ITC 2/52
Information Technology and Control
Vol. 52 / No. 2 / 2023
pp. 381-396
DOI 10.5755/j01.itc.52.2.33208

Breast Cancer Prognosis Based on Transfer Learning Techniques in Deep Neural Networks

Received 2023/01/15 | Accepted after revision 2023/02/28

HOW TO CITE: Diwakaran, M., Surendran, D. (2023). Breast Cancer Prognosis Based on Transfer Learning Techniques in Deep Neural Networks. *Information Technology and Control*, 52(2), 381-396. https://doi.org/10.5755/j01.itc.52.2.33208

Breast Cancer Prognosis Based on Transfer Learning Techniques in Deep Neural Networks

M. Diwakaran
Department of Information Technology, Sri Krishna College of Engineering and Technology, Coimbatore-641008, India

D. Surendran
Department of Information Technology, Karpagam College of Engineering, Coimbatore-641032, India

Corresponding author: diwakaranmres1@proton.me

Breast cancer is a major cause of death among women in both developed and underdeveloped countries. Early detection and diagnosis of breast cancer are crucial for patients to receive proper treatment and increase their chances of survival. To improve the automatic detection and diagnosis of breast cancer, a new deep learning model called "Breast Cancer Prognosis Based Transfer Learning (BCP-TL)" has been developed. This model uses transfer learning, which applies the knowledge gained from solving one problem to another relevant problem. The model is based on a pre-trained convolutional neural network (CNN) that extracts features from the mammographic image analysis society (MIAS) dataset. Four different CNN architectures were used in this model: AlexNet, Xception, ResNeXt, and Channel Boosted CNN. The performance of the model was evaluated using six metrics, including accuracy, sensitivity, specificity, precision, F1-score, and the area under the ROC curve (AUC). The combination of Xception and Channel Boosted CNN showed excellent performance. By combining essential features from multiple iterations, the Channel Boosted CNN can achieve higher accuracy in breast cancer diagnosis, with an overall accuracy of 98.96%. This highlights the potential of the BCP-TL model in effectively detecting and diagnosing breast cancer.

KEYWORDS: Breast cancer, deep learning, Convolutional Neural Networks, Transfer Learning, Deep Neural Networks.

1. Introduction

Many deadly diseases affect people across the entire world. Cancer is the second most common cause of mortality, as reported by the World Health Organization (WHO) [11]. Notably, the prevalence of female breast cancer is substantially higher in developing nations than in industrialized ones. For instance, 1.42 million people are diagnosed with breast cancer each year, and one-third of them pass away [16]. 10.5 million people worldwide die from cancer each year, while 2 million people survive the disease. Breast cancer is the most prevalent and the leading cause of mortality worldwide for women between the ages of 21 and 60 [27]. The likelihood of surviving breast cancer might rise to 83% if it is detected early [27].

To prevent cancer from spreading to the entire breast as well as other body areas, it is crucial to identify it early and begin treatment [4]. The proper first treatment may boost the breast cancer survival rate by up to 83% following an accurate and successful diagnosis. Breast cancer-related tumors can be classified as benign or malignant. Both groups have the potential to be malignant or not [28]. For instance, benign tumors cause abnormalities in the endothelial cells, but they cannot develop further. Therefore, these won't cause breast cancer. Malignant cells, on the other hand, can be considered tumor cells and are harmful due to their erratic proliferation in the body. In microscopic pictures, distinguishing between benign and malignant tumors requires careful analysis and classification [14].

Mammography and biopsy are the two frequently utilized screening techniques for the early identification of breast cancer. The clinician uses specialized breast images from mammography to look for early signs of cancer in female patients [32]. It has been noted that the fatality ratio has dropped due to the adoption of mammography for cancer detection. Another effective and reliable diagnosis approach for finding breast cancer is a biopsy. In this method, a physician examines a tissue specimen from the breast area impacted under a microscope to identify and classify the tumor. Currently, a biopsy is essential for both the diagnosis of breast cancer and other types of cancers as well [26].

The physician can distinguish between benign and malignant tumors through biopsy. Although the abnormalities in the endothelial cells that make up the benign lesion are indeed malignant, the majority of these aberrations cannot give rise to breast cancer [12]. These cells start their divisions improperly, develop erratically, and are classified as malignant or cancerous. Because regular and malignant cells have sporadic appearances, manually analyzing microscopic images is challenging and complex [5].

Numerous researchers have put up various ideas over the last few years for automated cell categorization for the clinical diagnosis of cancer from breast histology imagery [2]. In this connection, some scientists have focused on mitochondria analysis by extracting characteristics from nuclei to provide critical information for cell classification into benign and malignant types. Similar to how nuclei are segmented and classified, statistical sampling characteristics and training methods based on clustering are also used. Although histomorphological image analysis methods are evolving quickly, an automatic system is still essential to obtaining efficient and highly reliable findings [6].

These procedures are necessary to ensure uniformity in the outcomes during the observation process, improve neutrality, and provide the appropriate orientation toward personal items for diagnosis [18]. In traditional machine learning systems, the sophistication of tasks like pre-processing, fragmentation, extraction, classification, etc., reduces the system's performance in terms of precision and reliability.

Deep learning has been developed as a solution to the limitations of classic machine learning methods. It extracts pertinent data from unprocessed imagery and uses it in identification processes. Deep learning relies on general-purpose learning techniques to learn from data sets rather than directly tweaking features. Convolution Neural Network (CNN) has made significant progress in medical applications and image analysis in recent years [1]. Examples include the detection of tumors from dermoscopic examination, segmentation of neurological cellular membrane, recognition, and identification of lymphocytes, and characterization of volume in screening mammography.

The two-dimensional input-image layout specifically modifies the CNN architecture. The medical industry needs more data for a CNN training task.

The Transfer Learning (TL) technique using pre-trained image datasets, such as ImageNet, and the application of a fine-tuning technique are potential solutions to this issue. By integrating their knowledge, the TL principle can improve the performance of individual CNN designs. The main benefits of TL are increased prediction accuracy and accelerated training. A model transfer is a suitable TL technique that involves pre-training the network configuration using the original data, applying them to the feature space, and then fine-tuning them for better productivity [19].

Diagnosing breast cancer at an early stage using image features and ML algorithms is a challenging task due to several reasons:

1. Limited Image Quality: Early-stage breast cancer often results in subtle changes in the tissue structure, which may not be easily noticeable on mammograms or other imaging modalities. This can lead to limited image quality, making it difficult for ML algorithms to accurately identify and classify the changes as cancerous.
2. Class Imbalance: The number of cases of early-stage breast cancer is much lower compared to advanced breast cancer. This results in a class imbalance problem in training ML models, leading to a higher chance of false negatives and a lower overall accuracy of the model.
3. Overfitting: ML algorithms tend to overfit the training data if the dataset is small, leading to poor generalization on new unseen data, especially when it comes to early-stage breast cancer.
4. Complexity: Breast tissue is complex and varies greatly among individuals, making it difficult to develop ML algorithms that can accurately identify early-stage cancer in all cases.
5. Despite these challenges, advances in deep learning and computer vision techniques are continuously improving the accuracy of breast cancer diagnosis using image features and ML algorithms.

The main goals of this work are the automatic extraction of the afflicted region employing fragmentation, the minimization of training time, and improved prediction performance. In this regard, a system based on transfer learning for breast cancer detection and categorization is suggested. The suggested model contains two key phases. Several preprocessing approaches are used in the first phase to enhance the breast images and perform feature extraction. The learned parameters are then applied to classifying breast cancer using a pre-trained CNN, such as the AlexNet, ResNet, Xception, and Channel Boosted CNN.

1.1. Contributions to the Work

The major contributions of this research are as given below:

1. To propose a novel deep-learning model (BCP-TL) based on the transfer-learning technique for the effective and reliable prognosis of Breast Cancer.
2. To proficiently employ CNN architectures such as AlexNet, ResNet, Xception, and Channel Boosted CNN to classify between benign and malignant lesions with more accurate precision.
3. To demonstrate the performance efficacy of the proposed model by minimizing the training time and retrieving only the impacted regions from breast images.

1.2. Organization of the Paper

The remainder of the paper is organized as follows. Section 2 discusses the state-of-the-art works on Breast Cancer prognosis using Machine Learning and Deep Learning techniques. Section 3 briefs the proposed methodology for breast cancer classification by detailing the CNN architectures such as AlexNet, ResNet, Xception, and Channel Boosted CNN. Section 4 depicts the results and observations derived from the experimental analysis of the proposed model on the MIAS dataset. Section 5 summarizes the discussion on the performance of the proposed model. Section 6 concludes the present work.

2. Related Works

Over the past ten years, research into breast cancer detection has improved. Identifying malignant breast tissue and classifying tumors have been the focus of much research. Some researchers have chosen to develop diagnosis systems using content-based image retrieval techniques, which would have the advantage of providing practitioners with images found in a medical image database [8], whose content is known and comparable to image queries for which the phy-

sician would have doubts. The search time and proper similarity measurement between the requested image and those in the database are issues that this approach also brings up [7]. Using the expert information found in many mammograms that make up the imagery database, Tourassi et al. suggested a semantic search engine for cancer detection. To accomplish this, they use a perfectly matched pattern to identify imagery in the database comparable to the ROI request submitted by the system's user. They devised a decision method that successfully integrates many similarity measures on the best matches to evaluate whether the query ROI contains a malignancy (of any kind) or solely healthy tissue [20].

Authors in [10] favored extracting adjectives that described the edge's thickness, contour, and directionality. Tao et al. [9] combined the parameters linked to center pixel, form, and roughness to discover tumors similar to those in the ROI search and classify them as benign or cancerous. Zheng et al. [30] suggested a system that further interacts with the user to enhance visual similarity; in this system, the user is requested to assess the nature of speckled cancer in the search query so that the algorithm searches for similarities with equivalent degrees of supposition. The query base for this work was consecutively modified by deleting the ROIs that produced the lowest matching score [13].

George et al. [13] have opted to classify breast cancer and detect genetic material using stochastic neural networks and Support Vector Machines. Breast flow cytometry visuals were used in the experiments, and the results were compared based on their margins of error, exact identification rates, responsiveness, and precision. They assert that the outcomes obtained using their particular methods are significantly more potent and apply to several datasets [15]. Sharma et al. presented a thorough study on the categorization of breast cancer using traditional machine learning and deep learning techniques. They extracted the image attributes, texture feature histogram, and Correlation - based patterns to categorize them into benign and malignant lesions. Their suggested strategy produced accuracy levels between 94.36% and 94.39% [23].

A thorough study on the use of ML and DL applications to predict breast cancer was provided by Chugh et al. They reviewed the literature and studies linked to the categorization of breast cancer in great detail. They also emphasized both the advantages and disadvantages of these strategies' characteristics. By incorporating the findings of earlier investigations, the authors of this study concluded that deep learning approaches are considered more appropriate for categorizing breast cancer images when the datasets are more extensive [24]. A systematic and in-depth analysis of deep learning and machine learning methods for breast cancer identification and categorization using medical imaging was presented by Houssein et al. They displayed all of the most recent applications for diagnosing illnesses and the quick adoption of deep learning and machine learning in the medical industry [21]. In their research, Hamed et al. suggested employing models created using ML to classify breast cancer. They asserted that whereas clinicians, on average, diagnose and classify breast cancer with an accuracy of around 75%, their suggested approach does so with an accuracy of 92% [3]. Authors in [25] classified breast cancer pictures using 582 samples and 25 characteristics using the Wisconsin Breast Cancer Dataset. The Kaggle repository was used to get the dataset. They gauged the effectiveness of their labor based on its precision and accuracy. They used artificial neural networks, K Nearest Neighbor, logistic regression, and Support Vector Machines as their methods [22]. To achieve the desired effects, they have employed them independently. They could classify breast cancer imagery with the highest mean precision of 99.1% [29, 17, 31]. A comparative analysis of the methods in the literature for Breast Cancer diagnosis is given in Table 1.

2.1. Research Gap and Motivation of Research

The extensive literature review in this part concludes that deep and transfer learning is not very effectively utilized for early cancer detection. According to the research mentioned, the task can be difficult due to a lack of resources or competent and experienced personnel. Researchers from the medical field have already put in a lot of work, but their conclusions are only sometimes accurate. We have attempted to address these issues by enhancing the process of correctly classifying breast cancer imagery by merging the ideas of deep learning and transfer learning, allowing for the accurate and promising detection of breast cancer in its early stages.

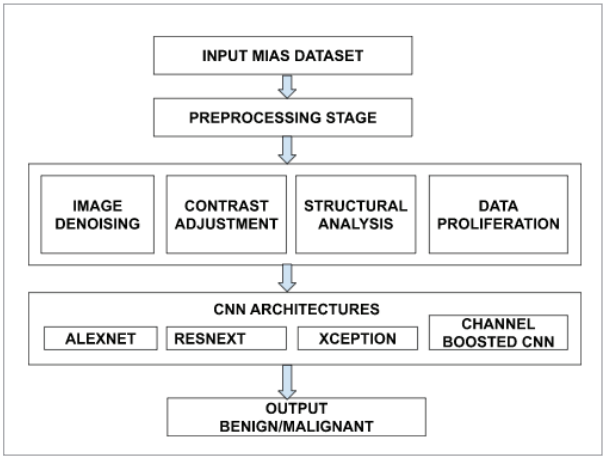
Table 1
Comparative analysis of the existing methods for CVD diagnosis

Reference	Techniques	Dataset	Category	Findings	Limitations
[8]	Random Forest	BreakHis dataset	Breast-mass classification	Sample distribution was balanced	Preprocessing was not performed
[20]	K Nearest Neighbour	Wisconsin breast cancer dataset	Feature extraction	Segmentation of tumors was done automatically	The model was not evaluated with the appropriate parameters
[10]	Naive Bayes and Decision tree	BreakHis dataset	Breast-mass classification	Multiple tasks were processed in parallel	Comparison of models was not implemented
[9]	Deep Belief Network	Wisconsin breast cancer dataset	Feature extraction	Redundant features were trimmed during feature selection	The model was not evaluated with the appropriate parameters
[30]	CNN	BreakHis dataset	Breast-mass classification	Image preprocessing was performed and its effects were analyzed	Sufficient training data was not present
[15]	AlexNet	METABRIC breast cancer dataset	Histopathology	Large datasets were used	Training time was not adequate
[23]	VGGNet-16	Mammogram images	Histopathology	Images collected from Internet sources were utilized	The quality of the collected images was not adequate

3. Proposed Methodology

This section discusses the proposed methodology for breast cancer prognosis using the transfer learning technique. Figure 1 represents the architecture of the proposed BCP-TL model.

Figure 1
Proposed architecture



3.1. Problem Statement

The problem considered for this research is a classification problem with multiple labels in which the input to the problem is a Region of Interest in the image denoted as,

$$a \epsilon \rho^{c} \; with c = a \, x \, a \, . \tag{1}$$

The output of the problem is considered as a group of labels denoted as in (2):

$$b \epsilon \beta = \{0,1\}^{D} . \tag{2}$$

The dataset for the classification problem is represented as shown in (3),

$$DS = \{(a_k, b_k)\}_{k=1}^{N} . \tag{3}$$

For any imagery in the input set, the target is considered from the output set:

$$b = [b_1, b_2, ..., b_D] \epsilon \, b_k . \tag{4}$$

Here, b_k = 1 if the kth data is equivalent to an else b_k = 0. The aim of this problem is to build a model that learns to make predictions using a label set β for any unknown image from ρ^c with the evaluation score as denoted in (5):

$$ev(a):\rho^c -> \rho^D. \quad (5)$$

A label feature transformation is used, which turns each label combination into a unified brand while taking into account any potential connections between labels. The multi-label problem is simplified into a single-label inter-classification problem, with the assortment of unique tags contained in the initial training data serving as the set of potential values for the modified class attribute.

In this problem, a set of different subsets of tags t of T are taken into account with the output space as represented in (6):

$$OS = \delta_{t \subset T} \rho^c. \quad (6)$$

For each category in the output space, the number of possibilities is denoted as in (7):

$$\rho_{t1} = \frac{s!}{(s-p)!} possibilities. \quad (7)$$

The network is altered to produce a set of binary variables suggesting potential ROI markings. An 8-dimensional output is created by swapping out the final completely connected layer for a convolution layer, and then we add a nonlinear sigmoid behavior. The anticipated probability distribution over the labels for each class is the ultimate result.

The loss function is optimized to obtain an additive process of the binary cross-entropy losses available, as shown in (8):

$$\gamma(a,b) = -\frac{1}{K}\Sigma_{i=1}^{D}\left[b_i \log b'_i + (1-b_i)\log(1-b'_i)\right]. \quad (8)$$

The value of b'_i in Equation (8) is represented as shown in (9):

$$b'_i = \frac{1}{1+e^{-\Delta a_i}}. \quad (9)$$

The optimization problem to solve the classification problem for a set of values (a,b) where a denotes an input image and b denotes the associated label is shown in (10):

$$b'_i = \frac{1}{1+e^{-\Delta a_i}}. \quad (10)$$

Here, $\rho(\Delta)$ denotes the regularization factor added to the optimization problem.

Mini-batch gradient descent function is employed to optimize the hyperparameters of the model as shown in (11):

$$\Delta\Omega(i) = \vartheta(i) MBG(t), \quad (11)$$

where $\vartheta(i)$ represents the learning rate of the model and $MBG(t)$ represents the function for Mini-batch gradient descent as shown in (12):

$$MBG(t) = \sigma(\Omega(i)). \quad (12)$$

The updated gradient value is added with a factor after adjustment of slopes is represented in (13):

$$\Delta\Omega(i) = \Phi\Delta\Omega(i-1) - \vartheta(i) MBG(t). \quad (13)$$

The learning rate of the model at the next iteration can be written as shown in (14):

$$\vartheta_1(i) = \vartheta_0 e^{-\mu z}. \quad (14)$$

The modified equation for Mini-batch gradient descent with diminishing gradient descent can be represented as in (15):

$$\begin{aligned}\Delta\Omega_1(i) &= \Phi\Delta\Omega_1(i-1) - \vartheta_1(i) MBG(t) \\ &= \Phi\Delta\Omega_1(i-1) - \vartheta_1(i)\sigma(\Omega(i))\end{aligned}. \quad (15)$$

3.2. Preprocessing Stage

Image preparation is crucial to reduce the restrictions on observing anomalies without an excessive amount of mammography effect. To cut down on computation time, the tumor patches in this study are automatically extracted using feature extraction methods before the learning process. Image denoising, contrast adjustment, and structural analysis performed before

feature extraction can enhance image quality and increase the accuracy of the feature extraction findings.

3.2.1. Image Denoising and Contrast Adjustment

Contrast enhancement is used in mammography scans to modify the contrast, making image anomalies more obvious. The mammography image's quantization noise is filtered using a two-dimensional average filter of 3X3 size. The contrast of the original image is improved across all levels using the traditional histogram-based method. This is achieved by successfully dispersing the image's most common gray level or by extending the image's spectrum.

3.2.2. Structural Analysis

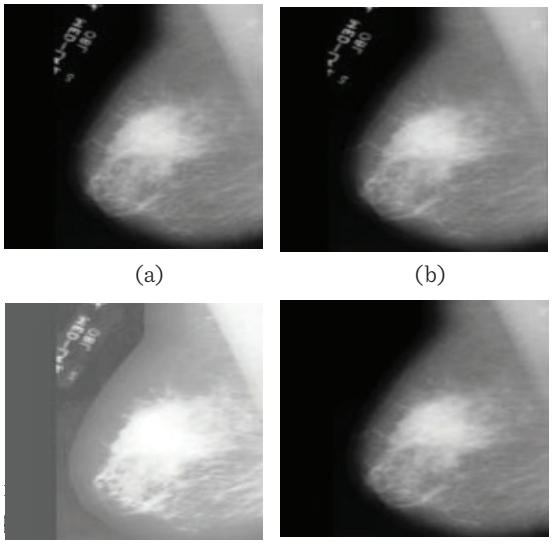
Before feature extraction, it is crucial to remove non-breast sections using structural analysis to ensure that the results are unaffected. When performing structural procedures, the structuring element is applied before the pertinent structures are recovered from the input image (SE). A minimum standard feature extraction method for dynamic edge selection can save computation time and allow the study to concentrate on the area most affected by cancer. This method produces a picture whose size matches the input. Each pixel's value is determined by the neighboring pixels and the matching pixel in the input. To fit the input size of the prie-trained Cnn model, the breast images are shrunk and transformed into three RGB channels.

3.2.3. Data Augmentation

Large datasets are preferable for Deep learning models than smaller ones. When training with very little data, overfitting can be avoided using one of the most common techniques for expanding the dataset: data augmentation. In this work, a set of adjustments can be used to add more images to the training data. To boost the input data, Data augmentation is used. The segmented images are then rotated to 90, 180, 270, and 360 degrees clockwise. Each rotating image is then laterally inverted. Eight images will result from a single input image in this fashion.

Figure 2 represents the results obtained after applying the preprocessing steps. In this figure a) represents the original image. b) represents the image after applying image denoising. c) represents the image after incorporating contrast adjustment. d) represents the result after employing structural analysis.

Figure 2
Proprocessing results



3.3. CNN Architectures

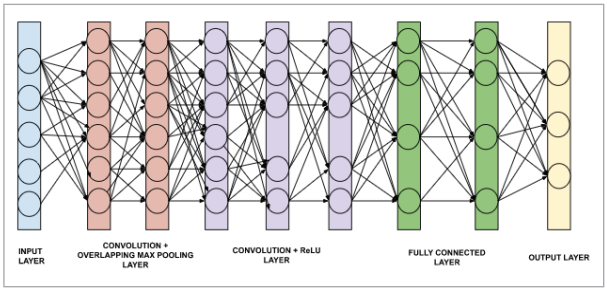
The following Convolutional Neural Network architectures are applied to the given dataset.

3.3.1. AlexNet

Five convolutional layers and three fully connected layers make up AlexNet. Filters that use several Convolutional Kernels to extract useful features from images. There are typically multiple kernels of the same size in one convolutional layer. For instance, AlexNet's first Conv Layer has 96 units, each 11x11x3 in size. The kernel's width and height are typically equal, and its depth is proportional to the number of channels. The architecture of the AlexNet model is depicted in Figure 3.

The Overlapping Max Pooling layers are placed after the initial two Convolutional layers. Direct connections exist between the following consecutive convolutional layers. The Overlapping Max Pooling layer, whose output is fed into a series of two fully connected layers, is placed after the fifth convolutional layer. After all the convolutional and fully linked layers, ReLU nonlinearity is applied. Before pooling, a local equalization step is done after the first and second convolution layers' ReLU nonlinearity.

Figure 3
AlexNet architecture

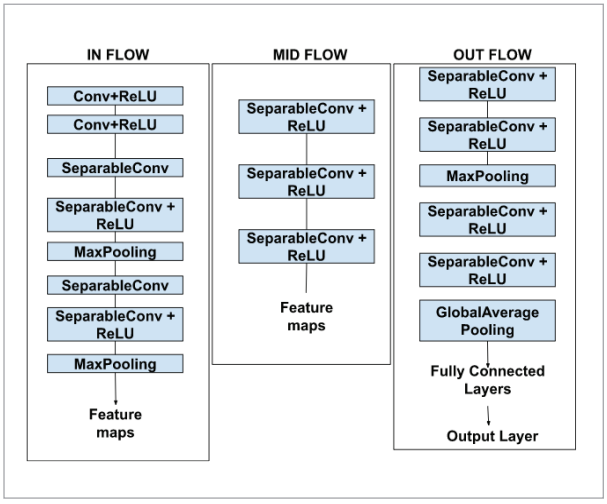


The width and height of the convolutions are often downsampled while maintaining the same depth using Max Pooling layers. Similar to Max Pool layers, overlapping Max Pool layers work the same way and differ only by the method by which the neighboring windows over which the maximum is computed overlap. The pooling windows are usually set to 3X3 in size and have a two stride between adjacent windows. Compared to employing non-overlapping pooling windows of size 22 with a stride of 2, which would yield the exact output dimensions, this overlapping pooling feature helped drastically cut the error rates. The AlexNet uses Rectified Linear Unit Nonlinearity, which is a crucial component. The standard method for training neural network models used to be tanh or sigmoid activation functions. Deep CNNs might be trained significantly more quickly utilizing ReLU nonlinearity than they could use saturating activation functions like tanh or sigmoid, according to AlexNet.

3.3.2. Xception

Depthwise Separable Convolutions are used in the deep convolutional neural network architecture known as Xception. Researchers from Google created it. According to Google, Inception modules in convolutional neural networks serve as a transitional stage between the depthwise separable convolution operation and ordinary convolution. In this context, a depthwise separable convolution can be viewed as an Inception module with the highest limit of columns. With Inception modules replaced with depthwise separable convolutions, Xception is a unique deep convolutional neural network design based on this result. Separable convolutions are replacements for convolutions that are ostensibly significantly faster to compute. The inflow, the mid flow, repeated eight times, and the outflow is all the steps the data must initially go through. Batch equalization comes after every Convolution and SeparableConvolution layer. The architecture of the Xception model is depicted in Figure 4.

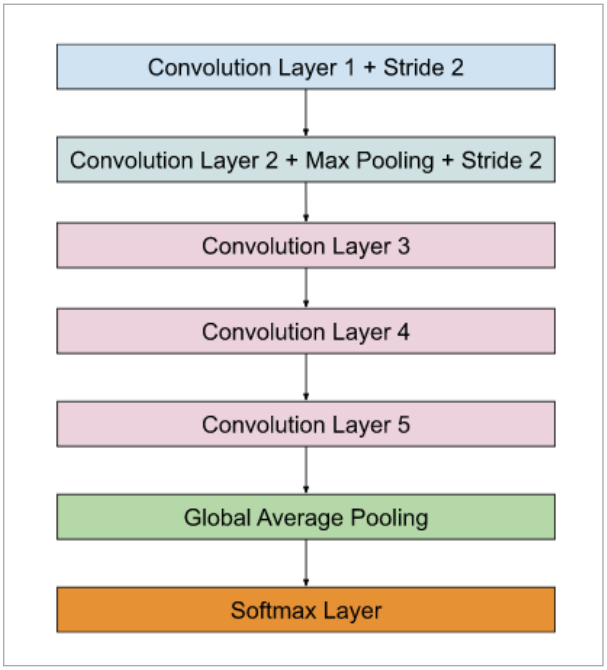
Figure 4
Xception architecture



3.3.3. ResNeXt

A homogeneous neural network called ResNeXt minimizes the amount of hyperparameters needed by traditional ResNet. They accomplish this by adding a third dimension—"cardinality"—to ResNet's width and depth. The number of transitions in the set is determined by cardinality. In this architecture, the transitions are repeated according to the specified cardinality and the outcomes are aggregated to achieve the final result. The divide-transit-append approach of Inception Network is paired with ResNet's repetition strategy in ResNeXt architecture. To put it another way, a network block divides the input, formats it as needed, and then combines it to provide an output where each unit has the same structure. Two rules specify ResNeXt's fundamental architecture. First, if the blocks create contour plots of the same resolution, they use the same set of hyperparameters. Second, if the contour plot is ever downscaled by a factor of 2, the block's width is multiplied by the same factor. The architecture of the ResNeXt model is depicted in Figure 5.

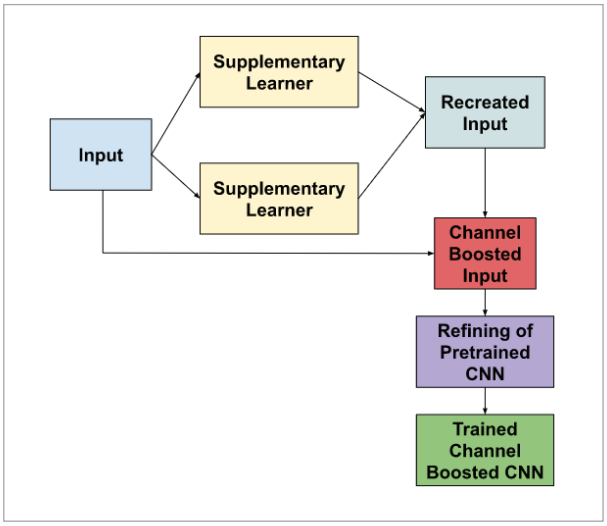
Figure 5
ResNeXt architecture



3.3.4. Channel Boosted CNN

The main motive behind the development of Channel Boosted CNN is to improve the capacity of the representations to a more significant extent. A Channel Boosted input is constructed by combining the essential features from the information taken from several iterations. Any input is first passed through a supplementary learner, which recreates the input and merges it with the original input to produce the boosted intake. The retrieved standard data information and internal and external invariance are added to the input data by supplementing information from numerous supplementary learners. Any meta-model may be used as the supplementary learner in a channel-boosted module; the choice will mostly rely on the nature of the task. The purpose of the supplementary learner is to extract complicated interpretations from the data distribution and so increase the ability of the Channel Boosted CNN to extract information. To create the Channel Boosted input, the data produced by supplementary learners are either concatenated with the input feature networks or some of the input vectors are substituted with rebuilt channels. The architecture of the Channel Boosted CNN model is depicted in Figure 6.

Figure 6
Channel Boosted CNN architecture



Pseudocode for channel-boosted CNN

1 Load mammographic images into the network
2 Preprocess images to correct for variations in brightness, contrast, and resolution
3 Initialize convolutional layers with filters
4 For each image:
  a Apply filters to extract features
  b Apply pooling layers to reduce spatial dimensions
  c Pass features through channel-boosted layer
  d Pass output of channel-boosted layer through fully connected layers
  e Apply softmax activation to generate a probability distribution over two classes (normal vs. abnormal)
  f Choose the class with the highest probability as a prediction
5 Train the network using a large dataset of
6 Mammographic images and corresponding labels
7 Update weights using a loss function (e.g., cross-entropy)
9 Repeat steps 4-6 until the desired accuracy is achieved

The benefits of transfer learning are applied to CNN training in the second phase to enhance adaptation

and shorten the time required for activity. The pre-trained CNN, which provides information via Transfer Learning, is further learned and refined using the boosted input. The network's capacity for learning is enhanced by this additional refining leveraging the pre-trained enabled channels. The use of Transfer Learning has two benefits. First, it lowers the cost of training and improves adaptation to the transferred feature maps from the pre-trained CNN. Secondly, the classifier's ability to represent complex classification issues is enhanced by supplementing the auxiliary channels made accessible by Transfer Learning from the already trained Deep Neural Networks with original feature maps.

4. Result

4.1. Experimental Setup

A computer with an Intel Core i5 CPU running at 2.40 GHz and 4 GB of memory is used for the investigations, which runs on a version of Windows 11 as the processor's operating system. The Python 3.9 language is used to implement the experiments. To implement and test the source code, Python packages like Pandas, Scikit-Learn, Numpy, and Matplotlib are incorporated. The Keras library is used to generate the deep learning model for the CNN architectures, while Theano Framework is used for the backend application.

4.2. Dataset Description

Mammographic Image Analysis Society (MIAS) gave the applicable mammography database used in this study. The primary motivation behind using the MIAS dataset for this research is that it is important to use a standard assessment database (data set) while evaluating an algorithm so that researchers can compare the outcomes immediately. Most mammographic databases are not accessible to the general public. MIAS is the database that is most often utilized for breast cancer-based research since it is the easiest to access the publicly available dataset.

The dataset used for the experimental purpose in this research can be accessed using the following link, https://www.kaggle.com/datasets/kmader/mias-mammography?resource=download.Each image is stored in portable gray map format with a 1024x1024 dimension. In-depth information about the mammography images are provided, including baseline muscle, the class of abnormalities detected, the tumor category, the coordinates of the abnormality center, and an approximation of the circle's circumference. The anomaly class is represented in six formats, as shown in Table 2. 322 imagery in total, divided into three categories—61 for mild cases, 52 for malignant cases, and 209 for standard cases—make up the MIAS dataset.

Table 2
Anomaly representations

Name of the anomaly	Represented Format
Calcification	CALC
Well-defined circumscribed masses	CIRC
Spiculated masses	SPIC
Other ill-defined masses	MISC
Architectural distortion	ARCH
Asymmetry	ASYM

The proposed system is effective for any stage of the patients. This system analyzes the background tissue to categorize it as Fatty, Fatty-glandular, or Dense-glandular. Further, it identifies the type of prevalent abnormality as per the classes specified in Table 2. In addition, the severity of the abnormality is also interpreted. These steps are applicable to patients with the disease at any stage.

4.3. Performance Metrics

The following metrics are used in order to evaluate the performance of the proposed model.

a Accuracy($Perf_{acc}$)

One of the simplest Classification metrics to use is accuracy, which is calculated as the proportion of accurate predictions to all other predictions.

$$Perf_{acc} = \frac{Accurate\ Predictions}{Other\ Predictions} x100\% \quad (16)$$

b Sensitivity($Perf_{sen}$)

This parameter seeks to determine the percentage of actual positives that were mistakenly detected. This is also known as recall.

$$Perf_{sen} = \frac{POS_{true}}{POS_{true} + NEG_{false}} x100\% \quad (17)$$

c Specificity($Perf_{spe}$)

The percentage of true negatives that the model correctly detects is known as specificity.

$$Perf_{spe} = \frac{NEG_{true}}{NEG_{true} + POS_{false}} x100\% \quad (18)$$

d Precision

The fraction of positive predictions that were accurate is determined by precision.

$$Perf_{prec} = \frac{POS_{true}}{POS_{true} + POS_{false}} x100\% \quad (19)$$

e F1-score

A binary classification model is evaluated using the F1 Score metric based on the predictions provided for the positive class.

$$Perf_{F1} = 2 * \frac{Perf_{prec} * Perf_{sen}}{Perf_{prec} + Perf_{sen}} x100\% \quad (20)$$

f Area Under Curve

AUC determines the performance across all levels and offers an overall measurement. AUC has a value between 0 and 1. AUC stands for the area under the curve, and a model with 100% incorrect predictions will have an AUC of 0.0, while a model with 100% correct predictions will have an AUC of 1.0.

4.4. Experimental Results

This section presents the results obtained from experiments using the proposed model on the MIAS dataset. Transfer Learning is applied to four CNN architectures: AlexNet, ResNet, Xception, and Channel Boosted CNN. The performance of these models is evaluated using the metrics specified in Section 4.2.

The performance of the CNN architectures in classifying benign and malignant tumors is evaluated, and the results are tabulated in Table 3. AlexNet architecture produces an accuracy of 93.32%, ResNeXt makes 94.2%, Xception 95.85%, and compared to the other architectures, Channel Boosted CNN has higher accuracy of 96.52%. AUC values for all the architectures are pretty close, with Channel Boosted CNN showing a slightly different value of 98.3%.

Table 3
Performance comparison of CNN architectures

Model	AlexNet	ResNeXt	Xception	Channel Boosted CNN
$Perf_{acc}$(%)	93.32	94.2	95.85	**96.52**
$Perf_{sen}$(%)	91.2	92.5	93.4	**94.2**
$Perf_{spe}$(%)	93.56	93.89	96.2	**97.1**
$Perf_{prec}$(%)	92.3	92.6	93.8	**94.5**
$Perf_{F1}$(%)	92.8	93.9	94.78	**96.4**
$Perf_{AUE}$(%)	98.2	98.2	98.2	**98.3**

Channel Boosted CNN offers better performance with increased data than other architectures. Figure 7 depicts the performance comparison of the CNN architectures for Breast Cancer Classification. The performance of the proposed architectures on Proliferated data is also evaluated, and the observations are shown in Table 4.

Channel Boosted CNN shows an accuracy of 97.45%, a sensitivity of 96.8%, a specificity of 97.8%, a precision of 96.5%, an F1 score of 96.8%, and an AUC of 99.1%. The accuracy obtained by the other models is

Figure 7
Performance comparison of CNN architectures

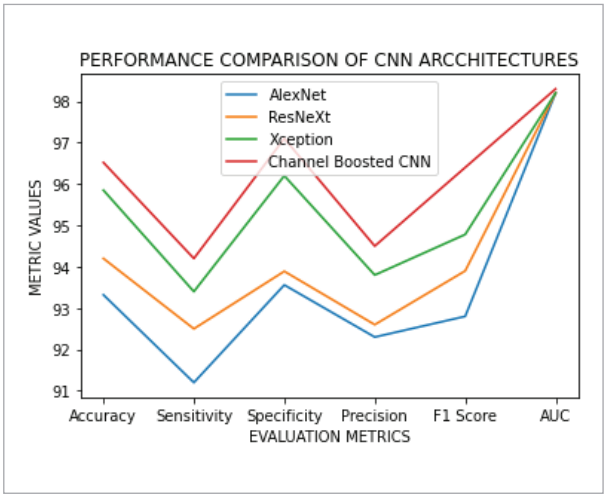


Table 4
Performance evaluation on proliferated data

Model	AlexNet	ResNeXt	Xcep-tion	Channel Boosted CNN
$Perf_{acc}$(%)	92.36	93.52	95.78	**97.45**
$Perf_{sen}$(%)	92.2	92.5	94.8	**96.8**
$Perf_{spe}$(%)	93.56	93.78	95.9	**97.8**
$Perf_{prec}$(%)	93.4	93.65	94.7	**96.5**
$Perf_{F1}$(%)	92.4	92.7	94.9	**96.8**
$Perf_{AUE}$(%)	97.4	97.8	97.9	**99.1**

comparatively low, with AlexNet 92.36%, ResNeXt 93.52%, and Xception 95.78%, respectively.

The models are also executed under two cross-validation techniques: five-fold cross-validation and ten-fold cross-validation. The analysis results are recorded in Tables 5 and 6 for five-fold and ten-fold cross-validation, respectively. For both cases, the Xception model performed better than the other models under consideration. AlexNet, ResNeXt, and Channel Boosted CNN produced an accuracy of 92.5%, 93.4%, and 94.5%, respectively. The Xception model made an accuracy of 96.78% for five-fold cross-validation.

Table 5
Performance comparison using 5 fold cross validation

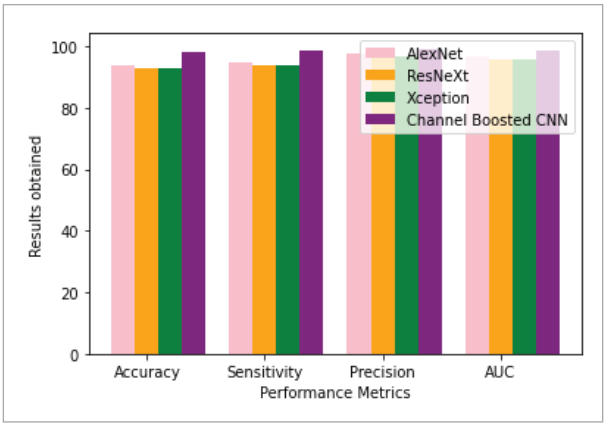
Model	AlexNet	ResNeXt	Xception	Channel Boosted CNN
$Perf_{acc}$(%)	92.5	93.4	**96.78**	94.5
$Perf_{sen}$(%)	93.6	92.4	**95.8**	93.7
$Perf_{spe}$(%)	92.8	93.6	**96.6**	94.8
$Perf_{prec}$(%)	93.4	92.7	**95.7**	93.8
$Perf_{F1}$(%)	93.5	94.7	**95.8**	94.2
$Perf_{AUE}$(%)	98.1	98.2	**98.9**	97.9

Table 6
Performance comparison using 5 fold cross validation

Model	AlexNet	ResNeXt	Xception	Channel Boosted CNN
$Perf_{acc}$(%)	93.6	94.9	97.5	96.8
$Perf_{sen}$(%)	92.8	93.7	96.8	95.8
$Perf_{spe}$(%)	93.6	94.5	97.2	96.9
$Perf_{prec}$(%)	92.7	93.8	96.7	95.8
$Perf_{F1}$(%)	93.4	94.66	97.5	96.9
$Perf_{AUE}$(%)	98.3	98.5	99.2	98.7

Though Channel Boosted CNN produced a higher performance for the original data and increased data, the Xception model demonstrated better performance under the cross-validation techniques. Similarly, for the ten-fold cross-validation technique, the Xception model produced an accuracy of 97.5%, a sensitivity of 96.8%, a Specificity of 97.2%, a Precision of 96.7%, an F1 score of 97.5%, and an AUC of 99.2%. Figure 8 compares the performance exhibited by each model under ten-fold cross-validation.

Figure 8
Performance comparison using 10 fold cross validation



Further, two models were combined and executed on the MIAS dataset to observe the performance, and the results are presented in Table 7. The models taken for experimental analysis are AlexNet with ResNeXt,

Table 7
Performance comparison on combined CNN architectures

Model	AlexNet+ ResNeXt	ResNeXt+ Xception	AlexNet+ Xception	ResNeXt+ Channel Boosted CNN	AlexNet+ Channel Boosted CNN	Xception+ Channel Boosted CNN
$Perf_{acc}$(%)	95.2	96.3	95.4	96.9	97.4	98.96
$Perf_{sen}$(%)	94.5	95.6	94.8	95.6	96.7	98.5
$Perf_{spe}$(%)	95.8	96.7	95.7	96.8	96.7	98.6
$Perf_{prec}$(%)	94.6	95.8	95.9	96.2	97.2	98.7
$Perf_{F1}$(%)	95.9	96.8	95.9	96.5	97.1	98.5
$Perf_{AUE}$(%)	98.6	98.5	98.3	98.9	99.3	99.7

ResNeXt with Xception, AlexNet with Xception, ResNeXt with Channel Boosted CNN, AlexNet with Channel Boosted CNN, and Xception with Channel Boosted CNN. Among these combinations of models, Xception with Channel Boosted CNN showed a higher accuracy of 98.96%. While AlexNet with Channel Boosted CNN showed an accuracy of 97.4%, AlexNet with ResNeXt showed the least accuracy of 95.2%. AUC values were also higher for Xception with the Channel Boosted CNN model with 99.7%.

The proposed model was compared with the existing models, and the results are as per the observations detailed in Table 8. Figure 9 compares proposed and current models in terms of performance metrics. Authors in [5] executed transfer learning for breast cancer prediction using VGG models and achieved an accuracy of 95.6%. The work proposed in [17] was based on Inception models, which produced an accuracy of 96.3%. The work in [21] employed InceptionNet and ResNet, which exhibited 97.5% accuracy. The researchers in [23] demonstrated an accuracy of 97.9% for breast cancer classification with the DenseNet model. However, the proposed model outperformed the other with an accuracy of 98.96%.

Figure 9
Performance of the proposed vs existing models

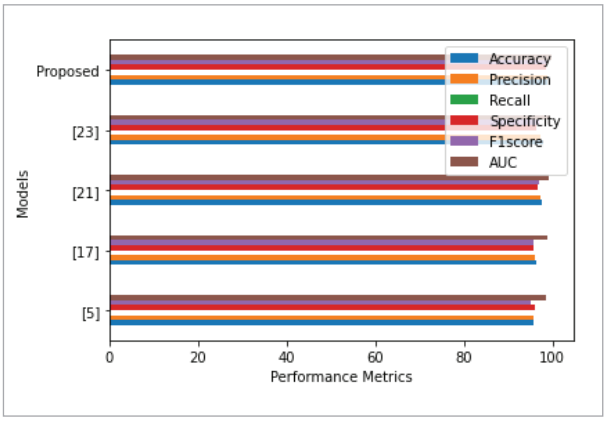


Table 8
Performance comparison between the proposed vs existing models

Model	[5]	[17]	[21]	[23]	Proposed Model
$Perf_{acc}$(%)	95.6	96.3	97.5	97.9	98.9
$Perf_{sen}$(%)	94.8	95.6	96.8	96.5	98.5
$Perf_{spe}$(%)	95.9	95.8	96.5	96.4	98.6
$Perf_{prec}$(%)	95.8	96.1	97.2	97.3	98.7
$Perf_{F1}$(%)	95.2	95.6	96.9	96.8	98.5
$Perf_{AUE}$(%)	98.6	98.7	99.2	99.4	99.7

In the proposed system, Data parallelism and model parallelism are coupled to create mixed parallelism, also known as hybrid parallelism, in order to obtain a higher level of parallelism. The Hybrid Parallelism Technique divides the CNN architectures so that convolutional layers and pooling layers can utilize data parallelism and fully connected levels can exploit model parallelism because the complexity of the various

CNN layers varies. This type of parallelism technique is adopted in order to train deeper networks using a sufficient quantity of resources. As a result, communication cost and overhead are minimized by a factor of 1.1-23.0 in comparison to these techniques being employed individually. Application of Hybrid parallelism in the proposed system also outperformed existing systems in terms of runtime performance.

5. Discussion

Investigations demonstrated that transfer learning, even between unrelated tasks, benefits our job of concern. Initializing with weights already taught is a decent technique to begin the learning process. The loaded weights will then be gradually refined to synchronize the network with the latest dataset. Resuming backpropagation on the layers with a slow learning rate allows for this. Overfitting is the outcome of too much refining, which produces subpar results. The standard method for refining freezes the network's top layers and propagates backward through a few of the final convolutional layers. This is based on the idea that whereas the last layers of a CNN tend to be more unique to the data, the initial layers learn generic features. Since the early convolutional layers are already optimized to learn generic features, especially when we are missing data to train on, there is no need to adjust their weights significantly. Instead, the goal is to have the last convolutional layers learn more data-specific features.

This concept is improved by refining the per-exponentially declining learning rate, which causes acclimation to occur more naturally and automatically than its opponent. The rate at which weights change for each network component can be regulated using the suggested method. In other words, some of the layers can be more or less responsive to change while refining. The final convolutional layers will be heavily represented, while the initial layers will be changed the least. Given that the intermediate convolutional layers may be in charge of picking up slightly more complicated characteristics than the initial layers, they need to be substantially altered. The finest method for assisting us in this is the auto-regressive learning rate.

The proposed technique reduced the risk of overfitting caused by the lack of data. Empirically, the model's performance was consistent across small and comparatively more significant datasets. The proposed model avoided overfitting by refining the declining learning rate for every layer, accounting for label associations during the categorization process, and using other techniques like feature extraction, regularization, and dropout. The suggested system can provide comprehensive observations, giving the physician a comprehensive view of the region of interest. This can aid in decision-making and eventually help him improve the accuracy of the diagnoses.

6. Conclusion

An innovative deep-learning model for enhancing the breast cancer classification outcomes on the MIAS dataset was put forth in this research. This approach is meant to aid physicians in the detection and diagnosis of Breast Cancer. The three types of MIAS images—benign, malignant, and normal—were separated. The original MIAS dataset underwent pre-processing to denoise noise, enhance breast picture contrast, remove non-breast regions, and identify the malignant area. To improve the CNN structure's performance, the data augmentation concept was also suggested. The experimental results on the MIAS dataset were performed under various combinations, and finally, Xception with Channel Boosted CNN produced higher accuracy of 98.96%.

Furthermore, a definite advancement over other current methods can be accomplished by incorporating CNN architectures with transfer learning into the screening mechanism. The sensitivity, specificity, precision, F1 score, and AUC values produced by the proposed model are 98.5%, 98.6%, 98.78%, 98.5%, and 99.7%, respectively. One limitation of the present research is that it uses only mammography images for experimental purposes. Future research will focus on finding strategies to use the unique characteristics of the non-linear problems to develop a joint image-label embedding that characterizes both the contextual attribute dependency and the relevance of the model relationship. In order to construct an effective, reliable, and potent computer-aided diagnosis system for early breast cancer diagnosis, we would also like to use imaging modalities other than mammography in the learning process.

References

1. Ahmad, S., Ur, Rehman, S., Iqbal, A., Farooq, R. K., Shahid, A., Ullah, M. I., Breast Cancer Research in Pakistan: A Bibliometric Analysis. SAGE Open, 2021, 11(3). https://doi.org/10.1177/21582440211046934
2. Agrawal, D. K., Kirar, B. S., Pachori, R. B. Automated Glaucoma Detection Using Quasi-Bivariate Variational Mode Decomposition From Fundus Images. IET Image Processing, 2019, 13(13), 2401-2408. https://doi.org/10.1049/iet-ipr.2019.0036
3. Al-antari, M. A., Al-masni, M. A., Choi, M. T., Han, S. M., Kim, T. S. A Fully Integrated Computer-Aided Diagnosis System for Digital X-Ray Mammograms via Deep Learning Detection, Segmentation, and Classification. International Journal of Medical Informatics, 2018, 117, 44-54. https://doi.org/10.1016/j.ijmedinf.2018.06.003
4. Brancati, N., De Pietro, G., Frucci, M., Riccio, D. A Deep Learning Approach for Breast Invasive Ductal Carcinoma Detection and Lymphoma Multi-Classification in Histological Images. IEEE Access, 2019, 7, 44709-44720. https://doi.org/10.1109/ACCESS.2019.2908724
5. Chaudhary, P. K., Pachori, R. B. Automatic Diagnosis of Glaucoma Using Two-Dimensional Fourier-Bessel Series Expansion Based Empirical Wavelet Transform. Biomedical Signal Processing and Control, 2021, 64, 102237. https://doi.org/10.1016/j.bspc.2020.102237
6. Chatterjee, S., Dey, D., Munshi, S. Empirical Wavelet Fractal Texture Analysis for Skin Disease Identification. IEEE, 2020, 819-822. https://doi.org/10.1109/TENSYMP50017.2020.9230698
7. Chougrad, H., Zouaki, H., Alheyane, O. Convolutional Neural Networks for Breast Cancer Screening: Transfer Learning with Exponential Decay. Proceedings of the NIPS-Machine Learning of Health Workshop, 2017, 1711(10752).
8. Chugh, G., Kumar, S., Singh, N. Survey on Machine Learning and Deep Learning Applications in Breast Cancer Diagnosis. Cognitive Computation, 2021, 13(6), 1451-1470. https://doi.org/10.1007/s12559-020-09813-6
9. Danaee, P., Ghaeini, R., Hendrix, D. A. A Deep Learning Approach for Cancer Detection and Relevant Gene Identification. In Proceedings in Biocomputing, 2017, 219-229.
10. Dhungel, N., Carneiro, G., Bradley, A. P. Automated Mass Detection in Mammograms Using Cascaded Deep Learning and Random Forests. In Proceedings of International Conference on Digital Image Computing: Techniques and Applications (DICTA), 2015, 1-8. https://doi.org/10.1109/DICTA.2015.7371234
11. Galassi, A., Lippi, M., Torroni, P. Attention, Please! A Critical Review of Neural Attention Models in Natural Language Processing. Arxiv, 2019. arXiv:1902.02181.
12. Ghosh, S. K., Ghosh, A. Classification of Gene Expression Patterns Using a Novel Type- 2 Fuzzy Multi Granulation-Based SVM Model for the Recognition of Cancer Mediating Biomarkers. Neural Computing and Applications, 2021, 33(9), 4263-4281. https://doi.org/10.1007/s00521-020-05241-7
13. Gupta, P., Kaur, Malhi, A. Using Deep Learning to Enhance Head and Neck Cancer Diagnosis and Classification. In 2018 IEEE International Conference on System, Computation, Automation and Networking (Icscan), 2018, 1-6. https://doi.org/10.1109/ICSCAN.2018.8541142
14. Kausar, T., MingJiang, W., Ashraf, M. A., Kausar, A. SmallMitosis: Small Size Mitotic Cells Detection in Breast Histopathology Images. IEEE Access, 2020. https://doi.org/10.1109/ACCESS.2020.3044625
15. Khan, S., Islam, N., Jan, Z., Ud Din, I., Rodrigues, J. J. P. C. A Novel Deep Learning Based Framework for the Detection and Classification of Breast Cancer Using Transfer Learning. Pattern Recognition Letters, 2019, 125, 1-6. https://doi.org/10.1016/j.patrec.2019.03.022
16. Man, R., Yang, P., Xu, B. Classification of Breast Cancer Histopathological Images Using Discriminative Patches Screened by Generative Adversarial Networks. IEEE Access, 2020, 8, 155362-155377. https://doi.org/10.1109/ACCESS.2020.3019327
17. Maqsood, S., Damaševičwasius, R., Maskeliūnas, R. TTCNN: A Breast Cancer Detection and Classification Towards Computer-Aided Diagnosis Using Digital Mammography in Early Stages. Applied Sciences, 2022, 12(7), 3273. https://doi.org/10.3390/app12073273
18. Nguyen, H. G., Blank, A., Dawson, H. E., Lugli, A., Zlobec, I. Classification of Colorectal Tissue Images from High Throughput Tissue Microarrays by Ensemble Deep Learning Methods. Scientific Reports, 2021, 11(1), 1-11. https://doi.org/10.1038/s41598-021-81352-y
19. Pandian, A. P. Identification and Classification of Cancer Cells Using Capsule Networks with Pathological Images, Journal of Artificial Intelligence and Capsule Networks, 2019, 1(1), 37-44. https://doi.org/10.36548/jaicn.2019.1.005

20. Platania, R., Shams, S., Yang, S., Zhang, J., Lee, K., Park, S. J. Automated Breast Cancer Diagnosis Using Deep Learning and Region of Interest Detection (BC-DROID). In Proceedings of 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, 2017, 536-543. https://doi.org/10.1145/3107411.3107484

21. Qiu, Y., Yan, S., Gundreddy, R. R., Wang, Y., Cheng, S., Liu, H., Zheng, B. A New Approach to Develop Computer-Aided Diagnosis Scheme of Breast Mass Classification Using Deep Learning Technology. Journal of X-Ray Science and Technology, 2017, 25(5), 751-763. https://doi.org/10.3233/XST-16226

22. Quist, J., Taylor, L., Staaf, J., Grigoriadis, A. Random Forest Modeling of High-Dimensional Mixed-Type Data for Breast Cancer Classification. Cancers, 2021, 13(5), 991. https://doi.org/10.3390/cancers13050991

23. Rakhlin, A., Shvets, A., Iglovikov, V., Kalinin, A. A. Deep Convolutional Neural Networks for Breast Cancer Histology Image Analysis. In Proceedings of International Conference on Image Analysis and Recognition. Cham, Switzerland: Springer, 2018, 737-744. https://doi.org/10.1007/978-3-319-93000-8_83

24. Rodríguez-Ruiz, A., Krupinski, E., Mordang, J. J., Schilling, K., Heywang-Köbrunner, S. H., Sechopoulos, I., Mann, R. M. Detection of Breast Cancer with Mammography: Effect of an Artificial Intelligence Support System. Radiology, 2019, 290(2), 305-314. https://doi.org/10.1148/radiol.2018181371

25. Shamy, S., Dheeba, J. A Research on Detection and Classification of Breast Cancer using K- Means GMM & CNN Algorithms. International Journal of Engineering and Advanced Technology, 2019, 11, 315-320.

26. Shu, X., Zhang, L., Wang, Z., Lv, Q., Yi, Z. Deep Neural Networks with Region-Based Pooling Structures for Mammographic Image Classification. IEEE Transactions on Medical Imaging, 2020, 39(6), 2246-2255. https://doi.org/10.1109/TMI.2020.2968397

27. Valkonen, M., Isola, J., Ylinen, O., Muhonen, V., Saxlin, A., Tolonen, T., Nykter, M., Ruusuvuori, P. Cytokeratin-Supervised Deep Learning for Automatic Recognition of Epithelial Cells in Breast Cancers Stained for ER, PR, and Ki-67. IEEE Transactions on Medical Imaging, 2020, 39(2), 534-542. https://doi.org/10.1109/TMI.2019.2933656

28. Wang, Y.I., Wang, N. A., Xu, M., Yu, J., Qin, C., Luo, X., Yang, X., Wang, T., Li, A., Ni, D. Deeply-Supervised Networks With Threshold Loss for Cancer Detection in Automated Breast Ultrasound. IEEE Transactions on Medical Imaging, 2020, 39(4), 866-876. https://doi.org/10.1109/TMI.2019.2936500

29. Wu, J., Hicks, C., Breast Cancer Type Classification Using Machine Learning. Journal of Personalized Medicine, 2021, 11(2), 61. https://doi.org/10.3390/jpm11020061

30. Yap, M. H., Pons, G., Marti, J., Ganau, S., Sentis, M., Zwiggelaar, R., Davison, A. K., Marti, R. Automated Breast Ultrasound Lesions Detection Using Convolutional Neural Networks. IEEE Journal of Biomedical and Health Informatics, 2018, 22(4), 1218-1226. https://doi.org/10.1109/JBHI.2017.2731873

31. Zebari, D. A., Ibrahim, D. A., Zeebaree, D. Q., Mohammed, M. A., Haron, H., Zebari, N. A., Damaševičius, R., Maskeliūnas, R. Breast Cancer Detection Using Mammogram Images with Improved Multi-Fractal Dimension Approach And Feature Fusion. Applied Sciences, 2021, 11(24), 12122. https://doi.org/10.3390/app112412122

32. Zheng, J., Lin, D., Gao, Z., Wang, S., He, M., Fan, J. Deep Learning Assisted Efficient Adaboost Algorithm for Breast Cancer Detection and Early Diagnosis. IEEE Access, 2020, 8, 96946-96954. https://doi.org/10.1109/ACCESS.2020.2993536