



Review

Deep transfer learning for IDC breast cancer detection using fast AI technique and Squeezenet architecture

Sushovan Chaudhury^{1,*}, Kartik Sau¹, Muhammad Attique Khan² and Mohammad Shabaz^{3,*}

¹ University of Engineering and Management, Kolkata, Department of Computer Science and Engineering, University Area, Plot No. III, B/5, New Town Rd, Action Area III, Newtown, Kolkata, West Bengal 700160, India

² Department of Computer Science, HITEC University, Taxila, Pakistan

³ Model Institute of Engineering and Technology Jammu, J&K, India

* **Correspondence:** Email: sushovan.chaudhury@gmail.com; bhatsab4@gmail.com; Tel: +917006413647.

Abstract: One of the most effective approaches for identifying breast cancer is histology, which is the meticulous inspection of tissues under a microscope. The kind of cancer cells, or whether they are cancerous (malignant) or non-cancerous, is typically determined by the type of tissue that is analyzed by the test performed by the technician (benign). The goal of this study was to automate IDC classification within breast cancer histology samples using a transfer learning technique. To improve our outcomes, we combined a Gradient Color Activation Mapping (Grad CAM) and image coloring mechanism with a discriminative fine-tuning methodology employing a one-cycle strategy using FastAI techniques. There have been lots of research studies related to deep transfer learning which use the same mechanism, but this report uses a transfer learning mechanism based on lightweight Squeeze Net architecture, a variant of CNN (Convolution neural network). This strategy demonstrates that fine-tuning on Squeeze Net makes it possible to achieve satisfactory results when transitioning generic features from natural images to medical images.

Keywords: histology; color activation mapping; FastAI; transfer learning; tumor; classification

1. Introduction

A distressing number of women diagnosed with breast cancer sadly succumb to the disease each

year. Breast cancer mortality rates can be lowered, and therapy efficacy improved with early diagnosis. The use of digital pathology as a quick and reliable method of diagnosing breast cancer is a promising area of research. Using deep learning, automated breast cancer detection has been achieved in recent years. With the help of transfer learning, deep learning can tackle novel issues with less specialized hardware and a smaller data set than usual. Due to this, deep transfer learning has received a lot of attention as a potential method for improving the accuracy and precision of breast cancer prediction models. Fast AI is a well-liked deep learning utility because of how simple it is to implement transfer learning models. Because of its small size and high effectiveness, Squeezenet's deep neural network is ideal for extracting characteristics from high-dimensional picture data. Particularly helpful for object recognition in computer vision. We suggest a deep transfer learning technique using the Fast AI methodology and the Squeezenet framework for efficient and accurate breast cancer identification in digital diagnostic pictures. They plan to use the freely accessible IDC breast cancer dataset to demonstrate the value of our approach.

The incidence of chronic, noncommunicable diseases has increased dramatically over the past few decades. In 2018, the Lancet magazine released a study named "Global Burden of Disease," which found that the majority of the rise in patients was due to nearly six different types of non-communicable diseases. In this case, cancer is a real risk. Cancer is an example of a noncommunicable illness in which aberrant cells multiply at an unchecked rate, eventually overpowering the body's healthy cells. The disease will advance from an apparently innocuous and treatable phase to a deadly phase if it is not detected and treated in a prompt way. For women, breast cancer is the most prevalent type of illness, and it is the sixteenth leading killer worldwide. Among women aged 50–59, it is the third most common, but it is the most common among women aged 15–49 [1]. As many as 23.6 million extra cases of malignancy are expected to be identified this year [2]. If a breast tumor is detected and removed before it develops into cancer, it will typically not spread beyond the breast's internal tissue. However, studies have found that women with seemingly harmless breast anomalies (such as asymmetry) are at a higher risk of developing breast cancer [3,4]. Breast cancer can develop for a variety of reasons. The breast is an organ located above the upper thoracic column and the chest muscles. Males and females alike have two breasts at birth, each with its own hormonal system that generates hormones, blood vessels, and triglycerides [5,6]. Newborns can't thrive and mature properly without breast milk. A woman's breast size primarily depends on the amount of breast tissue she already has [7,8]. By 2020 and 2022, one in four female cancer diagnoses is predicted to be breast cancer, and this proportion is only anticipated to grow [9]. Breast cancer prevalence is expected to increase globally in the coming years, according to recent cancer projections. Breast neoplasms, also known as tumors, occur when abnormal cell division results in unchecked tissue growth [10,11]. Breast cancer symptoms include the development of a malignancy in either breast or a change in the size or shape of either breast or a change in the breasts' supporting tissues. Mammograms could be useful for the early diagnosis of some diseases. Breast cancer is the second most common malignancy in women [12,13], after cutaneous cancer. Skin cancer is the most common type of malignancy in females. There is an increased risk of the disease occurring after age 50 in women. In rare cases, men can develop breast cancer as well. Less than one percent of women in the United States will develop breast cancer in their lifetime, but each year about 2,600 women will receive a diagnosis [14,15].

Furthermore, global trends indicate that we have not yet achieved in our effort to completely eliminate cancer. Breast cancer caused the deaths of 6,27,000 people in 2015, according to the World

Health Organization (WHO) (BC). The rise to 30 million is anticipated to occur within the next decade, nearly tripling the current number.

The development of breast cancer is a result of unchecked cell proliferation in the breast. Tumors develop when infectious cells proliferate and disseminate at rates far exceeding those of healthy cells. Lymph nodes are a sorting facility for these cells before they are dispersed throughout the body [16]. Multidetector CT's soft-tissue capabilities are adequate for daily precise identification [17,18], despite cardiac CT's tendency to miss breast lesions. Figure 1 shows some representative cell forms from the Kaggle invasive ductal carcinoma challenge (IDC).

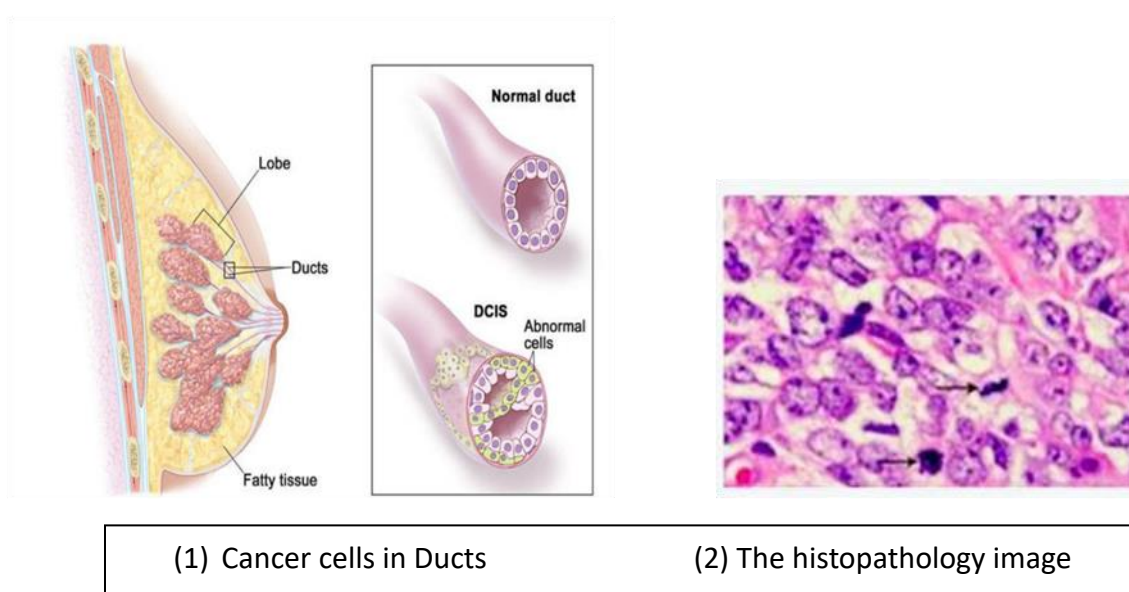


Figure 1. Appearances of Invasive Ductal Carcinoma (IDC) cells (From Kaggle datasets).

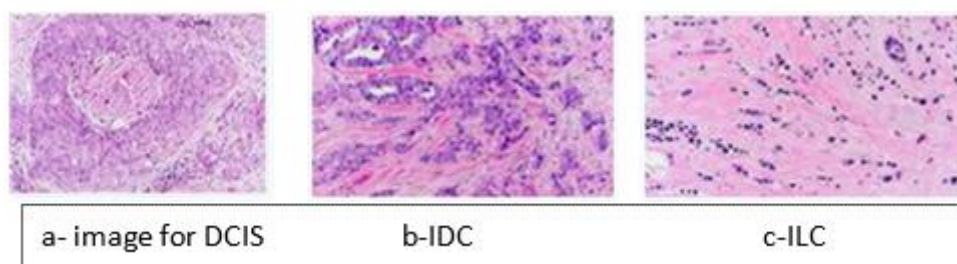


Figure 2. Three different types of BC.

There are three distinct varieties of BC, as depicted in Figure 2. As stated in the aforementioned reference [19], invasive ductal carcinoma and invasive lobular carcinoma are the two most prevalent subtypes of invasive breast cancer. There are two main types of breast cancer, and the vast majority of patients fall into one of these. Cancer Treatment Centers of America found that among women aged 45–60, invasive ductal carcinoma (IDC) was the most prevalent variety of breast cancer, while invasive lobular carcinoma (ILC) was the rarest. DCIS (Ductal Invasive Carcinoma in Situ) is a subtype of breast cancer that typically manifests at an early stage. Ductal invasive adenocarcinoma in situ (abbreviated as DCIS) is a type of invasive ductal although further cancerous growth is possible,

no cancer cells have been detected at this time. Some breast cancers have been linked to DNA harm. Of the two types of breast cancer, IDC is presently getting the most focus from experts. The many varieties of IDC are summarized in Table 1.

Table 1. Subtypes of IDC [19].

Type of Breast Cancer	Subtypes of its types
IDC (Benign)	Papillary, Tubular, Medullary, Cribriform, and Mucinous carcinomas
IDC (Malignant)	Fibroadenoma, Adenosis, Tubular Adenoma, and Phyllodes tumor

The invasiveness and frequency of tumor cells turn them into one of the most challenging malignancies to identify. As per today's scenario, there is a great demand for automation of the system which helps in classifying breast cancer histopathology images reliably. The demand arose due to the shortage of available medical professionals and a severe problem caused by inaccuracies in the result.

When IDC is found in breast cancer, it necessitates aggressive treatment, including operations and radiation treatments. To categorize slides as either positive or negative for malignancy, IDC pathology uses a microscope and manual evaluation of multiple slides. Due to the limitations of human cognition, this process takes a long time and is prone to making mistakes. These errors, however, can be easily corrected by computer vision techniques coupled with analysis of histopathological images.

Further, a few components like the misidentification of samples can lead to cancer development and the survival rate in such conditions gets too low. As a result, numerous Computer-Aided Diagnosis (CAD) systems for accurate and automated breast cancer diagnosis have been developed. In recent years, machine intelligence (MI) has revolutionized oncological research. MI has been shown in numerous investigations to correctly categorize tissue samples as benign or cancerous. Deep learning models, particularly those embedded with image interpretation tools like Convolutional neural networks (CNN), have been proven to accurately distinguish between cancerous and benign prostate biopsies from images.

Because of their built-in automatic feature extraction methodologies, CNNs, for example, have lately acquired prominence in detection and classification applications. Deep learning algorithms, on the other hand, have made significant advances in the picture categorization issue.

In the job of image analysis, deep learning-based systems beat classical machine learning approaches. However, the categorization of cancer histology images using machine learning algorithms required additional knowledge and effort. However, despite the attractiveness of machine learning (ML) jobs due to efficient accuracy, there is currently no mechanism for decoding a DL classification model.

2. Literature review

2.1. Related work

Few limitations are observed in the above studies in IDC detection using the DL mechanism.

Given the flaws, it is not guaranteed that the tedious process of detection will be carried out by

an expert with extremely accurate dimensions. Also, because larger and labeled datasets are not publicly available, the majority of the investigations were conducted with smaller datasets. These challenges need to be addressed. Hence, this paper is aimed toward the use of a combination of transfer learning-based SqueezeNet model to perform the classification of IDC with various hyper tuning parameters. For the related literature review we have devised Table 2 for easy reference.

This study also aims to provide a full one-stop solution for early and automatic tumor diagnosis utilizing entire slide images; early detection of cancer has always been helpful. We can now avoid cancer's worsening malignant condition with fewer efforts; credit goes to modern machine learning libraries. For this purpose, the 1-cycle policy and FastAI are being used. FastAI is a PyTorch-based open-source deep-learning package that offers high-level techniques for easy DL model training [30–32]. The feature extraction mechanism helps to retrieve the necessary attributes, from each image, by splitting the image into smaller tiles of equivalent size. Machine learning algorithms get these features as inputs and it accelerates the development of mathematical models that can identify tumorous areas in images. For the processing task of the images, there are several inbuilt libraries in Python programming language to ease the task. Authors of articles have done numerous works on reviews and research works on breast cancer image segmentation and classification as well as other medical image classification tasks [42–49]. By comparing them, we have identified that very less amount of work has been done on the breast histology dataset and whatever is done did plain classification. As such we identified the gap that not the whole part of the histology image bears the same importance while the classification of the lesions, as such we decided to use GRAD-CAM based Attention mechanism to give more importance to certain areas of the images during the model building.

Metaheuristic optimization techniques learn from the world around them. This section discusses some recently proposed algorithms and provides resources for further reading.

Calculated strategies: To arrive at the optimal solution, an arithmetic optimization algorithm (AOA) uses only addition, subtraction, multiplication, and division. Until the target population is achieved, new findings are generated by repeatedly applying statistical processes to different subsets of the group. Engineers, bankers, and energy sector professionals have all reported increased productivity thanks to AOA.

The Prairie Dog Optimization (PDO) software is an optimization tool that uses metaheuristic techniques inspired by the behavior of prairie dogs. It forms tight-knit groups, much like prairie dog colonies, in order to hunt for food and avoid being harmed. Many industries, including engineering, public transportation, and finance, have found success using PDO to address optimization problems.

The lightning-fast effectiveness with which untamed gazelles travel served as inspiration for the novel metaheuristic program known as the Gazelle Optimization program (GOA). It forages for food like an impala, accidentally finding the sweet spot between exploration and exploitation. Successful applications of GOA to optimization problems have been found in many different fields, such as engineering, economics, and biology.

The new Dwarf Mongoose Optimization technique was created using knowledge gained from studying wild dwarf mongooses (DMOA). When it comes to hunting, hiding from predators, and cooperating, it behaves like a colony of tiny mongooses. Many fields, including engineering, medicine, and finance, have found success in applying DMOA to their expansion problems.

The Aquila Optimizer is a state-of-the-art meta-heuristic optimization method motivated by the ways in which wild eagles solve problems. It attempts to be like an eagle in that it seeks out

sustenance but also keeps some of it for itself. There are many fields that have adopted Aquila to improve productivity, from engineering to business to finance.

Table 2. Related literature review.

Title of Paper /Reference	Author(s)	Year of publication	Algorithm used	Dataset Used	Task performed
“Machine learning techniques to diagnose breast cancer.; [20]”	“Alireza Osareh et al.”	2010	K-nearest Neighbours, SVM Classifier	“Breast Cancer Histology images”	Differentiate between Benign and Malignant Cancer.
“Deep Learning for Magnification Independent Breast Cancer Histopathology Image Classification; [21]”	“Neslihan Bayramoglu & Juho Kannala & Janne Heikkila”	2016	CNN performing single-task and multi-task	“BreakHis Dataset”	Prediction of malignancy
“Convolutional Neural Networks and the Analysis of Cancer Imagery; [22]”	“Chris Pearce”	2017	Basic CNN	“Tumour Proliferation Assessment Challenge 2016 datasets”	Classification
“Breast Cancer Detection In Mammogram Images Using Deep Learning Technique;[23]”	“D. Selvathi and A. Aarthy Poornila”	2017	Unsupervised deep learning technique	“Mammographic images from MIAS Mini Mammographic Database”	Detection
“Deep Learning Framework for Multi-class Breast Cancer Histology Image Classification; [24]”	“Yeeleng S. Vang(B), Zhen Chen, and Xiaohui Xie”	2018	GoogLeNet Inception V3	“Breast Cancer Histology images (BACH)”	Classification
“Transfer learning based histopathologic image classification for breast cancer detection; [25]”	“Erkan Deniz1, Abdulkadir Şefngür”	2018	AlexNet and VGG16 models	“BreakHis Dataset”	Classification
“Deep Learning Applied for Histological Diagnosis of Breast Cancer; [16]”	“Yasin Yari, Thuy V. Nguyen, and Hieu T. Nguyen”	2020	ResNet and DenseNet from DCNN combined with transfer learning	“BreakHis Dataset”	Classification

Continued on next page

Title of Paper /Reference	Author(s)	Year of publication	Algorithm used	Dataset Used	Task performed
“A Novel Approach to Classifying Breast Cancer Histopathology Biopsy Images Using Bilateral Knowledge Distillation and Label Smoothing Regularization;[26]”	S.Chaudhury etal.	2021	Transfer Learning using Bilateral Knowledge Distillation	“Breast Cancer Histology images (BACH)”	Classification
“Breast Cancer Calcifications: Identification Using a Novel Segmentation Approach; [27]”	S.Chaudhury etal.	2021	Breast Cancer detection using novel segmentation technique in mammographic images.	“Mammographic images from MIAS Mini Mammographic Database”	Detection
“Effective Image Processing and Segmentation-Based Machine Learning Techniques for Diagnosis of Breast Cancer; [28]”	S.Chaudhury etal.	2022	Breast Cancer detection using CLAHE, K-Means and classification using SVM, Random Forest method.	“Mammographic images from MIAS Mini Mammographic Database”	Detection and Classification
“Breast Cancer Histopathology Image Classification and Localization using Multiple Instance Learning; [29]”	Abhijit Patil etal.	2020	Attention based Multiple instance learning and localization.	BreakHIS and BACH dataset	Detection and Classification
“ A BERT encoding with Recurrent Neural Network and Long-Short Term Memory for breast cancer image classification,[50]”	Sushovan Chaudhury etal.	2023	Vision Transformer and BEiT based LSTM approach for classification	Breast Ultrasound Images	Detection and classification

An innovative form of meta-heuristic software, called RSA, was inspired by the behaviors of snakes in the environment. Similarly, to snakes, it seeks food, stays out of harm's way, and finds a happy medium between research and practical application. Manufacturing, healthcare, and finance are just a few of the sectors that have found RSA to be useful.

The Ebola Optimization Search Algorithm is a novel metaheuristic algorithm that takes its inspiration from the ecosystem of the Ebola virus (EOSA). This virus, like the Ebola virus, spreads from cell to cell by penetrating their walls. Efficiency problems in the mechanical, biochemical, and medical sectors are just a few examples of where EOSA has been put to use.

The aforementioned metaheuristic approaches have all demonstrated promising results in

addressing various planning challenges. Improving the characteristics of machine learning models, feature selection, and related tasks can be useful in conducting further research into and fine-tuning these strategies for dealing with breast cancer classification problems. Many of these strategies have been implemented by the authors of [51].

2.2. Primary Contributions based on Research Gaps

This research study's primary contributions can be described as follows:

- Introduction of the latest Deep Learning algorithms such as the FastAI platform for evaluating difficult breast histology biopsy images.
- Histopathology, or the microscopic examination of tissues, is one of the most important diagnostic procedures for the detection of breast cancer.
- The pathologist, who observes the cells usually, investigates the analysis of texture. This analysis determines the type of tissue, with a particular focus on the tumor-stroma ratio. The main contribution of the pathologist is the automation of the tissue classification task of histology samples of BC through the use of deep transfer learning mechanisms.
- For better and improved results, the planning continues, and a suggestion for the employment of discriminative fine-tuning methods integrated with a one-cycle policy emerges, with the last stage concluding with a recommendation for the use of color normalizing approaches. These methods were developed to maintain the notion useful.
- Squeeze Net, which highly supports memory-limited hardware, helps to determine the tissue type of a cell for the concerned pathologists for the classification results.

3. Materials and methods

Deep learning algorithms show high robustness when used on image datasets. The dataset for the classification of IDC consists of histopathological images categorized into two types: the first category is BC-affected cells, and the second category includes normal or unaffected cells. Since Image data is not simple hence, some preprocessing steps are carried out on the images before conducting channel extraction in this model. Particularly, this section discusses the following: The gradient-weighted class activation mapping (GradCAM), the SqueezeNet model's architecture, FastAI's deep learning framework, and the 1Cycle Policy.

3.1. Dataset description

They use the openly available dataset hosted on the Kaggle computing platform (<https://kaggle.com>) to complete the IDC categorization task. Information from 162 whole mount slide photographs of BC Specimens is the subject of the study reported in [31]. The 78,786 IDC-positive samples and the 198,738 IDC-negative samples were used to create these 277,524 patches. A total of 277,524 RGB 50x50 pixel digital image fragments were created from 162 H&E-stained breast histology samples.

These tiny spots, created from digital images of breast tissue, are helpful for distinguishing between different types of cancerous cells. The areas labeled with a 1 contain cells with the features typical of IDCs. [33–35] Eighty percent of the data is used in classroom teaching, while only twenty percent is put to use in evaluating student progress. The training set contains 7042 photos out of a

total of 8801. The suggested deep learning approach is implemented using Python, a popular computer language. For categorization and identifying jobs, this language employs a wide range of libraries and frameworks, including Tensorflow and Keras.

3.2. *FastAI*

A professional and powerful library with high-level source codes to work on images and other datasets are called FastAI [36]. It also provides academics with access to low-level data that may be used to construct innovative models and algorithms. FastAI contributed to the development of a collaborative interface for dealing with the most often used DL applications, such as computer vision, collaborative filtering, time-series, tabular data, and text analysis. An incredibly productive, configurable, and easily adaptable framework is the key design goal while working with FastAI. It is comprised of layered architecture in which the high-level APIs do not need to know the utilization of the lower-level API. In terms of unconnected abstractions, FastAI demonstrates common patterns of numerous deep learning and data processing algorithms. Python language along with the PyTorch library, are collectively used to express the abstractions legibly and precisely. FastAI is a popular open-source library built on top of PyTorch that provides high-level APIs for training state-of-the-art deep learning models with minimal code. It was developed by Jeremy Howard and Sylvain Gugger, who are both well-known figures in the deep learning community. The library offers a range of functionality, including pre-processing data, visualizing data, creating and training models, and deploying models. FastAI also includes implementations of state-of-the-art models such as ResNet, VGG, and Transformer, which can be easily fine-tuned on custom datasets.

One of the key features of FastAI is its emphasis on "best practices" for deep learning. The library includes a comprehensive set of tools and techniques for training models effectively, such as learning rate schedules, weight decay, and data augmentation. These practices are based on the latest research findings and have been shown to significantly improve model performance. FastAI also includes a number of built-in visualization tools to help users understand and debug their models. For example, the library provides easy-to-use methods for visualizing activations, gradients, and learning rates, which can be invaluable in identifying problems with model performance. Overall, FastAI is a powerful and user-friendly library that has been widely adopted in the deep learning community. Its focus on best practices and ease of use has made it a popular choice for both researchers and practitioners.

Overall, FastAI includes:

- Both a logical type hierarchy for tensors and a new kind of delivery mechanism for Python are presented.
- A computer vision library designed for GPUs that may be extended in pure Python.
- A new data block API

4. Proposed method

4.1. *Pre-processing and normalization*

Color variations in images are caused by a mismatch in the color responses of slide scanners, raw materials, and the varied production techniques employed by stain vendors and their procedures. This color variation further results in a problem that is very common while conducting image

analysis in histology. Stain normalization provides a solution to this issue. As a result, stain normalizing is essential for a pre-processing step before doing any histological image studies. Figure 3 represents the histological images of the breast as loaded from the dataset. 0 and 1 represent the benign and malignant cells respectively.

The normalization step takes an image as an input, divides it by 255, and hence gives the result value in the range from 0 to 1. This expedites the model's training process. Several problems related to image datasets like vanishing and exploding gradients are resolved with the help of normalization. The dataset of 90,000 images consists of 64583 images from class 0 and the rest of 25417 images from class 1. This is depicted in Figure 3.

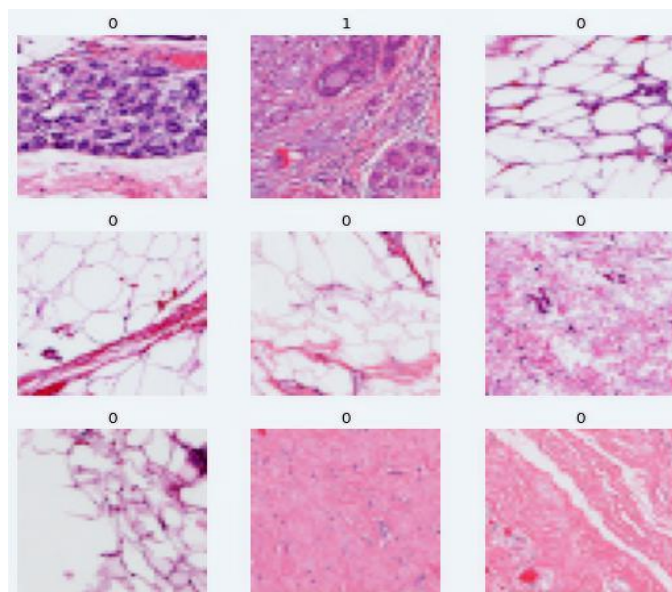


Figure 3. Loaded images from Dataset.

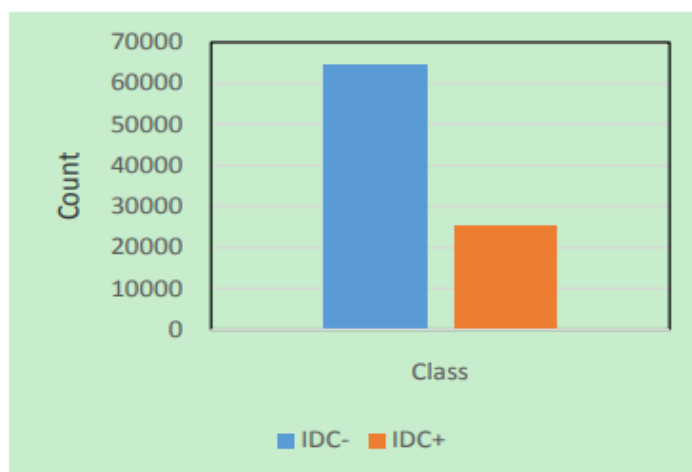


Figure 4. Data imbalance before random under-sampling.

This issue, known as data imbalance, leads to model biasing towards one particular class. A random under-sampling technique can be employed to solve this data imbalance problem. Samples

from the majority class are deleted to provide an equal number of samples for both the majority and minority classes. Figure 4 depicts the equal number of samples for the two classes after using the under-sampling strategy. In Figure 5, the blue bar specifies Invasive Ductal Carcinoma Negative (-) and the orange bar specifies Invasive Ductal Carcinoma Positive (+) after random under-sampling. The bar graph clearly shows samples in both classes are equal in size. The sample specifications for the random under-sampling technique have been summarized in Table 3.

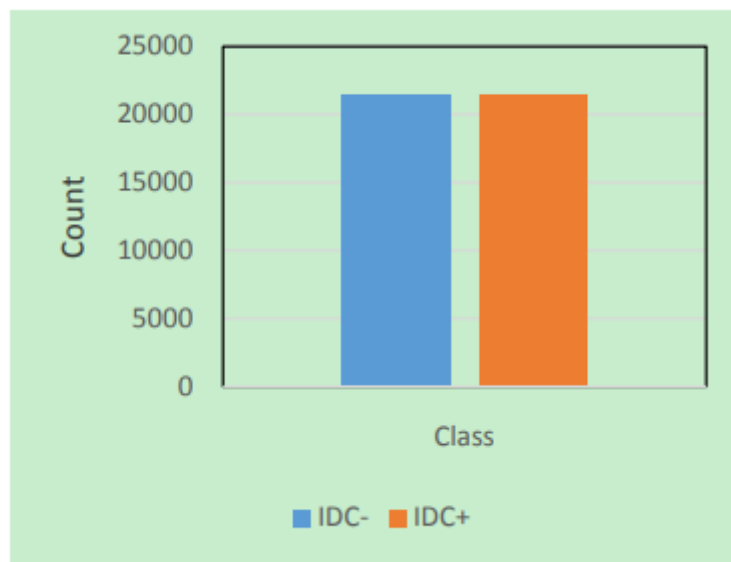


Figure 5. IDC- and IDC+ count after random under sampling.

Table 3. Sample specifications following random under-sampling technique.

Class	Labels	Training data	Testing data	Size	File extension
IDC positive	(+)	20336	5084	50x50	PNG
IDC negative	(-)	20336	5084	50x50	PNG

4.2. Data augmentation

When we have large datasets, deep learning models perform much better. Data augmentation or jittering is a popular approach to make our datasets bigger. Data augmentation is used to increase the amount of a dataset to alleviate the problem of limiting data size. Data augmentation raises the dataset's size to ten times its original one when the training dataset is very little. This helps minimize overfitting. Adding noise to incoming images or applying geometric modifications to them are two examples of data augmentation procedures. The following methods were applied step by step to the images to be augmented:

- (1) Rotate the images by 30 degrees in the clockwise direction.
- (2) Scaling of an image by 15%,
- (3) Horizontal and vertical flipping and
- (4) Gaussian noises are included with a mean (0) and variance (0.25).

Tensor image data is produced using a real-time data augmentation method and the

KerasImageDataGenerator package. One (1) input image's outcome from various data augmentation approaches are displayed in Figure 6. Figure 6 is a representation of the results of various augmentation techniques like horizontal, vertical flip, injection noise, etc on the dataset to increase the variance within the dataset and increase the number of images. This is done to ensure that there is enough variance, and the data is not overfitted during training.

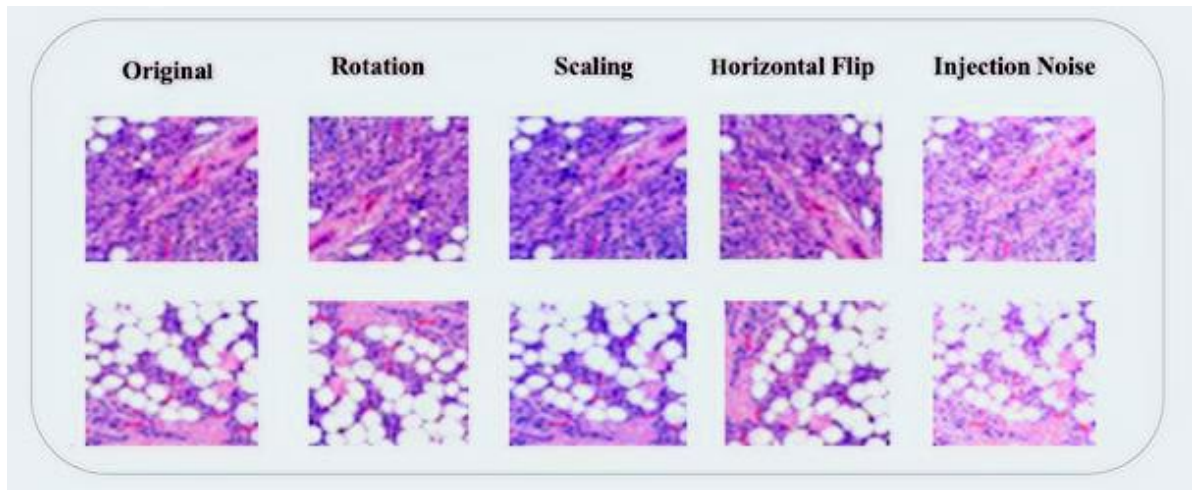


Figure 6. Transformation of Images after Augmentation [32].

4.3. Transfer learning using Squeeze Net

Transfer learning can involve natural learning. Transfer learning with fine-tuning is the process of modifying, fitting, and re-training a previously trained network based on new input; in other words, it transfers weights from the trained network that have a comparable design to the new network we wish to train with the new data. It's a powerful tool for dealing with a variety of deep learning issues. The concept of transfer learning is based on retraining a certain number of layers on the target dataset using the general characteristics obtained in earlier levels of the source dataset. The primary benefits of TL are decreased training time, increased neural network performance, and the elimination of data constraints [37–39].

Several issues arise when using transfer learning to medical image classification. One issue is that the amount of annotated data required to train CNNs is insufficient for medical image classification applications[40]. Due to the scarcity of data, large CNNs like ImageNet would struggle to prevent overfitting on these datasets. As a result, a great deal of regularization in various forms is required. Another of these issues is over parameterization, which refers to a network's large number of parameters. It will take longer to train a network with more trainable parameters, the more epochs it will require, and the more computation it will necessitate.

One answer to these difficulties may be to use lightweight architectures, which are less in size and have fewer parameters. The recently proposed lightweight architectures include “EfficientNet”, “SqueezeNet” (Figure 7 and Figure 8), and “MobileNet-v2”. There are also different variants of transfer learning. Feature extraction, weight initialization, and fine-tuning are all included as a part of this process. To properly utilize the generic information acquired in the first layers, fine-tuning is the process of freezing a specific number of layers in a model.

The latest and updated DNN (Deep Neural Network) architecture, which provides more accurate decisions for image processing applications, named Squeeze Net is used in this study [40]. Squeeze Net is a variant of CNN which is providing almost similar results and accuracy as ImageNet but uses lesser parameters than it. Replacing 3x3 filters with 1x1 filters and taking the down-sampling steps late in the network helped to maintain activation maps for the convolution layers, and reduction of input channels to 3x3 filters are important steps followed in this variant. A module named Fire module is bundled with all the above-listed strategies. On top of it, there are two layers named squeeze Convolutional layer and an expanded layer inbuilt in the Squeeze Net. Squeeze Net is a rather small model, with just 1,267,400 parameters and a 4.85 MB model size. The rectified linear unit (ReLU) was used as the activation function. Squeeze Net is a lightweight convolutional neural network architecture designed to reduce the number of parameters while maintaining high accuracy. It was introduced in 2016 and achieved Alex Net-level accuracy on ImageNet with only 50x fewer parameters. The small size of Squeeze Net makes it an attractive choice for applications where computational resources are limited, such as on mobile devices or embedded systems. Our experiments show that Squeeze Net achieved competitive performance compared to larger models while being easier to deploy and requiring less computational resources. Overall, the use of Squeeze Net demonstrates the importance of designing lightweight models that can be easily deployed in real-world scenarios, especially in medical imaging applications where resources may be limited.

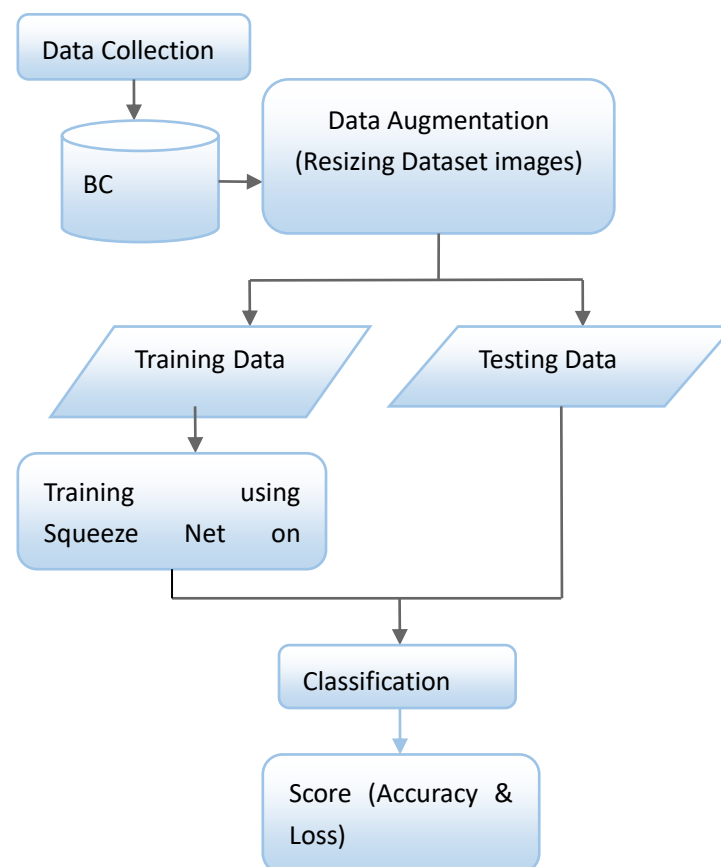


Figure 7. Transfer learning workflow Architecture.

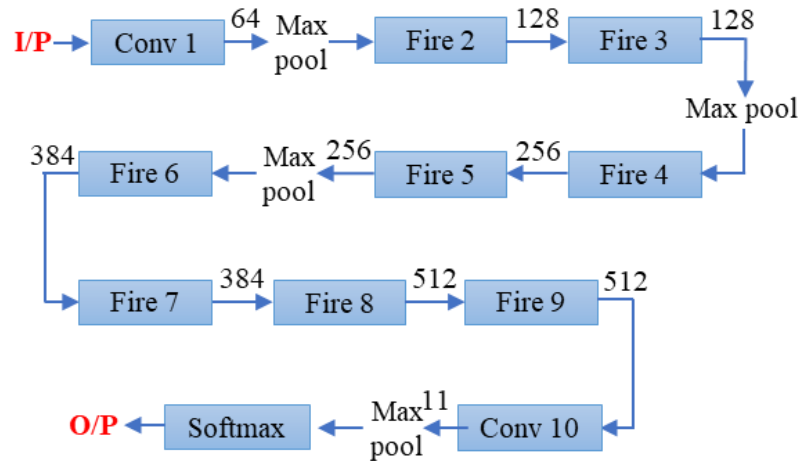


Figure 8. Transfer learning workflow Architecture Squeeze Net.

4.4. Determining the super convergence-optimal learning rate

Architecture's capacity to converge to global minima when applied to the loss function is known as the learning rate. A few tuning parameters which help to manage topology search are weight decay, momentum, learning rate, and batch size. These kinds of algorithms fear selecting high learning rates; hence, they use the optimizers like Adam, AdaGrad, AdaDelta, etc. These optimizers are distinct in that they start with a high global learning rate and progressively lower it on test sets until they reach a plateau. This entire method aids the network in achieving faster convergence. Layer group 4's associated weights are unfrozen and made to learn, while other layers are fine-tuned for two epochs. This stage is followed by the phase of super convergence. As part of a mock test [41], a network is trained on a wide variety of learning rates for 100 batches, and the learning versus loss curve is given (Figure 9).

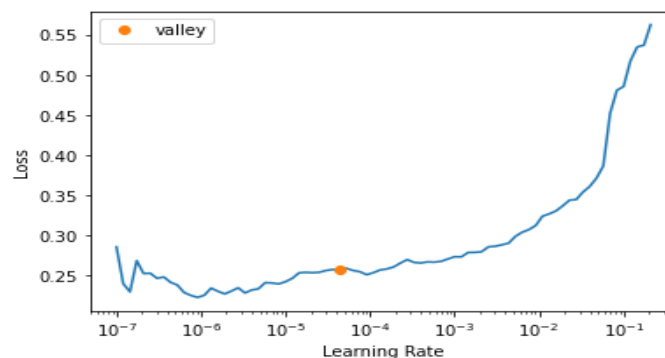


Figure 9. Training of the complete network over a variety of learning rates - Learning rate range test.

This provides us with a brief summary of the model's maximal learning rate (L_{max}); once it hits that threshold, the test or validation loss begins to climb, resulting in low accuracy. Learning rates between 0.0001 and 0.01 suffer a loss but learning rates greater than 0.01 begin to rise. In addition to the tuning parameters mentioned, the learning rate is a crucial factor in the performance of deep learning models. Choosing the right learning rate can be a challenge and often requires experimentation. A high learning rate can lead to unstable convergence and overshooting, while a

low learning rate can result in slow convergence and getting stuck in local minima. The use of optimizers such as Adam, AdaGrad, and AdaDelta has made it easier to find the optimal learning rate by dynamically adjusting the learning rate during training. The one-cycle policy with super convergence further enhances the optimization process by allowing the learning rate to vary cyclically, resulting in faster and more precise convergence to the global minima.

4.5. 1-Cycle Policy within Discriminative fine-tuning

With the one-cycle approach, you set a modest initial learning rate and gradually raise it until you hit the desired maximum. This technique is distinguished, among other things, by the high degree of agreement it accomplishes. After each iteration of data collection, the training rate is reset to reflect the new optimal value determined by the 1-cycle learning rate method. Therefore, rather than using a global learning rate, a cyclical learning rate with a nonuniform decline in weight is employed. During those two phases of the cycle, the learning rate may fluctuate. This improvement will allow the network to achieve coordination more quickly and precisely. The so-called "one-cycle policy" alternates between a moderate and high maximum learning rate (LR) during training. This rule is put into effect for just one workout at a time. This strategy entails setting the learning rate low to begin with, raising it to its maximum value, and then setting it back down again. Super convergence is a training strategy in machine learning that, when compared to traditional approaches, yields higher accuracy and a quicker rate of convergence to a decreased loss value. Combining the one-cycle approach with super convergence improves optimization effectiveness, speeds up convergence, and reduces overfitting. Changing the learning rate at regular periods allows the network to explore more potential weight combos and find the optimal answer more quickly. Training at a high learning rate can regularize the network and reduce overfitting while training at a low learning rate can maximize the weights and improve performance. One of the keys to the success of the single-cycle approach is cutting out unnecessary fat. To achieve this, the loss function can incorporate a regularization component that provides positive reinforcement for small weights and negative reinforcement for big weights. The goal is to improve generalization performance while minimizing the risk of overfitting to the data. When it comes to developing deep neural networks, the one-cycle approach and the super convergence method are both useful instruments. A quicker closure and better generalization performance may be possible with methods that employ a cyclic learning rate and weight decrease. These procedures can be used in many deep learning applications, including those dealing with image identification, NLP, and others.

4.6. Gradient-Weighted Class Activation Mapping (GRAD-CAM)

Recognition of failure modes and winning customers' loyalty and confidence is made possible only through the deep learning model's explanation. However, deep neural networks are very difficult to break down into simple, understandable parts. Class Activation Map (CAM) is an approved strategy for explaining deep learning model decisions, yet it is confined to only those architectures, in which feature maps directly precede the softmax layers. Grad CAM, a localization technique that uses class discrimination, produces visual explanations for any deep learning model. Before Grad CAM, simple CAM was employed, where weighted feature maps were produced by the Convolutional layer, then ReLU processes. It further uses this information to highlight the region's most responsible for the prediction. Through Grad-CAM, we can verify the working of our model

whose main finding is to visualize the appropriate patterns from the images and activation around those patterns. The output of the Grad-CAM mechanism is shown in Figure 10. As we can see, from the labels in the figure 10 A, B, C, D, E, F, G and H. The left images of each of the labels are normal patches of histology cells of the breast whereas the right images are the output of Grad-CAM. We know that histology cells have a critical limitation as the color combinations in the patches are hard to distinguish. The Grad-CAM approach clearly distinguishes the patches and more attentions are given to those areas which are more responsible for giving the accurate prediction that whether the cell is non-invasive, benign or malignant. Thus, this approach is a more efficient way for feature selection and attention-based classification is possible.

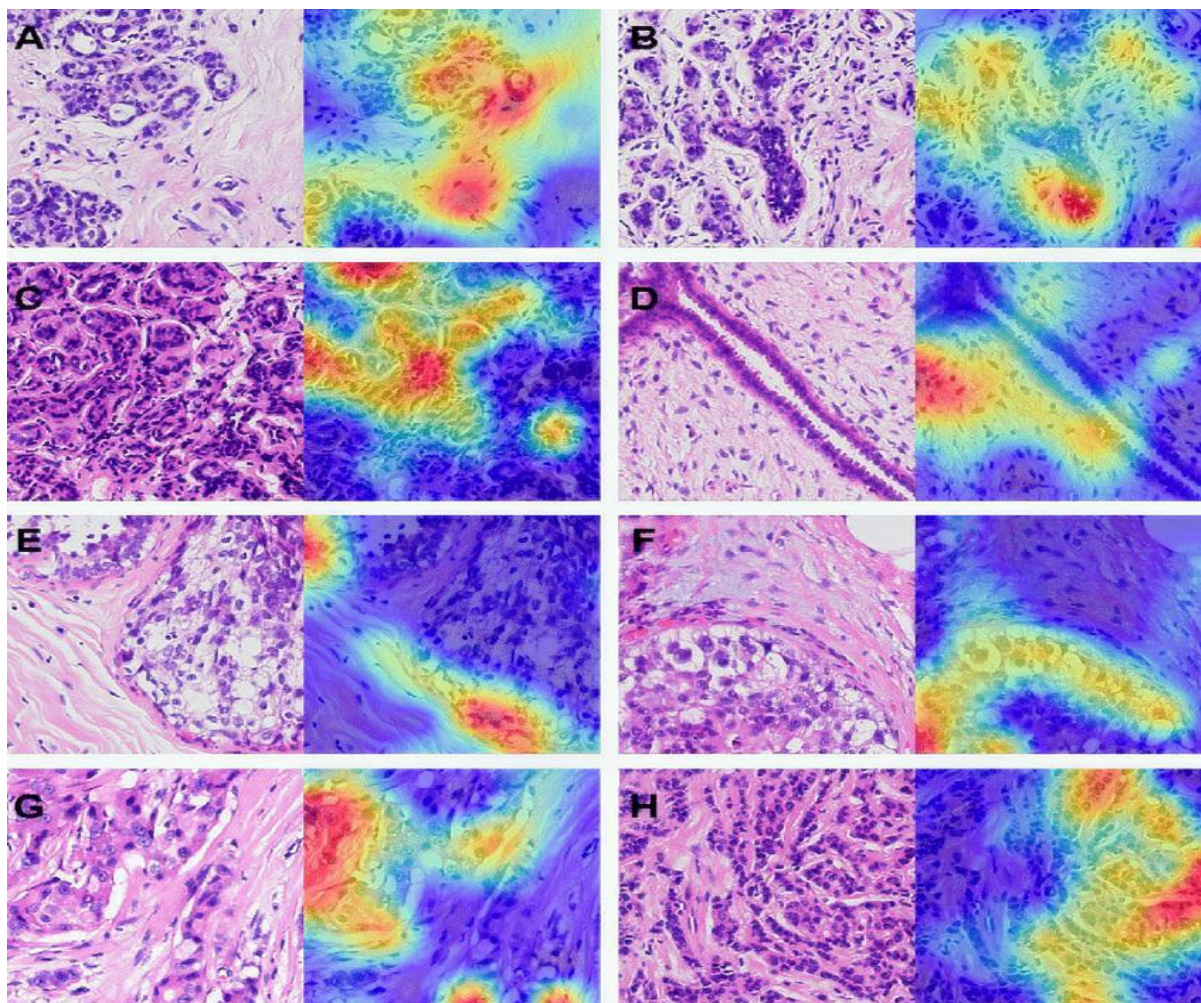


Figure 10. Grad-CAM images for breast tissues. Notes: (A) and (B) Left: Normal tissue images collected from patches (non-cancerous cells). Right: Views of these patches in Grad-CAM. (C) and (D) On left: Patches collected from benign lesions. Right: Presentation for these patches in GRAD-CAM. (E) and (F) Left: Patches collected from in situ carcinoma cells. Right: Presenting in situ cell patches using GRAD-CAM. (G) and (H) Left: Image Patches extracted from invasive carcinoma cells. Right: Grad-CAM's feature extracted image for invasive carcinoma.

The Grad-CAM approach is an effective method for feature selection and attention-based classification, as it helps to highlight the regions of an image that are most responsible for the

model's predictions. This is particularly important in the field of medical image analysis, where the accurate classification of images can have significant implications for patient care.

By using Grad-CAM, we were able to visualize the patterns in the images and the activation around those patterns. This allowed us to gain a better understanding of how the model was making its predictions, and to verify that it was indeed focusing on the most relevant features in the images. The limitations of histology images, which often have similar color combinations, can make it challenging to distinguish between different types of cells. However, by using Grad-CAM, we were able to overcome this limitation and more accurately classify the cells as non-invasive, benign, or malignant. Overall, the Grad-CAM approach is a valuable tool for improving the interpretability of deep learning models, particularly in the field of medical image analysis. By gaining a better understanding of how these models make their predictions, we can increase our confidence in their accuracy and reliability, and ultimately improve patient outcomes.

5. Results and discussion

In total, 162 whole-mount transparencies from breast cancer (BC) tissues are analyzed in this study. From these images, we extracted a dataset consisting of 277,524 regions of size (50x50) and split it 75:25 between a training set and a testing set.

All pictures were optimized for size using the refined Image Data archiving software before being uploaded to the Squeeze network. For the next part of the procedure, we used Adam as our planner, categorical-cross entropy as our cost function, and 0.001 as our learning rate over the course of 32 iterations. A GPU was used during the teaching process. Table 4 displays these early findings.

Table 4. The output obtained after GRAD CAM.

Epoch	Train loss	Valid loss	Error rate	Time
0	0.268622	0.249776	0.103434	04:24
1	0.229270	0.235605	0.097452	04:25

5.1. Confusion Matrix

The confusion matrix depicts true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) throughout the classification task (FN). The output is referred to as TP when the outcomes of actual 1 and the model are comparable. When the output of the model is 0 and the predicted outcome is negative, the output is referred to be TN.

The results are more accurate when there are greater actual successes and true failures. A true negative is when there is no tumor on a slide, while a true positive is when a tumor is discovered. A false alarm refers to the wrong identification of a tumor, whereas a false negative refers to the mistaken identification of no tumor.

Fine-tuned Squeeze Net was trained and Confusion Matrix as a result is obtained as in Figure 11. The confusion matrix as shown in figure clearly shows that the number of TP, TN, FP and FN are 12447, 37222, 2467 and 3368 respectively and the accuracy score of the model turns out to be 90.3% accurate by calculating $TP+TN$ over $TP+TN+FP+FN$ and a competitive Recall Score as required in Medical Image prediction as the target is to reduce False Negatives as much as possible.

We will try to see what the model predicted. In this case, the mistakes look reasonable (none of the mistakes seems naive). This is an indicator that our classifier is working correctly. In Figure 12,

the prediction, actual, loss and probability are clearly shown for few samples of the histology images after running the experiment using Squeeze Net Model.

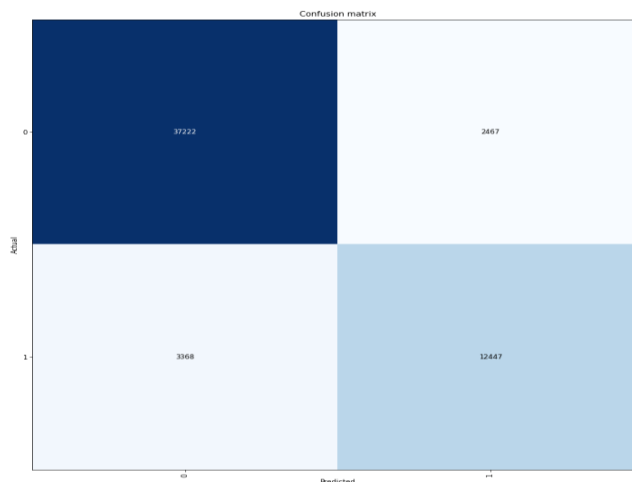


Figure 11. Confusion Matrix for the samples tested.

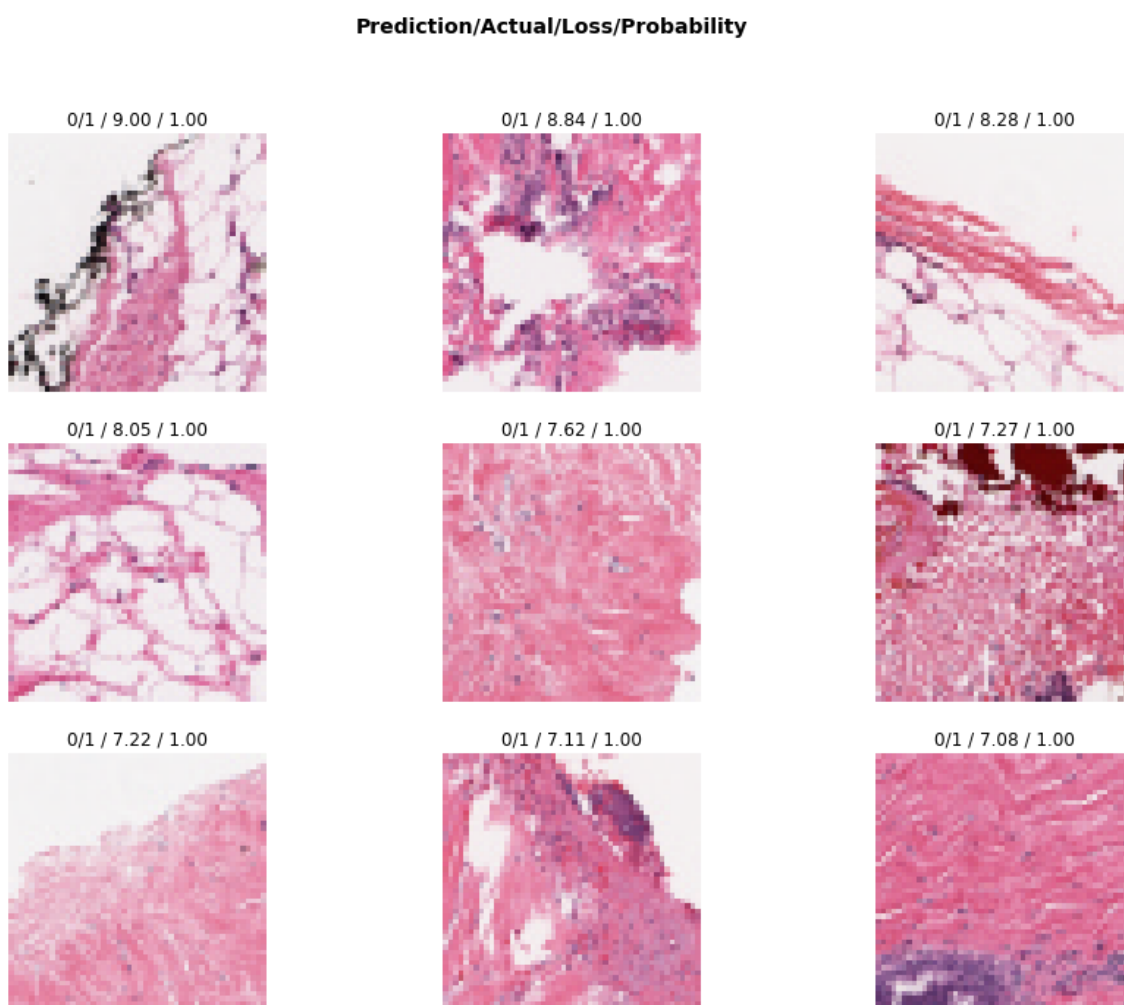


Figure 12. Outcomes as obtained after the testing of the Squeeze Net model.

The results of the fine-tuned Squeeze Net model for texture classification in IDC are quite promising. The confusion matrix obtained shows that the model has a high number of true positives (TP) and true negatives (TN) and a low number of false positives (FP) and false negatives (FN). This indicates that the model is performing well in correctly classifying the texture patterns of the histology images. The accuracy score of 90.3% achieved by the model is quite competitive, especially considering the small size of the Squeeze Net model. The use of a competitive recall score is particularly important in medical image prediction tasks, as reducing false negatives is critical for avoiding misdiagnosis or delayed diagnosis.

The model's performance is further confirmed by the visualization of some sample predictions in Figure 12. The predictions, actual class labels, loss, and probability scores are shown for a few examples. The results show that the model is capable of correctly identifying the texture patterns of the histology images in most cases, and the misclassifications seem reasonable and not naive. This suggests that the model is indeed learning meaningful features from the data and is not overfitting. The use of Squeeze Net in this study demonstrates the importance of designing lightweight models that can be easily deployed in real-world scenarios. The small size of the model not only allows for efficient deployment on mobile devices but also reduces the computational resources required for training and inference. This is especially important in medical imaging, where computational resources may be limited, and the models need to be efficient enough to provide real-time predictions.

Overall, the results of this study demonstrate the effectiveness of transfer learning with super convergence and lightweight models in texture classification for IDC. The use of data augmentation techniques and structure-preserving color normalization further enhances the performance of the model. The attention-based effective feature extraction technique provided by Grad-CAM also adds interpretability to the model's decision-making process.

The promising results obtained in this study open up new possibilities for using lightweight models in medical imaging applications, where computational resources may be limited, and real-time predictions are essential. The fine-tuned Squeeze Net model can be extended to other medical imaging tasks, such as tumor growth analysis, with the potential for a significant impact on improving healthcare in resource-constrained areas.

5.2. Future directions of work

This research enlightens the use of Lightweight deep learning models in biomedical image classification. The greatest drawback of deep learning models is that it requires huge computational resources in terms of memory and GPU. A Lightweight model ensures that the deep learning model learns fewer parameters and yet generates good feature maps using small-sized but powerful Kernels which makes the model very lightweight. The Gradient CAM approach using heat map enables an attention-based classification where more emphasis is given to those sections of the histological patches which are more responsible to determine the type of cell as benign, malignant or non-invasive. The variable learning rate ensures that the gradient is movable and does not get stuck in a local optimum. We propose this model and justify through examples and results that such a lightweight deep learning model can be effective for histological images of the other cancer types as all types of histological images have the same limitations due to the type of stains used. We hope that a heat map-based approach while using Grad-CAM and more optimizations using scheduling of learning rate will give research directions to future researchers while dealing with histopathology

images of other cancer types.

6. Conclusion

In this study, we used transfer learning with super convergence to produce the latest texture classification findings in IDC. The research also considers a DL model that is easy to deploy along with network visuals to help better explain and illustrate how neural networks make decisions. To illustrate the results, a small (4.8 MB) model named Squeeze Net is employed. The outcomes were improved using a variety of data augmentation techniques, including structure-preserving color normalization. In a nutshell, this research addresses the color similarity within the histology images, provides a Grad-CAM based solution for attention-based effective feature extraction technique for classification and also addresses the fact that deep learning models are hard to deploy. Hence a lightweight model has proposed whose network size is reasonably small and easy to deploy even on mobile devices. A variable learning rate is used so that the model does not get stuck in a plateau or local minima and can effectively reach the global minima during gradient descent. We want to use a similar dataset to analyze tumor growth using the learned network in the future. In conclusion, this study presents a novel approach to texture classification in IDC using transfer learning with super convergence. The proposed method not only achieves state-of-the-art results but also addresses the challenges of deploying deep learning models by utilizing a lightweight architecture and network visualization. The incorporation of data augmentation techniques and structure-preserving color normalization further improves the performance of the model. The use of a variable learning rate ensures efficient convergence to the global minima during gradient descent. Moreover, the Grad-CAM based solution for attention-based effective feature extraction technique provides a more interpretable solution for feature visualization and understanding the model's decision-making process. The proposed method can potentially be extended to other medical imaging tasks, such as tumor growth analysis, by fine-tuning the learned network. The use of a lightweight model that can be easily deployed on mobile devices can have a significant impact on improving healthcare in resource-constrained areas. Overall, this study contributes to advancing the field of medical image analysis by providing a more effective and interpretable solution for texture classification in IDC.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. A. Vulli, P. N. Srinivasu, M. S. K. Sashank, J. Shafi, J. Choi, M. F. Ijaz, Fine-Tuned DenseNet-169 for breast cancer metastasis prediction using FastAI and 1-cycle policy, *Sensors*, **22** (2022), 2988. <https://doi.org/10.3390/s22082988>
2. L. Benning, A. Peintner, L. Peintner, Advances in and the applicability of machine learning-based screening and early detection approaches for cancer: A primer, *Cancers (Basel)*, **14** (2022), 1–15. <https://doi.org/10.3390/cancers14030623>

3. T. M. C. Pereira, R. C. Conceição, V. Sencadas, R. Sebastião, Biometric recognition: A systematic review on electrocardiogram data acquisition methods, *Sensors*, **23** (2023), 1507. <https://doi.org/10.3390/s23031507>
4. T. Thakur, I. Batra, M. Luthra, S. Vimal, G. Dhiman, A. Malik, et al., Gene Expression-Assisted Cancer Prediction Techniques, *J. Healthcare Eng.*, **2021** (2021), 1–9. <https://doi.org/10.1155/2021/4242646>
5. G. K. Saini, H. Chouhan, S. Kori, A. Gupta, M. Shabaz, V. Jagota, et al., Recognition of human sentiment from image using machine learning, *Ann. RSCB*, (2021), 1802–1808. <https://www.annalsofrscb.ro/index.php/journal/article/view/4703>
6. S. N. H. Bukhari, A. Jain, E. Haq, M. A. Khder, R. Neware, J. Bhola, Machine learning-based ensemble model for Zika virus T-cell epitope prediction, *J. Healthcare Eng.*, **2021** (2021), 1–10. <https://doi.org/10.1155/2021/9591670>
7. S. Chaudhury, N. Shelke, K. Sau, B. Prasanalakshmi, M. Shabaz, A novel approach to classifying breast cancer histopathology biopsy images using bilateral knowledge distillation and label smoothing regularization, *Comput. Math. Methods Med.*, **2021** (2021), 1–11. <https://doi.org/10.1155/2021/4019358>
8. G. Kaur, S. Bhushan, D. Singh, Fusion in multimodal biometric system: A review, *Indian J. Sci. Technol.*, **10** (2017), 1–10. <https://doi.org/10.17485/ijst/2017/v10i19/114382>
9. R. L. Siegel, K. D. Miller, H. E. Fuchs, A. Jemal, Cancer statistics, 2022, *CA Cancer J. Clin.*, **72** (2022), 7–33. <https://doi.org/10.3322/caac.21708>
10. H. J. Hoffman, A. Khan, K. M. Ajmera, L. Zolfaghari, J. R. Schenfeld, P. H. Levine, Initial response to chemotherapy, not delay in diagnosis, predicts overall survival in inflammatory breast cancer cases, *Am. J. Clin. Oncol.*, **37** (2014), 315–321. <https://doi.org/10.1097/COC.0b013e318271b34b>
11. S. L. Bangare, G. Pradeepini, S. T. Patil, Regenerative pixel mode and tumour locus algorithm development for brain tumour analysis: A new computational technique for precise medical imaging, *Int. J. Biomed. Eng. Technol.*, **27** (2018), 76. <https://doi.org/10.1504/IJBET.2018.093087>
12. R. Kumar, D. Shringi, K. N. Bairwa, Enhancing the tribological behavior of hybrid Al6061 metal matrix composites through the incorporation of nickel and chromium nanoparticles, *Int. J. Adv. Eng. Manag. Sci.*, **6** (2020), 414–420. <https://doi.org/10.22161/ijaems.69.1>
13. R. F. Mustapa, R. Rifin, M. E. Mahadan, A. Zainuddin, Interactive water level control system simulator based on OMRON CX-Programmer and CX-Designer, *Int. J. Emerg. Technol. Adv. Eng.*, **11** (2021), 91–99. https://doi.org/10.46338/ijetae0921_11
14. A. Singh, Mathematical modeling language/tool with disciplinary as a solution strategy in the study of worldwide subjects, *J. Positive Psychol. Wellbeing*, **6** (2022), 1690–1698. <https://www.journalppw.com/index.php/jppw/article/view/2630>
15. V. Jagota, V. Bhatia, L. Vives, A. B. Prasad, ML-PASD, *Adv. Med. Diagn. Treatment Care*. (2021), 82–93. <https://doi.org/10.4018/978-1-7998-7460-7.ch006>
16. Y. Yari, T. V. Nguyen, H. T. Nguyen, Deep learning applied for histological diagnosis of breast cancer, *IEEE Access*, **8** (2020), 162432–162448. <https://doi.org/10.1109/access.2020.3021557>
17. A. Mehbodniya, I. Alam, S. Pande, R. Neware, K. P. Rane, M. Shabaz, M. V. Madhavan, Financial fraud detection in healthcare using machine learning and deep learning techniques, *Secur. Commun. Networks*, **2021** (2021), 1–8. <https://doi.org/10.1155/2021/9293877>

18. G. S. Sajja, Machine learning based detection of depression and anxiety, *Int. J. Computer Appl.*, **183** (2021), 20–23. <https://doi.org/10.5120/ijca2021921856>
19. R. Godasu, D. Zeng, K. Sutrave, Transfer learning in medical image classification: Challenges and opportunities, *MWAIS 2020 Proceedings*, (2020), 5–28. <https://aisel.aisnet.org/mwais2020/18>
20. A. Osareh, B. Shadgar, Machine learning techniques to diagnose breast cancer, *2010 5th Int. Symp. Heal. Informatics Bioinformatics*, **2010** (2010), 114–120. <https://doi.org/10.1109/HIBIT.2010.5478895>
21. N. Bayramoglu, J. Kannala, J. Heikkila, Deep learning for magnification independent breast cancer histopathology image classification, *23rd International Conference on Pattern Recognition (ICPR)*, (2016). <https://doi.org/10.1109/icpr.2016.7900002>
22. C. Pearce, Convolutional Neural Networks and the Analysis of Cancer Imagery, (2017). <http://cs231n.stanford.edu/reports/2017/pdfs/25.pdf>
23. A. Pillai, A. Nizam, M. Joshee, A. Pinto, S. Chavan, Breast cancer detection in mammograms using deep learning, *Adv. Intell. Syst. Comput.*, **1354** (2021), 121–127. https://doi.org/10.1007/978-981-16-2008-9_11
24. Y. S. Vang, Z. Chen, X. Xie, Deep learning framework for multi-class breast cancer histology image classification, *Lecture Notes Computer Sci.*, (2018), 914–922. https://doi.org/10.1007/978-3-319-93000-8_104
25. E. Deniz, A. Şengür, Z. Kadiroğlu, Y. Guo, V. Bajaj, Ü. Budak, Transfer learning based histopathologic image classification for breast cancer detection, *Health Inform. Sci. Syst.*, **6** (2018). <https://doi.org/10.1007/s13755-018-0057-x>
26. S. Chaudhury, N. Shelke, K. Sau, B. Prasanalakshmi, M. Shabaz, A novel approach to classifying breast cancer histopathology biopsy images using bilateral knowledge distillation and label smoothing regularization, *Comput. Math. Methods Med.*, **2021** (2021), 1–11. <https://doi.org/10.1155/2021/4019358>
27. S. Chaudhury, M. Rakhra, N. Memon, K. Sau, M. T. Ayana, Breast cancer calcifications: Identification using a novel segmentation approach, *Comput. Math. Methods Med.*, **2021** (2021), 1–13. <https://doi.org/10.1155/2021/9905808>
28. S. Chaudhury, A. N. Krishna, S. Gupta, K. S. Sankaran, S. Khan, K. Sau, A. Raghuvanshi, et al., Effective image processing and segmentation-based machine learning techniques for diagnosis of breast cancer, *Comput. Math. Methods Med.*, **2022** (2022), 1–6. <https://doi.org/10.1155/2022/6841334>
29. A. Patil, D. Tamboli, S. Meena, D. Anand, A. Sethi, Breast cancer histopathology image classification and localization using multiple instance learning, *2019 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE)*, (2019), 1–4. <https://doi.org/10.1109/WIECON-ECE48653.2019.9019916>
30. F. A. Spanhol, L. S. Oliveira, C. Petitjean, L. Heutte, A dataset for breast cancer histopathological image classification, *IEEE Transact. Biomed. Eng.*, **63** (2016), 1455–1462. <https://doi.org/10.1109/TBME.2015.2496264>
31. F. A. Spanhol, L. S. Oliveira, C. Petitjean, L. Heutte, Breast cancer histopathological image classification using Convolutional Neural Networks, *2016 International Joint Conference on Neural Networks (IJCNN)*, (2016), 2560–2567. <https://doi.org/10.1109/IJCNN.2016.7727519>

32. S.-C. B. Lo, S.-L. A. Lou, Jyh-Shyan Lin, M. T. Freedman, M. V. Chien, S. K. Mun, Artificial convolution neural network techniques and applications for lung nodule detection, *IEEE Transact. Med. Imag.*, **14** (1995), 711–718. <https://doi.org/10.1109/42.476112>
33. B. E. Bejnordi, M. Mullooly, R. M. Pfeiffer, S. Q. Fan, P. M. Vacek, D. L. Weaver, et al., Using deep convolutional neural networks to identify and classify tumor-associated stroma in diagnostic breast biopsies, *Modern Pathol.*, **31** (2018), 1502–1512. <https://doi.org/10.1038/s41379-018-0073-z>
34. J. Howard, S. Gugger, Fastai: A layered API for deep learning, *Information*, **11** (2020), 108. <https://doi.org/10.3390/info11020108>
35. A. Cruz-Roa, A. Basavanhally, F. González, H. Gilmore, M. Feldman, S. Ganesan, et al., Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks, *SPIE Proceed.*, **9041** (2014), 904103. <https://doi.org/10.1117/12.2043872>
36. M. Romano, A. A. Hernandez, Enhanced deep learning approach for predicting invasive ductal carcinoma from histopathology images, *2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD)*, (2019), 142–148. <https://doi.org/10.1109/ICAIBD.2019.8837044>
37. S. J. Pan, Q. Yang, A survey on transfer learning, *IEEE Transact. Knowled. Data Eng.*, **22** (2010), 1345–1359. <https://doi.org/10.1109/TKDE.2009.191>
38. A. M. Dawud, K. Yurtkan, H. Oztoprak, Application of deep learning in neuroradiology: Brain haemorrhage classification using transfer learning, *Comput. Intell. Neurosci.*, **2019** (2019), 1–12. <https://doi.org/10.1155/2019/4629859>
39. F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, K. Keutzer, SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size, (2016), 1–13. <http://arxiv.org/abs/1602.07360>.
40. E. H. Houssein, M. Dirar, L. Abualigah, W. M. Mohamed, An efficient equilibrium optimizer with support vector regression for stock market prediction, *Neural Comput. Appl.*, **34** (2021), 3165–3200. <https://doi.org/10.1007/s00521-021-06580-9>
41. A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet classification with deep convolutional neural networks, *Commun. ACM*, **60** (2017), 84–90. <https://doi.org/10.1145/3065386>
42. R. Ranjbarzadeh, S. Dorosti, S. J. Ghouschi, A. Caputo, E. B. Tirkolaei, S. S. Ali, et al., Breast tumor localization and segmentation using machine learning techniques: Overview of datasets, findings, and methods, *Comput. Biol. Med.*, **152** (2023), 106443. <https://doi.org/10.1016/j.compbiomed.2022.106443>
43. R. Ranjbarzadeh, N. T. Sarshar, S. J. Ghouschi, M. S. Esfahani, M. Parhizkar, Y. Pourasad, et al., MRFE-CNN: Multi-route feature extraction model for breast tumor segmentation in Mammograms using a convolutional neural network, *Ann. Operat. Res.*, (2022). <https://doi.org/10.1007/s10479-022-04755-8>
44. R. Ranjbarzadeh, A. B. Kasgari, S. J. Ghouschi, S. Anari, M. Naseri, M. Bendeche, Brain tumor segmentation based on deep learning and an attention mechanism using MRI multi-modalities brain images, *Sci. Rep.*, **11** (2021), 10930. <https://doi.org/10.1038/s41598-021-90428-8>
45. R. Ranjbarzadeh, S. J. Ghouschi, S. Anari, S. Safavi, N. T. Sarshar, E. B. Tirkolaei, et al., A deep learning approach for robust, multi-oriented, and curved text detection, *Cognit. Comput.*, (2022). <https://doi.org/10.1007/s12559-022-10072-w>

46. R. Ranjbarzadeh, A. Caputo, E. B. Tirkolaei, S. J. Ghouschi, M. Bendeche, Brain tumor segmentation of MRI images: A comprehensive review on the application of artificial intelligence tools, *Comput. Biol. Med.*, **152** (2023), 106405. <https://doi.org/10.1016/j.compbiomed.2022.106405>
47. R. Ranjbarzadeh, S. Dorosti, S. J. Ghouschi, S. Safavi, N. Razmjooy, N. T. Sarshar, et al., Nerve optic segmentation in CT images using a deep learning model and a texture descriptor, *Complex Intell. Syst.*, **8** (2022), 3543–3557. <https://doi.org/10.1007/s40747-022-00694-w>
48. A. Aghamohammadi, R. Ranjbarzadeh, F. Naiemi, M. Mogharrebi, S. Dorosti, M. Bendeche, TPCNN: Two-path convolutional neural network for tumor and liver segmentation in CT images using a novel encoding approach, *Expert Syst. Appl.*, **183** (2021), 115406. <https://doi.org/10.1016/j.eswa.2021.115406>
49. S. Anari, N. T. Sarshar, N. Mahjouri, S. Dorosti, A. Rezaie, Review of deep learning approaches for thyroid cancer diagnosis, *Math. Problems Eng.*, **2022** (2022), 1–8. <https://doi.org/10.1155/2022/5052435>
50. S. Chaudhury, K. Sau, A BERT encoding with recurrent neural network and long-short term memory for breast cancer image classification, *Decision Anal. J.*, **6** (2023), 100177. <https://doi.org/10.1016/j.dajour.2023.100177>
51. D. Singh, J. P. Choudhury, M. De, A comparative study of meta heuristic model to assess the type of breast cancer disease, *IETE J. Res.*, **68** (2022), 3683–3694. <https://doi.org/10.1080/03772063.2020.1775139>



AIMS Press

©2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)