Breast Cancer Diagnosis in Women Using Neural Networks and Deep Learning

Ojo Abayomi Fagbuaun¹, Olaya Folorunsho¹²*, Lawrence Bunmi Adewole¹ & Titilope Helen Akin-Olayemi³

¹Department of Computer Science, Faculty of Sciences, Federal University Oye Ekiti, Km 3, Oye-Afao Road, Oye Ekiti, 371104, Nigeria
²Unit for Data Science and Computing, School of Computer Science and Information Systems, North-West University, 11 Hoffman Street, Potchefstroom 2531, South Africa
³Department of Computer Science, Federal Polytechnic, Ado-Ikare Road, Ado-Ekiti, 360231, Nigeria
*E-mail: olaiya.folorunsho@fuoye.edu.ng

Abstract. Breast cancer is a deadly disease affecting women around the world. It can spread rapidly into other parts of the body, causing untimely death when undetected due to rapid growth and division of cells in the breast. Early diagnosis of this disease tends to increase the survival rate of women suffering from the disease. The use of technology to detect breast cancer in women has been explored over the years. A major drawback of most research in this area is low accuracy in the detection rate of breast cancer in women. This is partly due to the availability of few data sets to train classifiers and the lack of efficient algorithms that achieve optimal results. This research aimed to develop a model that uses a machine learning approach (convolution neural network) to detect breast cancer in women with significantly high accuracy. In this paper, a model was developed using 569 mammograms of various breasts diagnosed with benign and malignant cancers. The model achieved an accuracy of 98.25% and sensitivity of 99.5% after 80 iterations.

Keywords: breast-cancer; diagnosis; deep learning; mammography; neural network.

1 Introduction

As breakthroughs in technology continue to emerge in science and technology, their application in problem solving in the field of healthcare has also increased, despite the widespread fear of increased reliance on technology. Models are continuously being developed to imitate problem solving skills and decision making of human experts. Machine learning is a field of application that is continuously being leveraged to improve diagnostic processes in the healthcare system. Clinical decision support systems (CDSS) are a booming area where machine learning has found useful application. The huge number of deaths from malignant cancer resulting from poor or unavailable early detection of breast cancer has drawn the attention of scientists to the development of models that can
be used to improve upon the current diagnostic process, thereby identifying infections and promptly responding to the exigency of treatment of the disease. Furthermore, early detection is helpful in providing immediate medication and significantly improves patient care.

Uncontrolled growth and division of cells of the breast is an early signal of breast cancer. The disease is very prevalent among women, even though it is also common among men. In Sung, et al. [1], among all the various types of cancer, female breast cancer ranked highest with 2,261,419 new cases detected in the year 2020 alone. This accounts for about 11.7% of cancer cases worldwide. The new deaths that resulted from female breast cancer was put at 684,996, which accounts for about 6.9% of all deaths in the same year. This was a significant increase from the year 2018, when an estimated 2.1 million women were diagnosed with breast cancer and 626,679 died of the disease in the same year. As seen in Moody, et al. [2], early research on breast cancer proved that metastatic breast cancer is the major cause of death in women. Breast cancer is diagnosed in the healthcare system majorly by classifying tumors, which can be done in four ways: self-breast examination, near infrared fluorescence, mammography, and biopsy.

However, the most generally accepted ways are through mammography and biopsy (Smith, et al. [3] and Siegel, et al. [4]). Breast tumors are classified as benign or malignant. A benign lesion is classified as non-cancerous, while a malignant one is cancerous. The complexity of biopsy has made mammography to be the most accepted diagnosing approach (Chekkoury, et al. [5] and Carlson, et al. [6]). According to the World Health Organization (WHO) [7], emphasis should be on early diagnosis and treatment of breast cancer, because it can increase the survival rate of victims by over 80%.

In order to ensure early breast cancer diagnosis, radiologists have proposed various fast detection techniques. Recent advancements in artificial intelligence (AI) and machine learning (ML) have seen various computer scientists contribute by developing machine and deep learning techniques. Nevertheless, the issues of low accuracy caused by insufficient data, defective data pre-processing, and a lack of efficient algorithms still linger. The present research aimed to mitigate these issues by developing a model capable of detecting breast cancer with higher accuracy. Research has shown that early detection of breast cancer increases the survival rate of breast cancer patients, as quick treatment prevents it from becoming metastatic. Moreover, the availability and prompt accessibility of treatment after early detection of breast cancer makes it cheaper to treat cancer. This is important to reduce the mortality rate (Vargas, et al. [8]). An illustration of the hypothetical theory of breast cancer formation is presented in Figure 1.
Various deep learning methods have been used with great success in image analysis and recognition. Specific areas of application include analysis of chest radiographs for detection of pneumonia, and detection of tuberculosis in chest X-rays. In Nadeem, et al. [10], a comprehensive review, taxonomy, and future challenges of brain tumor analysis was carried out by using deep learning. Several deep learning techniques were reviewed by Ari, et al. [11], a deep convolutional neural network (DCNN) is presented in Liu, et al. [12], and a fully convolutional neural network (FCNN) is presented in Zhao et al. [13]. Even though these techniques achieved excellent results, a major drawback is that their performance is hindered by the dimension and size of the images along with a lack of available training data. Nevertheless, the learning ability of CNNs has produced excellent results when data for training is abundantly available as presented in Nie, et al. [14], Farrukh, et al. [15] and Amarapur, et al. [16]. In Khan, et al. [17], an intelligent model for identification of brain tumors using a deep learning approach is presented. A hierarchical deep learning approach was used in the classification of brain tumors into pituitary, glioma, and meningioma. The system achieved 92.12% accuracy, which shows that deep learning techniques can provide excellent performance in image classification and recognition. In
Siddiqui, et al. [18], intelligent prediction of the stages of breast cancer using deep learning and the Internet of Medical Things is proposed. The result of the experiment showed accuracy of 99.69% for detection of ductal carcinoma, 99.32% for lobular carcinoma, 98.96% for mucinous carcinoma, and 99.32 for papillary carcinoma breast cancer.

Going by the performance of deep learning techniques, the present research was aimed at solving the recurrent issue of low accuracy reported in early detection of breast cancer using mammograms, caused by various factors, such as the type of data set, the pre-processing data methods, the deep learning model implementation method, and the volume of data set used. This will reduce the mortality rate caused by the metastatic nature of breast cancer.

2 Related Works

The breast, a tissue that lies over the chest muscles, is composed of specialized glandular tissues with dense fibrous stroma, whose purpose is to produce milk. Multiple ducts in the breast connect milk-producing lobes to the nipple of the breast, and this secretion serves as a major food source for infants, as discussed in Ref. [19].

2.1 Mammography

Mammography entails using a dedicated medical imaging instrument that uses a low energy X-ray system to create an image of the interior part of human breasts, commonly known as a mammogram, that can be used to verify the presence or absence of benign or malignant cancers, as discussed by Rani [20]. Mammography examination, usually carried out by a radiologist, helps in the early diagnosis of breast diseases present in women. An illustrative instance of a mammogram is shown in Figure 2.

![Figure 2](image-url)
There are three major ways of carrying out a medical examination of human breasts using a mammogram:

1. **Film-screen mammography.** The captured X-ray images of the breasts are presented in black and white directly on a sheet of photographic film.

2. **Digital mammography.** The captured X-ray image of the breasts is directly recorded into a computer system. The X-ray image can be reviewed by radiological experts and then stored for future use, see Rampun et al. [21].

3. **3-D tomosynthesis mammography.** The physician can use this to distinguish between masses/tissues that may be cancerous and those which are not. It allows a proper and detailed view of the breasts in layers, thereby allowing radiologists a more accurate interpretation of the mammogram. It has been discovered that false-positive detection cases have been reduced drastically with the help of 3D mammography. It also ensures better accuracy in the early detection of breast cancer.

The importance of breastfeeding to infants cannot be overemphasized. However, various infections can affect the breast, the most common one being breast cancer. Breast cancer is a major health challenge, affecting women of all ages and races. As the most common disease affecting women across the globe, more than 570,000 deaths have been recorded from the disease since the year 2015. As reported by Stewart, et al. [22], more than 1.5 million women are diagnosed with cancer yearly across the globe. Due to its metastatic ability, breast cancer can migrate to other organs in the body, including the liver, the brain, the lungs, and any other part of the human body. This ease of spread makes it difficult for health workers to cure the disease. Despite the annual increase in breast cancer incidence across the globe, the mortality rate has reduced because of early screening to achieve early diagnosis. These and many other issues have prompted researchers to develop a scientific solution to mitigate the prevalence of breast cancer disease.

In Zuluaga-Gomez, et al. [23], a convolutional neural network (CNN) inspired method to diagnose breast cancer using thermal images with an optimization algorithm is proposed. The result shows that their system had an accuracy of 92%. However, the system lacked a sufficiently large data set, as only 57 instances of breast cancer were used. In Khan, et al. [24], a novel deep learning-based framework for breast cancer detection is proposed that uses transfer learning. Feature extraction was done with pre-trained GoogLeNet, Visual Geometry Group Network (VGGNet), and Residual Networks (ResNet). The images are input into a fully connected layer to classify both malignant and benign cancer cells, using average pooling classification.

The model recorded an accuracy of 97.52%. However, the data pre-processing method is questionable, as only 82 images were augmented to provide 8,000 images in total. In Mambou, et al. [25], infrared thermal imaging with a deep
learning method for the detection of breast cancer is proposed. The proposed model was trained using 67 images obtained through the Research Data Base (DMR). The DMR contained thermogram images obtained from the frontal region. The images were acquired with an FLIR SC-620 IR camera with a resolution of 640 x 480 picture elements (pixels). The model was tested using 12 breast cancer images, with each image augmented to generate 20 images each. The model achieved an accuracy of 88.5%.

In Asri, et al. [26], a performance comparison between different classification algorithms was carried out on the Wisconsin Breast Cancer data set. In Naji et al. [27], several machine learning algorithms were implemented, i.e., KNN, SVM, logistic regression, random forest, and decision tree. Data cleaning was carried out and features were extracted from the data. The data set was divided into 75% for training and 25% for testing. Out of all the algorithms implemented in the research, SVM reported the highest accuracy at 92.2%.

In Schaefer, et al. [28], a classification algorithm based on a fuzzy inferencing system was tested on 150 instances of breast cancer. The result showed an accuracy of 80%. A limitation of this research was the insufficient number of data used. In Sumathi, et al. [29], Wisconsin Breast Cancer Data (WBCD) was used as input into a genetic algorithm plus an adaptive resonance motivated neural network to diagnose breast cancer. In the experiment, 699 samples were trained and taken through the biopsy technique, with 16 missing data. A total of 683 samples with breast tumors were used in the research. 65% of the data showed evidence of benign tumors, while 35% showed evidence of malignant tumors. The researchers developed their model by combining probabilistic neural network (PNN) with multilayer perceptron (MLP). This research used a data set obtained through the biopsy technique, which is always cumbersome during training.

A hybridized computerized breast cancer detection model using a genetic algorithm and back propagation neural network (BPNN) is presented in Adam and Omar [30] as a faster model, reducing the overall detection time and increasing the accuracy of detection. Two data cleaning processes were implemented on the data set, and the model achieved an overall detection accuracy of 83.36%. Deep learning was introduced to improve the available machine learning techniques to extract relevant features from input images for an efficient classification process in LeCun, et al. [31]. A typical example of deep learning is the convolutional neural network (CNN), as those proposed by Lee, et al. [32], Li, et al. [33], and Liu, et al. [34]. Cloud-based breast cancer prediction empowered with soft-computing approaches is proposed by Farrukh, et al. in [15]. In their proposition, type-1 fuzzy logic breast cancer prediction (BCP-T1F) and support vector machine-based breast cancer prediction (BCP-SVM) were
implemented, defining the stage as well as the type of breast cancer a person is suffering from.

The total number of samples used in the experiment was 399. The accuracy of the proposed BCP-T1F-SVM after testing gave 98.33% sensitivity, 96.36% specificity, 97.06% accuracy, and a 2.94% miss rate. With more data used in the training stage, there is a very good probability that the accuracy of the model will be increased. A sample of the data set used for the model proposed in this research is presented in Figure 3. The data set consisted of 569 mammograms of various breasts diagnosed with benign and malign cancers.

(a)
(b)

Figure 3 Extract image from breast data set.

3 Data Normalization and Preprocessing

The size of images in the data set used was 1064 x 1064 pixels, which and this was rescaled to 256 x 256 before being fed to the neural network to ensure faster processing. The data set was converted to an array using the CV2 command in Python, and data cleaning was carried out using a special data science function. The Python 3.7 programming language and Jupyter editor via the Anaconda Distribution were used in implementing the proposed system. Google Colab and the TensorFlow Framework were used to develop the model. This research proposes the adoption of convolutional neural network (CNN) to achieve early breast cancer detection. The proposed architecture is presented in Figure 4.

---

Figure 4 Proposed model architecture.
CNN is a family of neural networks that produce excellent performance compared to other neural network models in image processing, speech processing, and audio processing applications. Neural networks are graphs with weights attached to each edge. It consists of layered sets where a layer consists of a set of nodes. It consists of an input layer and an output layer. The hidden layer lies between these two layers and varies from one type of network to another. When an input signal is applied, each node produces outputs that act as input to the hidden layers. A given node's output is derived by applying an activation function ($\psi$) to the inputs of the set of nodes that belong to the previous layers. The above process is called inference. The algorithm is described as follows:

$L$: total number of layers contained in the neural network
$l^{(k)}$: $k^{th}$ layer
$m^{(k)}$: nodes contained in $l^{(k)}$
$l^{(k)}_i$: node $i$ in layer $l^{(k)}$
$O^{(k)}$: the output vector representing the outputs of the nodes in $l^{(k)}$
$O^{(k)}_i$: layer $l^{(k)}$ output
$w^{(k)}$: matrix's weight that connects nodes in layer $l^{(k-1)}$ to nodes in layer $l^{(k)}$
$w^{(k)}_{ij}$: weight connecting nodes $l^{(k-1)}_i$ and $l^{(k)}_j$
$b^{(k)}$: bias for $l^{(k)}$
$\psi^{(k)}$: function to determine the outputs $O^{(k)}$ and $O^{(k-1)}$
$\sigma$: activation function
$X$: every input data of the data set
$Y$: every provided output of the data set
$\hat{Y}$: approximation of all outputs given all inputs
$x_n$: specific $n^{th}$ data input
$y_n$: specific $n^{th}$ output layer
$\hat{y}_n$: approximation of output $y_n$ given input $x_n$

4 Result and Discussion

The model developed in this research contained 8 interconnected convolution layers. The data set used contained 569 mammogram images in total. The data set was divided at a ratio of 90:10, representing the training set and the testing set, respectively. Thus, 512 images were used to train the model, while 57 images were used to test the system.

Table 1 shows an extract of our model during training. The system achieved an accuracy of 98.25% after 80 iterations. An extract of the system during iteration is shown in Figure 5.
<table>
<thead>
<tr>
<th>Layer (type)</th>
<th>Output Shape</th>
<th>Param #</th>
</tr>
</thead>
<tbody>
<tr>
<td>conv1d_1 (Conv1D)</td>
<td>(None, 16)</td>
<td>12</td>
</tr>
<tr>
<td>batch_normalization_1 (Batch)</td>
<td>(None, 16)</td>
<td>0</td>
</tr>
<tr>
<td>dropout_1 (Dropout)</td>
<td>(None, 16)</td>
<td>0</td>
</tr>
<tr>
<td>conv1d_2 (Conv1D)</td>
<td>(None, 32)</td>
<td>1952</td>
</tr>
<tr>
<td>batch_normalization_2 (Batch)</td>
<td>(None, 32)</td>
<td>12</td>
</tr>
<tr>
<td>dropout_2 (Dropout)</td>
<td>(None, 32)</td>
<td>0</td>
</tr>
<tr>
<td>Flatten_1 (Flatten)</td>
<td>(None, 32)</td>
<td>0</td>
</tr>
<tr>
<td>dense_1 (Dense)</td>
<td>(None, 1)</td>
<td>28704</td>
</tr>
<tr>
<td>dropout_3 (Dropout)</td>
<td>(None, 1)</td>
<td>0</td>
</tr>
<tr>
<td>dense_2 (Dense)</td>
<td>(None, 1)</td>
<td>55</td>
</tr>
</tbody>
</table>

Total params: 50,588
Trainable params: 23,587
Non-trainable params: 27

Figure 5 Iteration extract of model training.
The accuracy of the developed model is presented in Figure 6. The test loss of the developed model is presented in Figure 7.

The confusion matrix is displayed in Table 2, with 497 true positives, i.e., predictions that are true while confirmation also says they are true (malignant tumors detected as malignant); 6 true negatives, i.e., predictions that are not true while confirmation also says they are not true (benign tumors correctly identified as benign); 7 false positives, i.e., predictions that are true while confirmation says they are not true (malignant tumors detected as benign); 2 false negatives, i.e., predictions that are not true while confirmation says they are true (benign tumors classified as malignant).

<table>
<thead>
<tr>
<th>Actual: normal</th>
<th>Actual: cancerous</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted: normal</td>
<td>FN = 2</td>
<td>TN = 6</td>
</tr>
<tr>
<td>Predicted: cancerous</td>
<td>TP = 497</td>
<td>FP = 7</td>
</tr>
<tr>
<td>Total</td>
<td>499</td>
<td>13</td>
</tr>
</tbody>
</table>

These values were used to calculate and confirm the values for accuracy, sensitivity and specificity of our methods, as shown in Eqs. (1) to (3):

\[
\text{sensitivity} = \frac{TP}{TP+FN} \quad (1)
\]

\[
\text{specificity} = \frac{TN}{TN+FP} \quad (2)
\]

\[
\text{accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)
\]

From the above equations, the following parameters can be computed. Thus:

\[
\text{sensitivity} = \frac{497}{497+2} = 99.5\%, \ 182992919
\]

\[
\text{specificity} = \frac{6}{6+7} = 46.15\%
\]
The performance of the proposed model was evaluated with two existing algorithms, i.e., support vector machine (SVM) and multilayer perceptron (MLP). In order to train with these algorithms, we flattened our data. Figures 8, 9, and 10 show the accuracy, sensitivity and specificity comparisons of the proposed model through corresponding graphical representations, respectively.

![Figure 8](image1.png) **Figure 8** Accuracy comparison of the developed model.  

![Figure 9](image2.png) **Figure 9** Sensitivity comparison of the developed model.  

![Figure 10](image3.png) **Figure 10** Specificity comparison of developed model.

The model was compared with some existing models prior to its development as presented in Table 3.
<table>
<thead>
<tr>
<th>Study</th>
<th>Method</th>
<th>Proposed solution and preprocessing</th>
<th>Software/tools/language/libraries used for simulation and implementation</th>
<th>Evaluation (Accuracy %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zuluaga-Gomez, et al. [23], 2021</td>
<td>CNN</td>
<td>Color processing</td>
<td>ResNet50, SeResNet50, and Inception</td>
<td>92%</td>
</tr>
<tr>
<td>Naji, et al. [27], 2021</td>
<td>SVM</td>
<td>Not available</td>
<td>Python/SciKit-learn</td>
<td>97.2%</td>
</tr>
<tr>
<td>Khan, et al. [24], 2019</td>
<td>Deep learning</td>
<td>Not available</td>
<td>GoogLeNet, VGGNet, ResNet</td>
<td>97.52%</td>
</tr>
<tr>
<td>Mambou, et al. [25], 2018</td>
<td>Deep learning/SVM</td>
<td>RGB conversion</td>
<td>Inception-v3</td>
<td>88.5%</td>
</tr>
<tr>
<td>Asri, et al. [26], 2016</td>
<td>SVM</td>
<td>Not available</td>
<td>WEKA</td>
<td>97.28%</td>
</tr>
<tr>
<td>Fagbuagun, et al.   (our model)</td>
<td>CNN</td>
<td>Data cleaning</td>
<td>Python/Google Colab/TensorFlow</td>
<td>98.25%</td>
</tr>
</tbody>
</table>

5 Conclusion

In order to contribute to the recurrent issue of breast cancer detection accuracy, we developed an 8-layered convolutional neural network, trained with 512 non-augmented data sets containing mammograms of benign and malignant breast cancer, and tested with 57 mammograms. The model was evaluated using three standard performance metrics, i.e., accuracy, sensitivity, and specificity.

The evaluation showed that our model achieved an accuracy of 98.25%, sensitivity of 99.5% and specificity of 46.15%, which are relative improvements compared to existing systems. The model developed can still be improved upon because various factors, such as data set size and data pre-processing methods may lower the accuracy of the detection rate in early detection of breast cancer. Furthermore, it was discovered that the deep learning model implementation methods still affects breast cancer detection accuracy.

Abbreviation

\[ TP = \text{True positive} \]
\[ TN = \text{True negative} \]
\[ FN = \text{False negative} \]
FP = False positive
VGGNet = Visual Geometry Group Network
ResNet = Residual networks
PNN/MP = Probabilistic neural network/multilayer perceptron
WEKA = Waikato Environment for Knowledge Analysis

References


