

Deep Learning Algorithms for Diagnosis of Breast Cancer with Maximum Likelihood Estimation

Mehmet Akif Cifci(✉) and Zafer Aslan

Faculty of Engineering, Department of Computer Engineering, Istanbul Aydin University, 34295 Istanbul, Turkey
wwwakif@msn.com, zaferaslan@aydin.edu.tr

Abstract. Machine Learning (ML) and particularly Deep Learning (DL) continue to advance rapidly, attracting the attention of the health imaging community to apply these techniques to increase the precision of cancer screening. The most common cancer in women is breast cancer that affects more than 2 million women each year and causes the largest number of deaths from cancer in the female population.

This work provides state-of-the-art research on the contributions and new applications of DL for early diagnosis of breast cancer. Also, it emphasizes on how and which major applications of DL algorithms are going to be benefitted for early diagnosis of breast cancer for which CNNs, one of the DL architectures, will be used. In this study, a DL method to be used for diagnostic and prognostic analysis using the X-ray breast image dataset for breast cancer is studied. Based on the dataset, it is aimed to diagnose breast cancer at an early stage. Thus, it may take place before a clinical diagnosis. For the testing probability of the disease, 21400 X-ray breast images, both normal and cancer, were taken from USF mammography datasets. From these images, 70% is used as the training step, while 30% of images are benefitted for the testing step. After the implementation of the architecture, VGG16 has achieved an overall accuracy of 96.77%, with 97.04% sensitivity and 96.74% as AUC, while Inception-v4 has an overall accuracy of 96.67%, with 96.03% sensitivity and 99.88% as AUC. These results show the high value of using DL for early diagnosis of breast cancer. The results are promising.

Keywords: CNN · Deep learning · Transfer learning · Breast cancer · X-ray

1 Introduction

Breast Cancer is an important public health problem worldwide which is more common among women in the world [1]. Digital mammography is the normal treatment for the diagnosis of breast cancer today; in the field of medical diagnosis, different techniques are used for the classification issue. However, mammography has some limitations, such as difficulty in tumor detection in young patients or in massive cancers [2].

The image extraction function is a significant step in classifying mammograms. Extract these attributes using digital image processing. Breast cancer is a form of

© Springer Nature Switzerland AG 2020
O. Gervasi et al. (Eds.): ICCSA 2020, LNCS 12250, pp. 486–502, 2020.
https://doi.org/10.1007/978-3-030-58802-1_35

cancer that occurs in breast tissues and often occurs on the internal surfaces of the milk ducts [3]. Although the prevalence of human cases is high in women, breast cancer may also occur in men. The properties of cancer determine treatment; surgery, drug therapy radiation, and immunotherapy can be included. Prognosis and survival rates vary widely depending on the type of cancer, its stages, treatment, and the patient's geographic location. If not detected in the early stages, the probability of cure for breast cancer falls significantly [4]. The methodology of Deep Learning (DL) focuses on implementing nonlinear transformations and high-level model abstractions on large datasets. Especially recent developments in DL architectures in myriad fields have already made significant contributions to Artificial Intelligence (AI).

Radiologists nowadays diagnose the breast tumor manually in mammographic or other radiological images, which is a time-consuming and error-prone procedure due to its small size and different lesion shapes, and these manual images are impaired by the existence of low contrast and unclear boundaries between the normal tissues surrounding it [5]. It has been proven that even the most experienced radiologists miss 10–30% of breast cancers during a routine screening in manual diagnosis [6]. When the fatalities are taken into consideration, such as in the USA, breast cancer the leading cause for deaths while it is number one worldwide, and automatic detection becomes a must. In the meantime, the automatic diagnosis is done with Convolutional Neural Networks (CNNs), which attract attention and become popular in image classification. CNNs are multi-layered structures, which is simple when training parameters are less. CNNs are more like biological neural networks, thanks to its weight-sharing network structure in which there are layers called convolution, activation, pooling, and fully connected layers. CNNs have a loss function like softmax in the last layer. Traditional neural networks contain trainable weights; generally, these layers are one-dimensional. The neurons in the layer are completely connected to neighboring neurons [7]. CNNs have been the most prominent developments in the fields of computer vision and AI, which are a sort of mixture of genetics, mathematics, and computer science [8]. Later 2012 was the first year in which CNN rose in popularity when Alex Krizhevsky used an 8-layer CNN to win that year's ImageNet competition (later referred to as AlexNet) [5]. Since then CNN has been the focus of interest. It has been utilized by many companies such as Google, Facebook, etc. CNNs are influenced biologically by the visual cortex which has small cell regions sensitive to different visual field regions. For instance, some neurons fired on vertical edges and some on horizontal or diagonal edges.

Similar tasks for different components in a system are often used by machines and are the basis behind CNNs. A common belief in the profound learning community is that successful profound learning models cannot be produced without a large number of data. Though data is a vital part of networking, the concept of transfer learning reduces data requirements if the data is already processed. Transfer learning is a preformed model and a small dataset that can be modified to fine-tune the model [9]. The idea is that the pre-trained model will be an extractor of features.

For many physicians, the precise diagnosis of a cancerous tumor remains a challenging task. The advent of modern medical technology and large volumes of patient data paved the way for the development of new cancer prediction and cancer detection strategies. Although the data evaluation obtained from the admission of patients and

physicians contributes significantly to the diagnosis process, additional resources may be applied to promote accurate diagnosis.

Machine Learning (ML) techniques are aimed at removing potential diagnostic errors and providing a fast way to analyze large sections of the data. ML is an artificial information subfield of AI allowing learning without explicit programming, to learn a specific task by displaying it with datasets providing them to learn through experience. In recent decades ML approaches in designing predictive models to promote efficient decision-making have become widespread which make early detection and cancer prognosis possible. In research into cancer, these techniques may be used to classify various patterns in a sample, deciding whether cancer is malignant or benign [10]. The performance of such techniques can be evaluated according to classification, recall, specificity, and accuracy of the area under ROC. In this study, a DL model is projected for the automatic diagnosis of breast cancer. The studied model scenario has an end-to-end architecture without any feature extraction methods and requires raw chest X-ray images to restore the diagnosis [6]. This model is trained with 21400 chest X-ray images, which are not in a regular form taken from USF mammography datasets [11] containing 61% cancerous and 39% percent normal images.

Consequently, the proposed model is therefore intended to be used effectively. Radiologists will not have to wait a long time to determine whether or not a woman has breast cancer. The recovery cycle of patients, which can significantly minimize deaths, will also be a priority.

2 Literature Review

Many ML algorithms are used to get the same solution for breast cancer though most of them have drawbacks and disadvantages because of the CAD system or lack of datasets. It is believed that an algorithm should not only adjust the input instead it should regulate it with the best performance. The input must be given in the best way possible as well. For measurement of tumor analysis data, clinical results are used. The knowledge for this problem solution has been carefully analyzed and studied. It is extremely useful for data classification of patients. In [12], the authors suggested that Wilkins' scale-space techniques would use the original image to create a rougher resolution picture along with a Gaussian center. The method has a major drawback: the locations of the "semantically important" borders on ground scales are difficult to get accurately, and a class of algorithms is introduced that uses a diffusion technique. The algorithm consists of simple local operations on the image, which require parallel hardware applications. While [13], a new two-dimensional method for the medical ultrasound image was developed. Compared to typical images of the basic and second harmonic, the spatial resolution of the devolved image is much higher. Instead of rising harmonic base data, broadband radio frequency image data are decoded. The method was smart rather than the simple harmonic process. The operation has been verified by a fictional vision of captured image data and clinical image data. ML approaches in the past few decades to promote positive decision-making were becoming common in the development of predictive models. In [14], this technique can be used to classify different models in a dataset and thereby determine whether the cancer is malignant or

benign. Such techniques can be evaluated based on the particular classification, reminder, accuracy, and the Receiver Operating Characteristics (ROC) region. [15] studied a method based on CNN for the calculation of breast density. The CNN was educated in image features from the picture patches collected from all mammograms and graded as thick, fatty tissues. CNN was trained using local and worldwide statistical tools. The deep neural network (DNN) algorithm for classification of the breast densities in DMs was studied in [16]. The study included 20,000 screening mammograms labeled with a breast density of four levels. The Multiview data was used to discern breast densities by a scratch-based CNN with thick convolutional layers.

According to [17] the precise elimination of boundaries has been an important preprocessing step for the computerized analysis of the chest ultrasound. First, from an ultrasonic image, a rectangular region of interest is selected manually, followed by preprocessing for a reduction in noise and region-based image enhancement. The initial tumor edge is then achieved by using a transform wavelet. In the segmented regions, the authors of [18] suggested a technique of unattended DL for classifying the breast density and risk. The system uses a simple sparse autoencoder to learn the features. Three rating codes have been used for mammographic density scores: pm, fatty breast, and dense breast tissues. Two groups were considered for the mammographic texture score. This ranking was used as a criterion for the section of the breast tissue. The dike mark showed that the divided against the ground truth is fine. Three separate datasets were trained and checked, and the results showed a positive relation to the values that experts manually collected. In order to assist the risk scoring radiologists, the researchers of [19] proposed a CNN based density estimation method. The visual analog ranking of unrevealed images is calculated by CNN. In comparison with 2 independent readers in a clinical environment, the approach showed a good correlation with similar indices. The authors in [20] research a variety of objective methods of assessing image consistency, which was originally proposed for natural images and videos for medical images. Their key approach is to discuss the main questions arising from the acquisition processes of various imagery approaches, for objective quality assessment of still X-ray images. The majority of automated systems, therefore, concentrated on characterizing the epithelial regions of the brain for cancer detection.

3 Materials and Methods

ML models are historically trained in performing useful tasks on the basis of manually designed functions derived from the raw data or from features learned from other models. Through deep learning, computers automatically learn helpful representations and features from the raw data, bypassing this challenging manual. In comparison with traditional ML and AI, DL algorithms have dramatically improved performance [17, 21]. DL has become very common in the medical imaging community in order to detect or diagnose conditions such as lung cancer, brain tumor detections, and segmentation, and is commonly used in various fields, such as image recognition, natural language therapy etc. In ML the most important step is classification in which the decision-making process is done to decide whether an X-ray image is normal or cancerous. I.e. benign, or malignant. On the other hand, DL is an ML-implementation

technique that is particularly useful in order to learn patterns from unstructured datasets. DL enables multi-layered models of data representation to be learned at different levels of abstraction. These techniques have greatly developed emerging technologies in a range of other fields including speech recognition, visual object recognition, object detection, and drug discovery and genomics. It thrives in every area where DL is implemented. CNN's, one of the deepest learning systems, is one of the most related problems for image recognition [22]. Regardless of their ability to interpret images and segmentation, CNNs attract focus. A CNN is a DL-type consisting of many hidden layers, including the overlay, the pooling layer ReLU, and the fully linked layer. CNN shares weights in the convolutional layer, which reduces the memory base and improves network capacity. A modern CNN is usually built by stacking the convolutional layer on top of the input and one or more layers that are completely connected to the classification output in order to link them [23]. Most of the time, maximum pooling layers between convolutional layers are used to improve alternating invariance and reduce map dimensions. In this analysis, the VGG network and Inception-v4 are used as two common CNN structures (see Fig. 1). Furthermore, for each task, we train the model separately. We use the cross-entropic loss function to train the model. Adam's technique is used to obtain optimized model parameters values. Entire implementation was done in Python language.

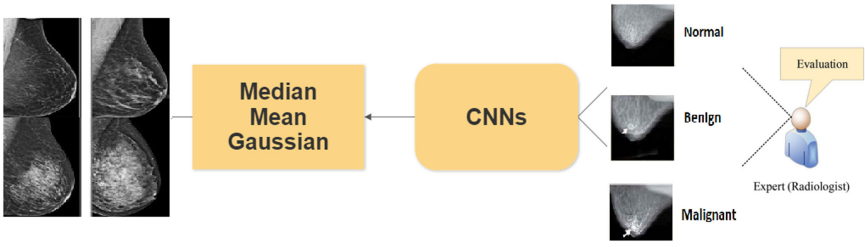

Fig. 1. The proposed method of architecture is presented.

In Fig. 1 as seen, the X-ray images are taken into pre-processed step to cope with the imbalanced data aftermath it is CNNs techniques are applied to process the data, and the output is tuned seamlessly. No area of interest for suspect lesions in the images has been given for the system. In the final step, the data is categorized into three classes as normal, benign, or malignant. Finally, it is ready for radiologists to evaluate.

3.1 Image Dataset

This work is focused on the Digital Database for Mammography Screenings (DDSM) which is a resource for the development of computer algorithms for screen-based use of mammographic image analysis research [11]. Each study contains two images per breast (age at study, ACR abnormality rate of breast density, and image (scanner, spatial resolution)). Moreover, the dataset includes normal, binomial, Poisson exponential Gauss distribution, gamma, inverse Gauss distribution, half binomial, and half

Poisson distributions. Standard functions; in addition to the unit, logit, log, reverse, and force functions, those used by distribution are also included. Precise reference information on pixel-level positions and types of suspect areas applies to photos comprising suspicious areas. Besides, the software is available for viewing mammograms and accurate pictures and for measuring the output of automated image analysis algorithms. The tissue analysis is based on the difference between high and low gray concentrations. Relevant mammographic image parameters related to the tissue allow us to identify them in a normal or abnormal way. Figure 2 shows different types of breast images.

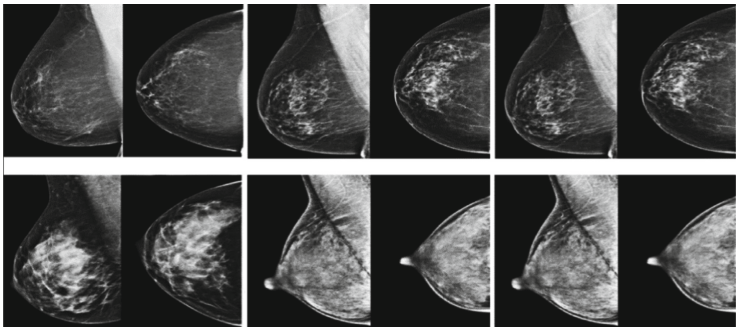

Fig. 2. Different X-ray images of breast dataset

3.2 Maximum Likelihood Estimation

One of the estimating the parameters of a statistical model is Maximum Likelihood Estimation (MLE), which is when applied to the X-ray image dataset provides better estimates for the model's parameters as MLE is benefitted for distinguishing between tumor and non-tumor cells [24]. MLE is the most consistent parameter estimate with sampled data and maximizes likelihood functions. MLE selects the set of values of model parameters that maximize the likelihood function, which has a unified approach; this has been described in the normal distribution and many other problems. Originally established by R.A. Fisher in the 1920s [25], MLE states that the desired possibility distribution is the one that makes the observed data "most likely." In other words, it is to investigate the value of the parameter vector, which maximizes probability functions. If a population has been known to follow a normal distribution, but the mean and variance are unknown, MLE can be used to approximate this with a small population sample by defining common mean and variance values, such that observation is the most likely outcome.

3.3 Pre-processing

Usually, data is inconsistent and also contains incomplete values, irrelevant values, diagrams, etc. Raw data must be parsed, collected, analyzed, and pre-processed before modeling and analysis. This is generally called a munging of data. Data that are often

imputed for missing data is a method used to fill or remove the missing values. Pre-processing data is a technique for data collection that converts raw information into a comprehensible format. Real-world data is often inconsistent, incomplete, and is possible to contain various errors.

The main aim of pre-processing is to enhance the image quality by eliminating or reducing unnecessary sections of the background of X-ray images to make it ready for further processing. Mammograms are complex, interpretable medical images. Preprocessing is, therefore, essential if the quality is to be improved. The mammogram will be prepared for the next phase. Noise and high-frequency components emitted by filters.

Mean Filtering

Mean filter is used to enhance image quality [26]. Here, each pixel was replaced by the filter with the average density value in the area. It reduces local variance and is easy to accomplish. The typical filter is used to distort an image to eliminate noise, as seen in Fig. 3b. This transforms the mean pixel values into an estimation of the kernel n x n. The center variable pixel intensity is then substituted by the average. It reduces any image noise and smoothes the image edges. What is benefitted from the mean filter is that the filter itself gives the average intensity value of an image used. Especially mammographic images containing microcalcification have a higher mean rather than normal ones. All in all, the mean filter is a simple spatial sliding window filter, which replaces the center in the window with the medium of all pixel values in the window. Typically, the window or kernel is square but can take any shape [27].

I) Averaging operations cause an image to blur and a position of the blur effect [28]. II) When the impulse noise distorted image is applied to an average operation, the impulse noise is decaying and dispersed, but not removed from the noise [29]. III) The mean value of all neighborhood pixels was significantly influenced by a single pixel with a non-representative value [30].

Median Filtering

The median filter is a conventional, nonlinear filter that is especially effective in removing noise from the impulse [31]. The pixel centered in a given window is replaced by the median of this window. As shown in Fig. 3(d), the application of a median filter on a highly polluted image eliminates spikes and thus significantly increases the ratio of signal to noise. The median filter is a nonlinear filter and is effective in removing salt and pepper noise. The median tends to maintain the sharpness of the image edges while removing noise. Median filter shall be applied in a moving window fashion to each part of a featured channel. An input image consisting of RGB channels, for example, corresponds to 3 feature channels; a set of features produced after the convolution typically contains several channel numbers. Then, the sequence median formed by all elements is found in that patch. Most of the median filter I) Center-weighted median filter [32]. II) weighted median filter III) Max-median filter increases the effect of window size on the median filtering noise, which is effectively removed [33].

Gaussian Filtering

A Gaussian filter, a linear filter is usually used to blur the image or to reduce noise [34]. Mostly it is used to blur edges and reduce contrast. Gaussian filters are used to correct

the weaknesses of X-ray images. That is, the Gaussian filter is applied to every X-ray image to reduce all the noise. Aftermath, X-ray images will have the same feature. When blurring affects X-ray images, it reduces the visualization and visibility of small components in the image [35]. Therefore, it is necessary to apply the Gaussian filter to restore the image from its corrupted version. Here It is used to isolate and quantify the vascular contents of the breast (see Fig. 3c).

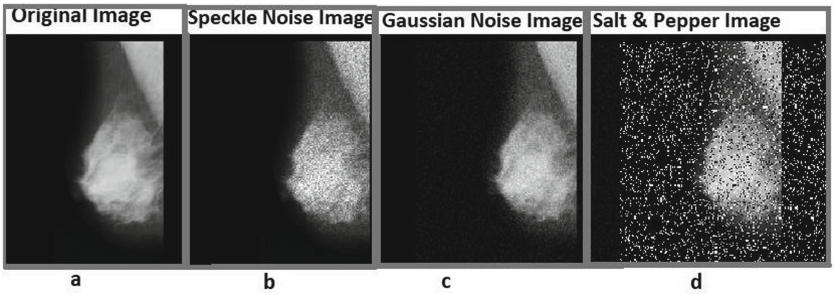

Fig. 3. X-ray images after pre-processing step

4 The Proposed Architecture

In the use of CNN architecture, the proposed DL framework is important. Two DL networks are involved in the system: Inception-v4 and VGG16, the novel suggested for diagnostic and predicational analysis of the breast cancer. DL is a family of hierarchical neural networks that aim to learn how to map the raw data to the desired outcome. The DL model's computer units are specified as layers and built into them to simulate a human brain deduction process. Convolution, bundling, active usage, and batch normalization are the principal machine formulas as described in the supplementary one.

From the 21400 X-ray images, we randomly selected 15000 images as training data and the rest 6400 images as testing data. Among the 15000 images for training data. Then the proposed data labeling method was applied to these 15000 data, and the majority of the unlabeled data were automatically labeled by the algorithm. The initially labeled data and newly labeled data were used to train the architecture.

4.1 Metrics Used in Health Check Systems for Evaluation

Different performance metrics are often used to investigate the performance of different models, such as specificity, accuracy, sensitivity (Table 1).

Table 1. Metrics used for evaluation

Metrics	Description	Formula
Sensitivity: It is the ability of a test to accurately.		$\frac{(TP)}{(TP+FN)}$
Specificity: It refers to how well a test identifies patients who have or do not have the disease.		$\frac{(TN)}{(TN+FP)}$
Accuracy: It is the relationship between the number of samples correctly identified and the number of samples.		$\frac{(TP+TN)}{(TP+FP+FN+TN)}$

4.2 VGG16

Known as its simplicity, VGG16 is one of the best networks; thus, its architecture is simple and deep. It primarily consists of an alternation between convolution layers and dropout layers [36]. VGG16 was the first to use a large number of small 3 × 3 filters in each convolutional layer and combine them in a sequence to emulate the effect of larger receiving areas.

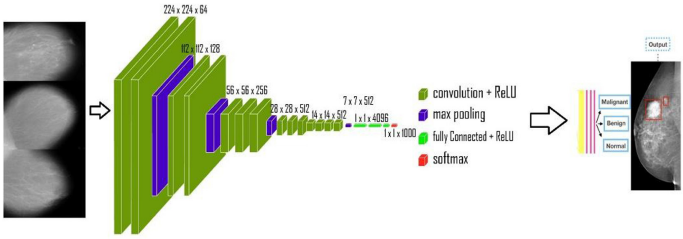

Fig. 4. VGG16 architecture used

However simple in network architecture, exponentially increasing cores lead to higher computation time and a larger size model, which is very expensive in terms of memory and computing costs. The applied VGG16 architecture consists of 13 convolutional layers, five pooling layers.

As seen in Fig. 4. X-ray images are passed through a stack of convolution layers with filters with a receiving area of 3 × 3 [37]. The convolution step is fixed at 1 pixel; The spatial filling of the convolutional layer input is done so that the spatial resolution

is maintained after the convolution. Spatial pooling is performed by five max-pooling layers that follow some convoluted layers. Max pooling is performed on a window from step 2 to 2 × 2 pixels.

4.3 Inception-v4

In computer vision, Inception is a striking deep neural network architecture [38]. The startup algorithm performs much better for built-in or mobile computing devices. The way to increase accuracy in deep CNN is to increase the level of work and the number of units in average sizes at each level. Initially, 1 × 1, 3 × 3, 5 × 5 convolution filter and 3 × 3 maximum affinity filter is used. The maximum bonding process in the convolution layer is very effective [30]. The initial is optimal sparse architecture and 3 ∼ 10 × faster than other architecture.

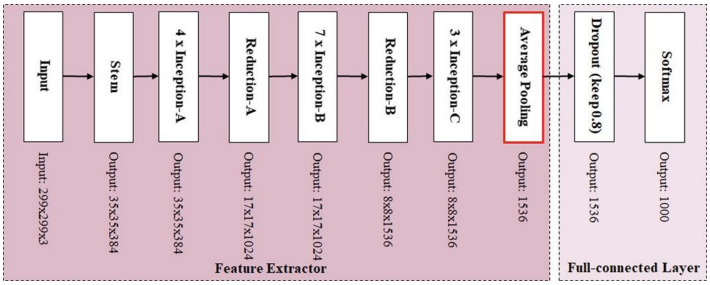

Fig. 5. Inception v4 model of the architecture [39]

From Fig. 5 it can be seen easily that Inception v4 consists of two parts a full layer and an extractor feature that has a lot of convolutional blocks such as 1 stem block, 4 Inception-A blocks, 7 Inception-B blocks, 3 Inception-C blocks, and 1 Average Pooling layer. 1 dropout block and 1 softmax layer are combined with a full-connected layer. Precisely, the stem module is used to exploit the Conv and Max-pool blocks to convert 299 × 299 × 3 image shapes into 35 × 35 × 384 image shapes, which is the Inception-A block input. Alternatively, Inception-A, Inception-B, Inception-C blocks only use Conv and Avg Pooling to convolute higher abstract features of images. Whereas the Inceptions of the same type have the same structure and are directly connected in sequence, concepts of a different type need to be connected by a grid reduction module [40]. For instance, the Reduction-A grid-reduction module, which converts a shape of 35 × 35 to a shape of 17 × 17, is used to connect the Inception-A block and Inception-B block. In addition, the Reduction-B grid-reduction module, which converts 17 × 17 shapes to 8 x 8 shapes, is used to connect the Inception-B block and Inception-C block [41]. The output of the Inception-C block is converted to the 1-Dimension of the 1536 property by the average pool layer [42].

5 Experimental Results and Discussion

This section explains the parameters and presents the results that helped the three classifications explored in this article. We performed experiments with X-ray images to identify and diagnose breast cancer in various separate scenarios. The proposed model was trained to classify X-ray images into three categories: benign, normal, malignant. The efficiency of the proposed model is displayed in Fig. 4 and 5 using VGG16 and Inception v4. The findings show that at the start of the training, there is not a sharp increase in loss values at the beginning of the training. The main reason for this slight increase and decrease is attributed to the number of X-ray data. However, when the proposed model analyzes all X-ray images each cycle of training, the quick ups and downs in the latter part of the training are gradually reduced.

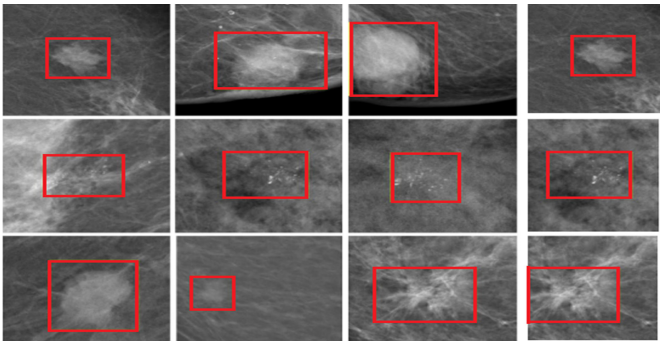

Fig. 6. X-ray images diagnosed as malignant

We fine-tune all the image networks on the training set directly without using ROI annotations and evaluate model performance using per-image validation the area under the curve (AUC) scores. The two best models are used for transfer learning. Figure 6 shows how the images are diagnosed according to their situation based on the model applied.

Table 2. Recall values found after CNN applied.

	VGG16	Inception v4
Benign	97.3	96.8
Malignant	96.34	95.7
Normal	97.5	96.4

5.1 Recall

Recall, known as sensitivity, is the rate of the positive observations accurately predicted to be positive. This measure is especially desirable because of how many observations are correctly diagnosed on the breast images dataset. Accurately identifying a malignant tumor is the main goal of the study. Recall can be thought of as the capacity of a model to identify all of the data points of interest in a dataset that reflects the proportion of real positives that have been correctly identified [43].

The recall values for all two techniques are shown in Table 2. In the model Inception v4 shows a higher result when compared to VGG16

5.2 Accuracy

Accuracy is characterized as a measurement similarity to the norm, or true value, i.e., a highly accurate navigation system may provide measurements that are very close to the norm, true or known values. A strong case for accuracy includes measurements of statistical variation and variance [44]. The accuracy is generally measured by the standard deviation of error. The accuracy of the classifier is a measure of how accurately the classifier can classify cases into their proper categories. It is obtained by dividing the right estimates by a total number of samples in the dataset. It is worth noting that the accuracy is largely dependent on the threshold chosen by the classifier and, therefore, can vary for different testing sets [45]. Therefore, comparing different classifiers is not the best approach, but can provide an overview of the class. As shown in Table 1, the equation can also be used for calculating accuracy where TP and TN reflect one-to-one True Positive and True Negative values [46]. P and N represent the positive and negative population, respectively, of cases of malignant and benign. Accuracy is the number of data points across all properly measured data points. In other words, the number of real positive and true negative is determined by the number of true positive, negative, and false negatives.

Confidence-based accuracy is characterized as a truly positive as well as a true negative ratio. It illustrates how well the architecture handles constructive insights but does not perceive them as negative anticipations. Figure 7 and Table 3 show the accuracy values and VGG16 has the highest performance [47].

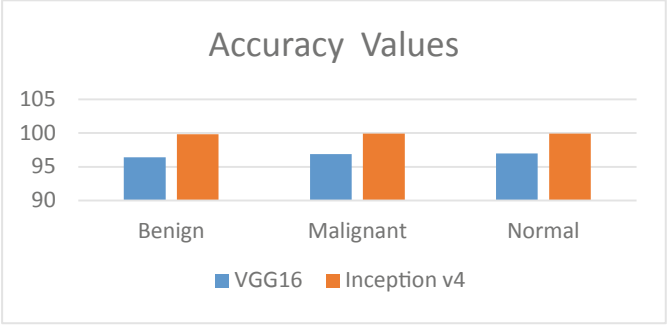

Fig. 7. Accuracy values found after CNN applied

The accuracy values for all two techniques are shown in Table 3. The performance of the different methods is measured by calculating the accuracy.

Table 3. Accuracy values found after CNN applied.

	VGG16	Inception v4
Benign	98.01	97.62
Malignant	95.11	95.8
Normal	97.2	96.6

5.3 ROC Area

ROC is a method commonly used in bioinformatics to test classifier performance [48]. There are four possible results in a dataset; The positive sample is counted as TP when classified correctly, FN when classified incorrectly, TN when classified correctly, and FP when classified incorrectly.

As Fig. 8 shows the efficiency of the classifier, it is a way to display the balance between the cost and the usefulness of a classifier. ROC is one of the most widely used and valuable data mining efficiency indicators. This 2D diagram displays the result of AUC.

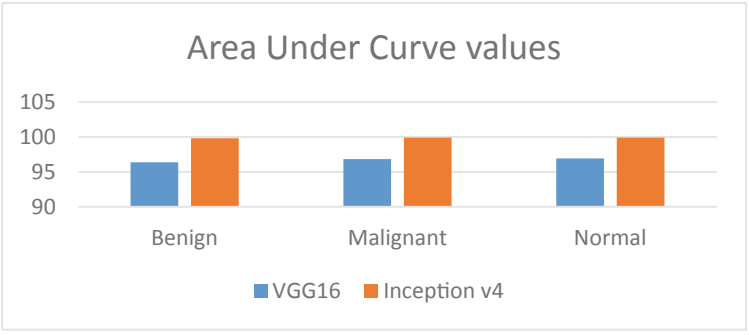

Fig. 8. The area under curve values found after CNN applied

In the upper left quarter of a plot with a high TP ratio and a low FP ratio, a point on the ROC is more preferred. A dot on the crossline y = x reflects a weak predictive classifier based on the random number of TP and FPs [49]. The area under the ROC graph reflects the classifier's performance as displayed in Fig. 8. This is achieved by dividing the area under the graph by the total area of the graph. Values close to 1 indicate the higher performance of the classifier. Percentage values for the ROC area of all two techniques.

6 Result

In this paper, DL architecture for early diagnosis of breast cancer by using a benchmark X-ray images dataset is proposed. Breast cancer is the most dangerous disease as a form of cancer among women. The progress of cancer detection and prediction is, therefore, critical for a healthy life. In this paper, we have discussed two popular ML techniques for Breast Cancer diagnosis. The proposed model shows better results when two methods are taken into consideration. DL is much better than the traditional classification approaches for image classification process and effectively reduced the FN rate with high accuracy, especially when using the VGG16 method. It finishes the test with 96.77% compared to Inception-v4 with and accuracy 96.67%, which is very close to the previous result. The results of the proposed model have a high accuracy of breast cancer images.

DL method to be used to diagnose breast cancer can efficiently and accurately calculate the 21400 collected X-ray images, which soon can be applied to laboratory X-ray images as well. The architecture can be used to screen a large number of suspected people's X-ray data sets to save people's lives and to save limited medical resources and can help the optimization of the diagnosis process. For future work the proposed architecture can constantly learn, adapt, and upgrade.

A more stimulating and a necessary approach for future research would be focusing on differentiating patients who show mild symptoms rather than severe ones, while these symptoms may or may not be displayed correctly on X-ray images.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Waheed, K.B., et al.: Breast cancers missed during screening in a tertiary-care hospital mammography facility. Ann. Saudi Med. (2019). http://doi.org/10.5144/0256-4947.2019.236
2. Mello-Thoms, C.R., Gandomkar, Z., Brennan, P.C.: A cognitive approach to determine the benefits of pairing radiologists in mammogram reading (2018)
3. Arevalo, J., González, F.A., Ramos-Pollán, R., Oliveira, J.L., Guevara Lopez, M.A.: Representation learning for mammography mass lesion classification with convolutional neural networks. Comput. Methods Programs Biomed. (2016)
4. Litjens, G., et al.: A survey on deep learning in medical image analysis. Med. Image Anal. (2017). http://doi.org/10.1016/j.media.2017.07.005
5. Routray, I., Rath, N.P.: Textural feature based classification of mammogram images using ANN. In: 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT 2018) (2018)
6. Samant, N., Sonar, P.: Mammogram classification in transform domain. In: 2018 5th International Conference on Signal Processing and Integrated Networks (2018)
7. Keren, L., et al.: Paired cycle-GAN-based image correction for quantitative cone-beam computed tomography. Med. Phys. **7**(1), 654–663 (2018)
8. Cifci, M.A.: Deep learning model for diagnosis of corona virus disease from CT images. Int. J. Sci. Eng. Res. **11**(4) (2020)
9. Chougrad, H., Zouaki, H., Alheyane, O.: Deep convolutional neural networks for breast cancer screening. Comput. Methods Programs Biomed. (2018)

10. Campanella, G., et al.: Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nat. Med. (2019). http://doi.org/10.1038/s41591-019-0508
11. Cogan, T., Cogan, M., Tamil, L.: RAMS: remote and automatic mammogram screening. Comput. Biol. Med. (2019). http://doi.org/10.1016/j.compbiomed.2019.01.024
12. Robertson, S., Azizpour, H., Smith, K., Hartman, J.: Digital image analysis in breast pathology—from image processing techniques to artificial intelligence. Transl. Res. (2018). http://doi.org/10.1016/j.trsl.2017.10.010
13. Wan, T., Cao, J., Chen, J., Qin, Z.: Automated grading of breast cancer histopathology using cascaded ensemble with combination of multi-level image features. Neurocomputing (2017). http://doi.org/10.1016/j.neucom.2016.05.084
14. Yue, W., Wang, Z., Chen, H., Payne, A., Liu, X.: Machine learning with applications in breast cancer diagnosis and prognosis. Designs **2**(2), 13 (2018)
15. Bejnordi, B.E., et al.: Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA - J. Am. Med. **318**(22), 2199–2210 (2017)
16. Mambou, S.J., Maresova, P., Krejcar, O., Selamat, A., Kuca, K.: Breast cancer detection using infrared thermal imaging and a deep learning model. Sensors **18**(9), 2799 (2018)
17. Al Nahid, A., Mehrabi, M.A., Kong, Y.: Histopathological breast cancer image classification by deep neural network techniques guided by local clustering. Biomed Res. Int. (2018). https://doi.org/10.1155/2018/2362108
18. Tammina, S., et al.: Automated classification of lung diseases in computed tomography images using a wavelet based convolutional neural network. J. Biomed. Sci. Eng. **11**(10), 263–274 (2019). https://doi.org/10.4236/jbise.2018.1110022
19. Gao, F., et al.: SD-CNN: a shallow-deep CNN for improved breast cancer diagnosis. Comput. Med. Imaging Graph. (2018). http://doi.org/10.1016/j.compmedimag.2018.09.004
20. Radiological Society of North America: Ultrasound – Breast. Radiology Info (2016)
21. Turkki, R., Linder, N., Kovanen, P.E., Pellinen, T., Lundin, J.: Antibody-supervised deep learning for quantification of tumor-infiltrating immune cells in hematoxylin and eosin stained breast cancer samples. J. Pathol. Inform. (2016)
22. Al Nahid, A., Kong, Y.: Involvement of machine learning for breast cancer image classification: a survey. Comput. Math. Methods Med. (2017)
23. Zhu, C., et al.: Deep learning system to screen coronavirus disease 2019 pneumonia. Appl. Intell. **9**(September), 330 (2020)
24. Nandagopal, V., Geeitha, S., Kumar, K.V., Anbarasi, J.: Feasible analysis of gene expression – a computational based classification for breast cancer. Meas. J. Int. Meas. Confed. **140**, 120–125 (2019). https://doi.org/10.1016/j.measurement.2019.03.015
25. Zou, S., et al.: Predicting lung nodule malignancies by combining deep convolutional neural network and handcrafted features. Sci. Rep. **2**(1), 374–382 (2020). https://doi.org/10.1088/1361-6560/ab326a
26. Becker, A.S., Mueller, M., Stoffel, E., Marcon, M., Ghafoor, S., Boss, A.: Classification of breast cancer in ultrasound imaging using a generic deep learning analysis software: a pilot study. Br. J. Radiol. **91**(xxxx), 20170576 (2018)
27. Anderson, R., Li, H., Ji, Y., Liu, P., Giger, M.L.: Evaluating deep learning techniques for dynamic contrast-enhanced MRI in the diagnosis of breast cancer. In: Medical Imaging 2019: Computer-Aided Diagnosis, vol. 10950, p. 1095006. International Society for Optics and Photonics, March 2019
28. Saha, M., Chakraborty, C., Racoceanu, D.: Efficient deep learning model for mitosis detection using breast histopathology images. Comput. Med. Imaging Graph. **64**, 29–40 (2018)

29. Mambou, S.J., Maresova, P., Krejcar, O., Selamat, A., Kuca, K.: Breast cancer detection using infrared thermal imaging and a deep learning model. Sensors **18**(9), 27 (2018)
30. Golatkar, A., Anand, D., Sethi, A.: Classification of breast cancer histology using deep learning. In: Campilho, A., Karray, F., ter Haar Romeny, B. (eds.) ICIAR 2018. LNCS, vol. 10882, pp. 837–844. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-93000-8_95
31. Delen, D., Walker, G., Kadam, A.: Predicting breast cancer survivability: a comparison of three data mining methods. Artif. Intell. Med. **34**(2), 113–127 (2005)
32. Zhang, D., Zou, L., Zhou, X., He, F.: Integrating feature selection and feature extraction methods with deep learning to predict clinical outcome of breast cancer. IEEE Access **6**, 28936–28944 (2018)
33. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. arXiv: 1512.03385 [cs] (2015)
34. Das, A., Acharya, U.R., Panda, S.S., Sabut, S.: Deep learning based liver cancer detection using watershed transform and Gaussian mixture model techniques. Cogn. Syst. Res. **54**, 165–175 (2019)
35. Choukroun, Y., et al.: Mammogram classification and abnormality detection from nonlocal labels using deep multiple instance neural network. In: Eurographics Workshop on Visual Computing for Biology and Medicine (2017)
36. Lee, G.F., Hoogi, A., Rubin, D.: Curated breast imaging subset of DDSM. Cancer Imaging Arch (2016)
37. Torres-Galván, J.C., Guevara, E., González, F.J.: Comparison of deep learning architectures for pre-screening of breast cancer thermograms. In: 2019 Photonics North (PN), pp. 1–2. IEEE, May 2019
38. Broadwater, D.R., Smith, N.E.: A fine-tuned inception v3 constitutional neural network (CNN) architecture accurately distinguishes between benign and malignant breast histology, no. 18133, 59 MDW San Antonio United States (2018)
39. Pham, T.-C., Luong, C.-M., Visani, M., Hoang, V.-D.: Deep CNN and data augmentation for skin lesion classification. In: Nguyen, N.T., Hoang, D.H., Hong, T.-P., Pham, H., Trawiński, B. (eds.) ACIIDS 2018. LNCS (LNAI), vol. 10752, pp. 573–582. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75420-8_54
40. Deng, J., Ma, Y., Deng-ao, L., Zhao, J., Liu, Y., Zhang, H.: Classification of breast density categories based on SE-attention neural networks. Comput. Methods Programs Biomed., 105489 (2020)
41. Wang, Y., Choi, E.J., Choi, Y., Zhang, H., Jin, G.Y., Ko, S.B.: Breast cancer classification in automated breast ultrasound using multiview convolutional neural network with transfer learning. Ultrasound Med. Biol. (2020)
42. Coccia, M.: Deep learning technology for detection of lung and breast cancer: application in clinical practice (2019)
43. Clancy, K., Zhang, L., Mohamed, A., Aboutalib, S., Berg, W., Wu, S.: Deep learning for identifying breast cancer malignancy and false recalls: a robustness study on training strategy. In: Medical Imaging 2019: Computer-Aided Diagnosis, vol. 10950, p. 1095005. International Society for Optics and Photonics, March 2019
44. Thrall, J.H., et al.: Artificial intelligence and machine learning in radiology: opportunities, challenges, pitfalls, and criteria for success. J. Am. Coll. Radiol. **15**(3), 504–508 (2018)
45. Samala, R.K., Chan, H.P., Hadjiiski, L., Helvie, M.A., Richter, C.D., Cha, K.H.: Breast cancer diagnosis in digital breast tomosynthesis: effects of training sample size on multi-stage transfer learning using deep neural nets. IEEE Trans. Med. Imaging **38**(3), 686–696 (2018)
46. Abdel-Zaher, A.M., Eldeib, A.M.: Breast cancer classification using deep belief networks. Expert Syst. Appl. **46**, 139–144 (2016)

47. Qi, X., et al.: Automated diagnosis of breast ultrasonography images using deep neural networks. Med. Image Anal. **52**, 185–198 (2019)
48. Yuan, X., Xie, L., Abouelenien, M.: A regularized ensemble framework of deep learning for cancer detection from multi-class, imbalanced training data (2018)
49. Wang, J., Yang, Y.: A context-sensitive deep learning approach for microcalcification detection in mammograms. Pattern Recogn. **78**, 12–22 (2018)