# A Hybridized Early Detection, Classification and Diagnosis of Breast Cancer Using Deep Learning Algorithm

Abubakar Mohammed[1*], Mohammed Ganga[2], Musa Wakil Bara[1] and Ibrahim Sani[3]

[1]Department of Computer Science

[2]Registry Department

[3]Department of Statistics

Mai Idris Alooma Polytechnic Geidam, Yobe State, Nigeria

[*]Corresponding author: gadaka2015@gmail.com

**Abstract**

There is a rise in the cases of breast cancer in low income population like northern Nigeria. Breast cancer is considered as one the major killer diseases among women of child bearing age (WCBA) (Aslan et al., 2018). There have been many researches on the identification and diagnosis of breast cancer disease for decades however, some of these researches are manual based and are inefficient due to their time consumption. With the recent advancement in ICT, machine learning algorithms are used to classify images (Kaji & Kida, 2019). A great achievement was made in classifying and detecting breast cancer using traditional machine learning algorithms such as Decision Tree and Artificial Neural Network (ANN) (Higa, 2018).

Despite the performance of these traditional machine learning algorithms in cancer prediction and diagnosis, there are common limitations that need to be addressed. These limitations include manual feature selection, fewer number of classes in classifying tumour (usually being classified into two classes) and their inability to classify larger dataset on time (Yari & Nguyen, 2020). This research is aimed at improving the performance of traditional machine learning algorithms by using a Deep Learning algorithms. Deep learning algorithm gives a promising result in image classification and can therefore extend the number of classes to more than the usual two (2) and solve the problem of fewer classes.

**Keywords:** Breast cancer, machine learning, deep learning.

## Introduction

One of the common disease among women around the world is the Breast cancer, this disease is considered as the second largest prevalent type of cancer which cause deaths among women of child bearing age (Higa, 2018). The disease is invasive among women with 35 years of age and older (Hamsagayathri, 2017).

The rise of the number of breast cancer cases in low income country like Nigeria is alarming (Zubair, 2009). Early detection, classification and diagnosis of breast cancer is undeniable. Even though there are many researches on the detection and classification of cancer disease, the breast cancer remains the complicated type of cancer. Manual method of detecting the cancer is inefficient due to time consumption.

However, recent advancement in ICT, machine learning algorithms used to classify images were employed to detect cancer and resulted in a great progress (Kaji & Kida, 2019). A great achievement was made in classifying and detecting breast cancer using traditional machine learning algorithms such as Decision Tree and Artificial Neural Network (ANN) (Higa, 2018). With the application of machine learning algorithms in breast cancer, one algorithm can perform the work of 20 pathologists (Sheykhmousa et al., 2020), this has made it more than necessary to apply these algorithms for speedy diagnosis and cost savings.

Even though these machine learning algorithms have promoted the classification and detection of breast cancer, they are however, lacking some features to make them work as desired. Among the limitations of these traditional machine learning algorithms are the number of classes (limited to only two classes), manual feature selection and their inability to classify larger dataset on time (Yari & Nguyen, 2020). The volume, variety, veracity and value of modern dataset which make the dataset to be called a big data is another threat to these traditional machine learning algorithms (Yadav & Jadhav, 2019).

With these challenges at hand, it is undeniable to improve the performance of these algorithms so as to fill the mentioned gaps.

This research is therefore aimed at improving the performance of the mentioned traditional machine learning algorithms by using a Deep Learning algorithms. Deep learning algorithm gives a promising result in image classification and can therefore extend the number of classes to more than the usual two (2) and solve the problem of fewer classes (Yadav & Jadhav, 2019).

The rest of the paper was organized as follows: In section 2, related scientific literature on breast cancer was presented. Research methodology including the description of our proposed system's architecture was presented in section 3. We present our experimental results and a discussion about them in section 4. Finally, in section 5, we presented the conclusions and recommended for future work.

## Related Literature Review

There is a rise in the cases of breast cancer in low income population like northern Nigeria. Breast cancer is considered as one the major killer diseases among women of child bearing age (Aslan et al., 2018). There have been many researches on the identification and diagnosis of breast cancer disease for decades however, some of these researches are manual based and are inefficient due to their time consumption. With the recent advancement in ICT, machine learning algorithms are used to classify images (Kaji & Kida, 2019). This makes it suitable for classifying tumours to cancer tissues. A great achievement was made in classifying and detecting breast cancer using traditional machine learning algorithms such as Decision Tree and Artificial Neural Network (ANN) (Higa, 2018).

With the application of machine learning algorithms in breast cancer, one algorithm can perform the work of 20 pathologists (Sheykhmousa et al., 2020), this has made it more than necessary to apply these algorithms for speedy diagnosis and cost savings.

Despite the performance of these traditional machine learning algorithms in cancer prediction and diagnosis, there are common limitations that need to be addressed. These limitations include manual feature selection, fewer number of classes in classifying tumour (usually being classified into two classes) and their inability to classify larger dataset on time (Yari & Nguyen, 2020).

It is considered to be the best choice for discovering complex architecture in high-dimensional data by employing back propagation algorithm. According to (Dargan et al., 2019), there are significant advancements and tremendous performance in numerous applications brought by deep learning, the widely used domains of deep learning are business, science and government which further includes adaptive testing, biological image classification, computer vision, cancer detection, natural language processing, object detection, face recognition, handwriting recognition, speech recognition, stock market analysis, smart city and many more.

The two famous classes of breast cancer are the benign and malignant tumors (Mekha, 2019)

**Benign tumors**

Benign tumors are not cancer. They can grow anywhere in the body but cannot invade to other areas of body. They often respond to the treatment and can be easily removed.

**Malignant tumors**

Malignant tumors are cancerous, which start from abnormal cells that highly grow in out of control and have ability to spread to other organs by circulating system or lymphatic system. We call that spread Metastasis. The common symptoms of breast cancer include a lump in the breast has changed the shape of the breast skin with fluid flow from the dimpled nipples or skin has red flakes

**Methodology**

This section describes the steps, processes, materials and methods that were employed to properly design and implement the proposed a hybridized early detection, classification and diagnosis of breast cancer using deep learning algorithms

Data sources, data types, data collection, coding and analysis are also explained. Technologies used to devise the system are also stated in this section.

**Steps of the research**

The following steps were used to design and develop the proposed model.

1. Feasibility study: this includes economic feasibility, is the research economically important to design the system? It is obvious that early detection, classification and diagnosis of breast cancer is a demanding goal in medical sector and especially in developing countries like Nigeria, setting aside a token amount to enhance this aspect is a welcome development notably to rural clients and with intervention of Tertiary Education Trust Fund, the research is economically the research is feasible.

2. Technical feasibility: are there technologies that can be used to develop the proposed system? With the recent advancement in the field of artificial intelligence and machine learning, the research is technically feasible.

Operational Feasibility: with little training and for the fact that, even the rural users can use

mobile phones, it takes a matter of days to train the users how to use the system. The research is therefore operationally feasible.

**Data Source, Collection and Analysis**

Wisconsin Breast Cancer (WBC) Dataset: this is breast cancer (diagnostic) dataset from UCI machine learning repository, which records the measurement for breast cancer cases. The dataset is available at https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(original)) to impute with missing data values by average value before using the classification algorithm. There are two classes, benign and malignant.

**Data Preparation**

This is a method of preparing the data ready for analysis, to use deep learning for data analysis, the data must be well prepared. Data preparation include data pre-processing which involves data cleaning. Data cleaning removes noise from the original data, all unwanted data or data that can affect the accuracy of the analysis are hereby removed from the data. The preprocessed data is shown in Fig. 1.

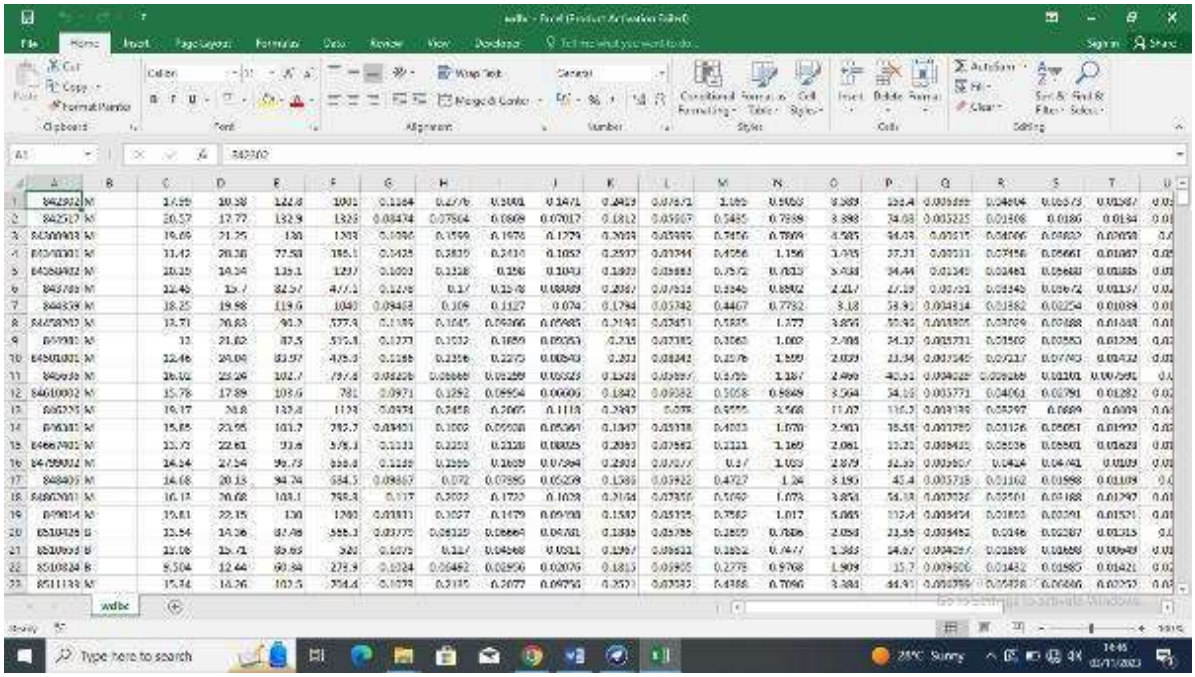

Figure 1 Preprocessed Data

**Additional Information**

Features are computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. They describe characteristics of the cell nuclei present in the image. A few of the images can be found at http://www.cs.wisc.edu/~street/images/

**Materials and Methods**

Materials required to develop the proposed system is **MATLAB,** is a data analysis software that is used to analyse the dataset collected from Wisconsin Dataset.

The proposed system is made up of three components that include Data Collection, Data Interpretation and application as shown in Fig. 2.

The U-net is well-known CNN that is developed for image segmentation applications (Rashed et al., 2019). It is a composition of successive convolution and

deconvolution layers as encoder-decoder architect. U-net provide an interesting architecture for image features extraction and used later in several imaging applications. Here, O-net which is a combination of two U-nets connected at the encoding level and disconnected at the decoding was used. This architecture enables us to apply convolution/deconvolution processes with different kernel structures that aims at finding features with different size/contrast. The proposed architecture of O-net is shown in figure. 2 and details of feature size at each network layer is shown in Table 1. The convolution process is implemented with kernel size of (3x3) pixels, (1x1) padding and (1x1) stride, except for the layers from 9 to 19 where two branches (up and bottom) are exists.

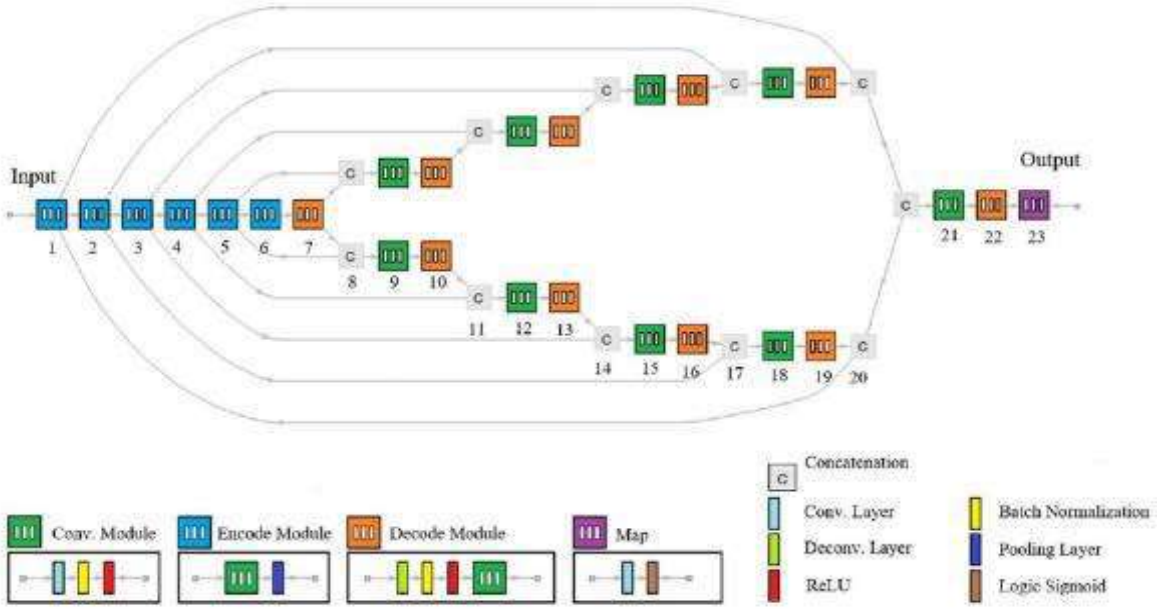


Figure 2: Architecture of the proposed System

Table 1. Network architecture of O-net.

| Layer | Feature size | Layer | Feature size |
|---|---|---|---|
| input | 1024×1024 | | |
| 1 | 8×512×512 | 13 | 32×128×128 |
| 2 | 16×256×256 | 14 | 2×32×128×128 |
| 3 | 32×128×128 | 15 | 32×128×128 |
| 4 | 64×64×64 | 16 | 16×256×256 |
| 5 | 128×32×32 | 17 | 2×16×256×256 |
| 6 | 256×16×16 | 18 | 16×256×256 |
| 7 | 128×32×32 | 19 | 8×512×512 |
| 8 | 2×128×32×32 | 20 | 2×8×512×512 |
| 9 | 128×32×32 | 21 | 4×8×512×512 |
| 10 | 64×64×64 | 22 | 8×512×512 |
| 11 | 2×64×64×64 | 23 | 4×1024×1024 |
| 12 | 64×64×64 | output | 1024×1024 |

**Results and Discussion**

The proposed CNN is implemented using MATLAB on a hp Laptop 4 Core i3 CPU @ 3.60 GHz, 64 GB memory and 3 *n*Vidia GeForce GTX 1080 GPU. Computations are conducted using GPUs to speed up the training. The network is trained for maximum 50 iterations with batch size of 4 images. The training can stop before reaching the maximum training iterations if the loss value difference is less than 10-3. We have used the default training optimizer algorithm, which is ADAM optimizer and images are randomly shuffled before training. Two networks are trained, one is for the set of microcalcifications and the other is for the set of masses. The network training time is less than 100 minutes (pre-processing is excluded).

The proposed O-net structure shows a promising results in challenging diagnosis problem. However, these are just initial results that need further justification to clearly identify the performance optimized parameters. For example, a more complex network structure with additional kernel size may contribute to the improvement of classification accuracy. Also, the identification of training set is still computed manually by selecting specific ROIs that include the abnormalities. Automatic identification of the ROI, or processing the raw mammograms is more useful for practical and clinical use.

**Performance Evaluation of CNN and other Classifiers**

The evaluated performance of the different classification methods for breast cancer into benign tumor and malignant tumor classes was performed on 698 number of dataset.

In Table 2 shows comparison of several classification models performance in breast cancer.

We used deep learning with activation function to classify and make ten-fold cross validation. Moreover, we compared methods with Naïve Bayes(NB and Random Forest(RF).

The performance evaluation compared the CNN algorithms with two other classifiers namely Naïve Bayes (NB) and Random Forest (RF) as shown in Table 2, that use the deep learning with exprectifier activation function. Accuracy of 96.99% which was better than other classification models was obtained. Random Forest and Naïve Bayes results for benign showed 98.00% and 98.65% respectively in precision, and 92.52% and 91.44% precision for malignant.

Table 2. Performance of breast cancer classification methods

| Methods | Precession | | Recall | | |
|---|---|---|---|---|---|
| | Benign | Malignant | Benign | Malignant | ACC |
| Deep Learning (CNN) | 98.44 | 94.35 | 96.94 | 97.10 | **96.99** |
| Naïve Bayes | 98.64 | 91.44 | 95.19 | 97.51 | 95.99 |
| Random Forest | 98.65 | 92.52 | 95.84 | 97.51 | 96.42 |

**Conclusion**

Deep learning approach has shown a promising result over the two powerful traditional Machine learning techniques, the Random Forest and Naïve Bayes. The comparison showed that CNN outperformed the machine learning algorithms with 96.99% against 96.42% of Random Forest.

**Future Work**

In the future, we are proposing to deploy the model into a mobile App for simplicity. The essence of deploying the model on mobile is to gain an offline access to remote users/clients. Once the mobile application (Mobile App) is installed on the mobile device, it would be use offline. This would avoid the cost of data purchase and also avoids problem of low network signal. Fig.3 shows how the model would be used on a Mobile App.

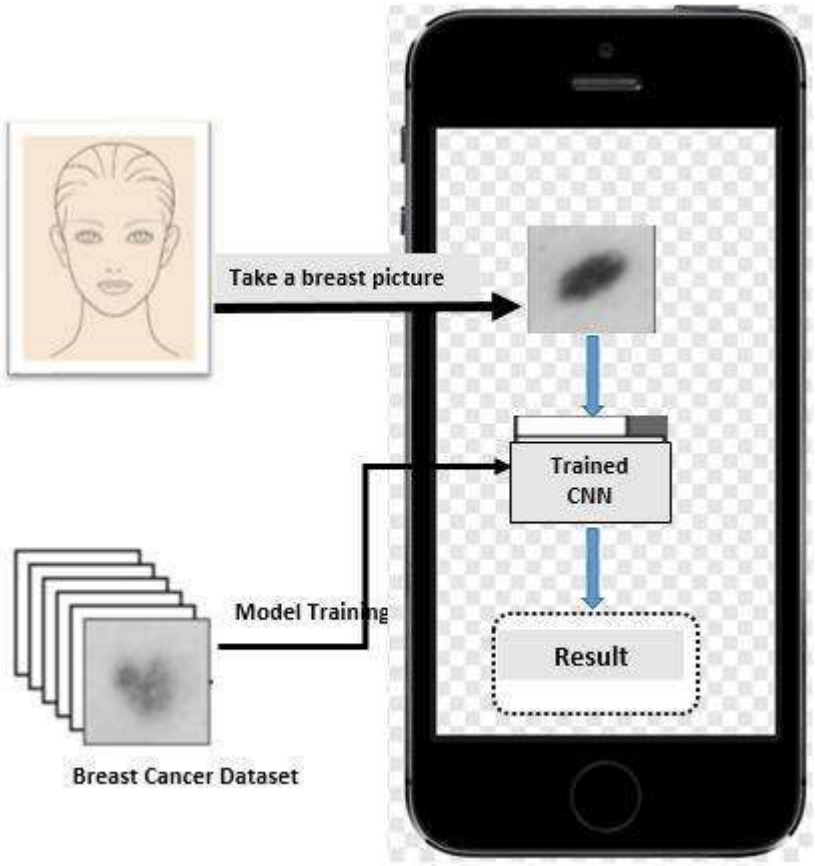


Figure 3 Deployment of the hybridized Model on Mobile App (proposed)

**Acknowledgement**

The researchers wish to extend their acknowledgement to the Tertiary Education Trust Fund (TETFund) Nigeria for sponsoring this research work. We also extend our gratitude to the management of Mai Idris Alooma Polytechnic Geidam for their efforts to secure the IBR intervention.

**References**

Aslan, M. F., Celik, Y., Sabanci, K., & Durdu, A. (2018). *Intelligent Systems and Applications in Engineering Breast Cancer Diagnosis by Different Machine Learning Methods Using Blood Analysis Data*. 2–6. https://doi.org/10.1039/b000000x

Dargan, S., Kumar, M., Rohit, M., & Gulshan, A. (2019). A Survey of Deep Learning and Its Applications : A New Paradigm to Machine Learning. *Archives of Computational Methods in Engineering*, *0123456789*. https://doi.org/10.1007/s11831-019-09344-w

Hamsagayathri, P. (2017). *Priority Based Decision Tree Classifier for Breast Cancer Detection*. 4–9.

Higa, A. (2018). *Diagnosis of Breast Cancer using Decision Tree and Artificial Neural Network Algorithms*. *7*(01), 23–27.

Kaji, S., & Kida, S. (2019). *Overview of image-to-image translation by use of deep neural networks :*

Mekha, P. (2019). Deep Learning Algorithms for Predicting Breast Cancer Based on Tumor Cells. *2019 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT-NCON)*, 343–346.

Rashed, E., Abou, M. S., & Seoud, E. (2019). *Deep learning approach for breast cancer diagnosis*.

Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., & Member, S. (2020). *Support Vector Machine Versus Random Forest for Remote Sensing Image Classification : A Meta-Analysis and Systematic Review*. *13*, 6308–6325.

Yadav, S. S., & Jadhav, S. M. (2019). Deep convolutional neural network based medical image classification for disease diagnosis. *Journal of Big Data*. https://doi.org/10.1186/s40537-019-0276-2

Yari, Y., & Nguyen, T. V. (2020). *Deep Learning Applied for Histological Diagnosis of Breast Cancer*. *8*. https://doi.org/10.1109/ACCESS.2020.3021557