.00264

Dromain, C., Boyer, B., Ferré, R., Canale, S., Delaloge, S., & Balleyguier, C. (2013). Computed-aided diagnosis (CAD) in the detection of breast cancer. *European Journal of Radiology*, *82*(3), 417–423. https://doi.org/10.1016/j.ejrad.2012.03.005

Gandomkar, Z., Brennan, P. C., & Mello-Thoms, C. (2018). Mudern: Multicategory classification of breast histopathological image using deep residual networks. *Artificial Intelligence in Medicine*, *88*, 14–24. https://doi.org/10.1016/j.artmed.2018.04.005

Gupta, K., & Chawla, N. (2020). Analysis of histopathological images for prediction of breast cancer using traditional classifiers with pre-trained CNN. *Procedia Computer Science*, *167*, 878–889. https://doi.org/10.1016/j.procs.2020.03.427

Gupta, V., & Bhavsar, A. (2018). Sequential modeling of deep features for breast cancer histopathological image classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 2335–23357). https://doi.org/10.1109/CVPRW.2018.00302

Gurcan, M. N., Boucheron, L. E., Can, A., Madabhushi, A., Rajpoot, N. M., & Yener, B. (2009). Histopathological image analysis: A review. *IEEE Reviews in Biomedical Engineering*, *2*, 147–171. https://doi.org/10.1109/RBME.2009.2034865

Hameed, Z., Zahia, S., Garcia-Zapirain, B., Javier Aguirre, J., & María Vanegas, A. (2020). Breast cancer histopathology image classification using an ensemble of deep learning models. *Sensors*, *20*(16), 4373. https://doi.org/10.3390/s20164373

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770–778). https://doi.org/10.1109/CVPR.2016.90

Hosna, A., Merry, E., Gyalmo, J., Alom, Z., Aung, Z., & Azim, M. A. (2022). Transfer learning: A friendly introduction. *Journal of Big Data*, *9*(1), 102. https://doi.org/10.1186/s40537-022-00652-w

Hu, J., Shen, L., & Sun, G. (2018). In *Squeeze-and-excitation networks IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 7132–7141). https://doi.org/10.1109/CVPR.2018.00745

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). In *Densely Connected Convolutional Networks IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2261–2269). https://doi.org/10.1109/CVPR.2017.243

Jannesari, M., Habibzadeh, M., Aboulkheyr, H., Khosravi, P., Elemento, O., Totonchi, M., & Hajirasouliha, I. (2018). Breast cancer histopathological image classification: A deep learning approach. In *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 2405–2412). https://doi.org/10.1109/BIBM.2018.8621307

Jiang, Y., Chen, L., Zhang, H., & Xiao, X. (2019). Breast cancer histopathological image classification using convolutional neural networks with small SE-ResNet module. *PLoS One*, *14*(3), e0214587. https://doi.org/10.1371/journal.pone.0214587

Kassani, S. H., Kassani, P. H., Wesolowski, M. J., Schneider, K. A., & Deters, R. (2019). Breast cancer diagnosis with transfer learning and global pooling. In *International Conference on Information and Communication Technology Convergence (ICTC)* (pp. 519–524). IEEE Publications. https://doi.org/10.1109/ICTC46691.2019.8939878

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

Lakhani, S. R., Ellis, I. O., Schnitt, S., Tan, P. H., & van de Vijver, M. (2012). In *Who classification of tumours of the breast* (4th ed., p. 4). IARC Publishing.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lim, M., Kim, D., Chung, D., Lim, H., & Kwon, Y. (2018). Deep convolution neural networks for medical image analysis. *International Journal of Engineering and Technology*, *7*(3.33), 115–119. https://doi.org/10.14419/ijet.v7i3.33.18588

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., et al. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, *42*, 60–88. https://doi.org/10.1016/j.media.2017.07.005

Macenko, M., Niethammer, M., Marron, J. S., Borland, D., Woosley, J. T., Guan, X., et al. (2009). A method for normalizing histology slides for quantitative analysis. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (pp. 1107–1110). IEEE Publications. https://doi.org/10.1109/ISBI.2009.5193250

Man, R., Yang, P., & Xu, B. (2020). Classification of breast cancer histopathological images using discriminative patches screened by generative adversarial networks. *IEEE Access*, *8*, 155362–155377. https://doi.org/10.1109/ACCESS.2020.3019327

Mertzanidou, T., Hipwell, J. H., Reis, S., Hawkes, D. J., Ehteshami Bejnordi, B., Dalmis, M., et al. (2017). 3d volume reconstruction from serial breast specimen radiographs for mapping between histology and 3d whole specimen imaging. *Medical Physics*, *44*(3), 935–948. https://doi.org/10.1002/mp.12077

Motlagh, M. H., Jannesari, M., Aboulkheyr, H., Khosravi, P., Elemento, O., Totonchi, M., & Hajirasouliha, I. (2018). Breast cancer histopathological image classification: A deep learning approach. *bioRxiv*, 242818. https://doi.org/10.1101/242818

Nawaz, M., Sewissy, A. A., & Hassan, T. H. A. (2018). Multi-class breast cancer classification using deep learning convolutional neural network. *International Journal of Advanced Computer Science and Applications*, *9*(6), 316–332. https://doi.org/10.14569/IJACSA.2018.090645

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., et al. (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*, *32*.

Reinhard, E., Adhikhmin, M., Gooch, B., & Shirley, P. (2001). Color transfer between images. *IEEE Computer Graphics and Applications*, *21*(4), 34–41. https://doi.org/10.1109/38.946629

Shahidi, F., Mohd Daud, S. M., Abas, H., Ahmad, N. A., & Maarop, N. (2020). Breast cancer classification using deep learning approaches and histopathology image: A comparison study. *IEEE Access*, *8*, 187531–187552. https://doi.org/10.1109/ACCESS.2020.3029881

Spanhol, F. A., Oliveira, L. S., Petitjean, C., & Heutte, L. (2016). A data set for breast cancer histopathological image classification. *IEEE Transactions on Biomedical Engineering*, *63*(7), 1455–1462. https://doi.org/10.1109/TBME.2015.2496264

Veta, M., Pluim, J. P., van Diest, P. J., & Viergever, M. A. (2014). Breast cancer histopathology image analysis: A review. *IEEE Transactions on Biomedical Engineering*, *61*(5), 1400–1411. https://doi.org/10.1109/TBME.2014.2303852

World Health Organization (2021). Breast cancer now most common form of cancer: WHO taking action. Retrieved from https://www.who.int/news/item/03-02-2021-breast-cancer-now-most-common-form-of-cancer-who-taking-action

Xie, J., Liu, R., Luttrell, J., & Zhang, C. (2019). Deep learning based analysis of histopathological images of breast cancer. *Frontiers in Genetics*, *10*, 80. https://doi.org/10.3389/fgene.2019.00080

Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 5987–5995). https://doi.org/10.1109/CVPR.2017.634

Yuan, E., & Suh, J. (2018). Neural stain normalization and unsupervised classification of cell nuclei in histopathological breast cancer images. arXiv preprint arXiv:1811.03815.

Zagoruyko, S., & Komodakis, N. (2016). Wide residual networks. arXiv preprint arXiv:1605.07146. https://doi.org/10.5244/C.30.87

Zheng, Y., Li, C., Zhou, X., Chen, H., Xu, H., Li, Y., et al. (2023). Application of transfer learning and ensemble learning in image-level classification for breast histopathology. *Intelligent Medicine*, *3*(2), 115–128. https://doi.org/10.1016/j.imed.2022.05.004

Zoph, B., & Le, Q. V. (2016). Neural architecture search with reinforcement learning. arXiv preprint arXiv:1611.01578.