

## A Computer Vision Model for Early Detection and Classification of Breast Cancer Using Deep Learning Algorithms

Alhaji Audu Goni\* , Usman Ngudama and Musa Wakil Bara

Department of Computer Science, Mai Idris Aloomo Polytechnic Geidam, Yobe State.

Corresponding author: [aagoni2016@gmail.com](mailto:aagoni2016@gmail.com)

### **Abstract**

Breast cancer remains a leading cause of global mortality, though early intervention significantly improves prognoses. The conventional screening technique, X-ray mammography, presents challenges for the early identification of lesions. The compression involved in the imaging process often obscures subtle abnormalities within dense breast tissue. Furthermore, the considerable inter- and intra-patient variability of breast anatomy complicates accurate diagnosis when relying on manually engineered features. Deep learning, a branch of machine learning demanding substantial computational resources, has demonstrated remarkable efficacy in complex, intelligence-driven tasks. This paper introduces a novel neural network architecture, derived from the U-net model, designed for the effective and early detection of breast cancer. The results, which show high sensitivity and specificity, suggest the model's potential utility in clinical environments.

**Keywords:** *U-net, mortality, breast, 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 and resulted in a great progress (Kaji & Kida, 2023). 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 in, 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 mention 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 remaining section of the research proposal is organized as follows: the problem statement followed the introduction section, then the objectives of the research are stated immediately after the problem statement. Literature of related work follows the objects section and finally methodology and expected results respectively.

## **Problem Statement**

The recent research on the application of machine learning algorithms for breast cancer classification and diagnosis is faced with the following problems:

1. Lack of early detection of breast cancer exposes the patient to risk their lives;
2. Classification techniques are based on handcrafted features techniques, and their performance relies upon the chosen features (Alzubaidi et el, 2020).
3. Lack of deployment of the model makes it difficult to be accessed easily by clients (Dai & Chapman, 2019).

## **Objectives of the Study**

This research which is aimed to develop an improved model for breast cancer detection, classification and diagnosis has following objectives:

1. To develop a cost-effective model that detects breast cancer at early stage using Deep learning algorithm;
2. To employ a Deep Learning Algorithm for automated features selection techniques;
3. To Improve the Computational power needed to capture and classify big dataset on time and
4. To deploy the model on mobile device for use by clients of low income countries like Nigeria.

**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 time consuming.

With the recent advancement in ICT, machine learning algorithms are used to classify images (Kaji & Kida, 2023), this classification can be used to classify cancer tissues such as breast cancer. A great achievement was made in classifying and detecting breast cancer using traditional machine learning algorithms. The work of (Bharat et al., 2018) has applied Decision Tree and Artificial Neural Network for breast cancer detection and diagnosis, there was an accuracy of over 75% in his classification. Priority based decision tree classifier algorithm and J48 showed a relatively good accuracy in breast cancer classification in (Hamsagayathri, 2017)

Support Vector Machine and Random Forest were used on 45 dataset to classify breast cancer in tumours and the accuracy classification of 80% was achieved (Sheykhmousa et al., 2020). Decision tree algorithms J48, Function Tree, Random Forest Tree, AD Alternating Decision Tree, Decision stump and Best First are applied to are applied to 569 dataset with 357 benign and 212 malignant cases with 32 attributes to test and proof the differences among the classification methods or algorithms and have achieved accuracy between 80%to 97.5% (Al-salihy, 2017).

**Table 1:** Summary of Traditional Machine Learning Algorithms used in Breast Cancer Prediction and their Limitations

S/No	Reference	Algorithm Used	Evaluation Algorithm	Research Findings	Issues/Open Opportunity
1.	(Agarap, 2018)s	GRU-SVM	NN, LR, MLP, Softmax Regression and SVM	MLP outperformed the proposed GRU-SVM	Lacks the implementation of cross-validation, hence low accuracy.
2.	(Bharat et al., 2018)	SVM	KNN, CART and Naïve Bayes	Highest Accuracy of 99% was obtained	Use on dataset with on two classes.
3.	(Al-salihy, 2017)	ELM, SVM, K-NN & ANN	HOMA	High accuracy of 83% was achieved	No data normalization

4.	(Higa, 2018)	Decision Tree and ANN	Not stated	Reliable result with 95.4% accuracy	Lacks data normalization which can affect the accuracy of the algorithms
5.	(Hamsagayathri, 2017)	J48	Not stated	Accuracy of 95% was made	Model not evaluated in comparison with other classification
6.	(Zhang et al., 2019)	Fast Fuzzy C-Means	Compared with K-Means and C-Means	Accuracy of 90% was achieved	Dataset was classified in to only two classes
7.	(Zheng, 2013)	Hybrid of K-Means and C-Means	Not stated	Achieved high accuracy of 97.38%	As in
8.	(Aslan et al., 2018)	J48	Not stated	Accuracy of 95% was made	Model not evaluated in comparison with other classification

Even though the traditional machine learning algorithms have done well in breast cancer classification, there are many limitations that include time consumption, manual feature selection and limited number of classes (Lee et al., 2018) and (Yadav & Jadhav, 2019).

To improve the performance of traditional machine learning algorithms, a new machine learning algorithm called deep learning was proposed by different researchers and an average improvement was achieved (Maier et al., 2019). Deep learning shows promising results in image classifications including that of breast cancer classification (Dargan et al., 2019).

Among the advantages of Deep learning algorithms are their ability to eliminate manual feature selection in dataset, extend the number of classes in to four, handle big dataset and above all minimizes time (Golden, 2017).

**Definition of Some Terms in the Study**

A. Machine Learning: this is the study of computer algorithms that can improve automatically through experience and by the use of data. It is part of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions within minimal human intervention. Machine learning evolved from left to right from supervised learning to Reinforcement Deep learning as shown in Fig.

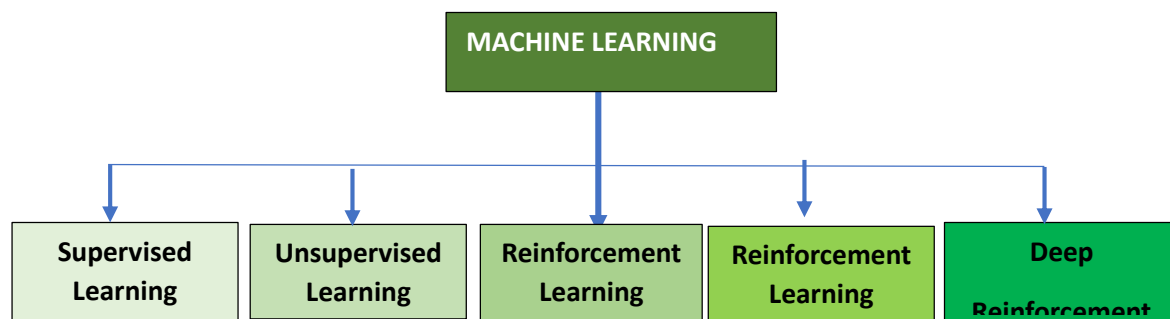


Figure 1: Evolution and classification of Machine Learning

a. Supervised Learning is a machine learning method where a computer learns from examples that have pre-defined answers. Think of it like teaching with flashcards: you show the model many pairs of data (an input 'X' and its correct output 'Y') until it learns the pattern connecting them.

After this training phase, the model's understanding is tested using a separate set of known examples to see if it can correctly predict the Y's from the X's. If its accuracy meets a target threshold, the training is complete. The model is then ready to be used for its real purpose: making predictions on entirely new data where the answer isn't known in advance.

b. Unsupervised Learning: In unsupervised learning, we do not specify a target variable to the machine, rather we ask machine "What can you tell me about X?". More specifically, we may ask questions such as given a huge data set X, "What are the five best groups we can make out of X?" or "What features occur together most frequently in X?". To arrive at the answers to such questions, you can understand that the number of data points that the machine would require to deduce a strategy would be very large.

c. Reinforcement Learning: this is an ML method based on rewarding desired behaviour and/or pushing undesired ones.

d. Deep Learning: The deep learning is a model based on Artificial Neural Networks (ANN), more specifically Convolutional Neural Networks (CNN)s. There are several architectures used in deep learning such as deep neural networks, deep belief networks, recurrent neural networks, and convolutional neural networks.

These networks have been successfully applied in solving the problems of computer vision, speech recognition, natural language

processing, bioinformatics, drug design, medical image analysis, and games. There are several other fields in which deep learning is proactively applied. The deep learning requires huge processing power and humongous data, which is generally easily available these days.

e. The Deep Reinforcement Learning (DRL) combines the techniques of both deep and reinforcement learning. The reinforcement learning algorithms like Q-learning are now combined with deep learning to create a powerful DRL model. The technique has been with a great success in the fields of robotics, video games, finance and healthcare. Many previously unsolvable problems are now solved by creating DRL models. There is lots of research going on in this area and this is very actively pursued by the industries.

**Image Classification:** this is the process of categorizing and labelling group of pixels or vectors within an image based on specific rules.

**Breast Cancer:** this is a type of cancer that starts in the breast. It can start in one or both breasts. Cancer starts when cells begin to grow out. The cancer forms in either the lobules or the ducts of the breast. When breasts grow abnormally, the following symptoms begin to appear

- i. Breast lump or thickening that feels different from the surrounding tissue;
- ii. Change in the size, shape or appearance of breast;
- iii. Changes to the skin over the breast such as dimpling;
- iv. A newly inverted nipple;
- v. Peeling, scaling, crusting or flaking of the pigmented area of skin surrounding the nipple (areola) or breast skin

This section describes the steps, processes, materials and methods that would be 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.

## **Methodology**

### **Steps of the research**

The following steps would be used to design and develop the smart agriculture model for wheat production.

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 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.

3. Operational Feasibility: with little training and for the fact that, even the rural farmers use mobile phones, it takes a matter of days to train the farmers how to use the system. The research is therefore operationally feasible.

### **Data Source, Collection and Analysis**

The dataset used for this research is the 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+\(original\)\)](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 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

### **Materials and Methods**

Materials required to develop the proposed system include:

1. MATLAB: this is a data analysis software that is used to analyse the data collected either from the Wisconsin Dataset or the locally collected dataset.

### **Architecture of the proposed system**

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

This study utilizes the U-net CNN, a well-established encoder-decoder architecture for image segmentation based on successive convolutional and deconvolutional layers (Rashed et al., 2019). Its efficacy in feature extraction has led to its widespread use in imaging applications.

**Proposed O-net Architecture:** We introduce the O-net, a new architecture formed by two U-nets connected at the encoding level but disconnected during decoding. This

configuration allows for concurrent convolution/deconvolution processes with different kernels, specifically aimed at capturing features of varying sizes and contrasts.

**Implementation Details:** The convolutional layers predominantly use a 3x3 kernel with 1x1

padding and stride. A structural deviation occurs in layers 9 through 19, where the architecture splits into independent "up" and "bottom" branches. The complete O-net architecture is depicted in Figure 2, and the corresponding feature dimensions for each layer are enumerated in Table 1.

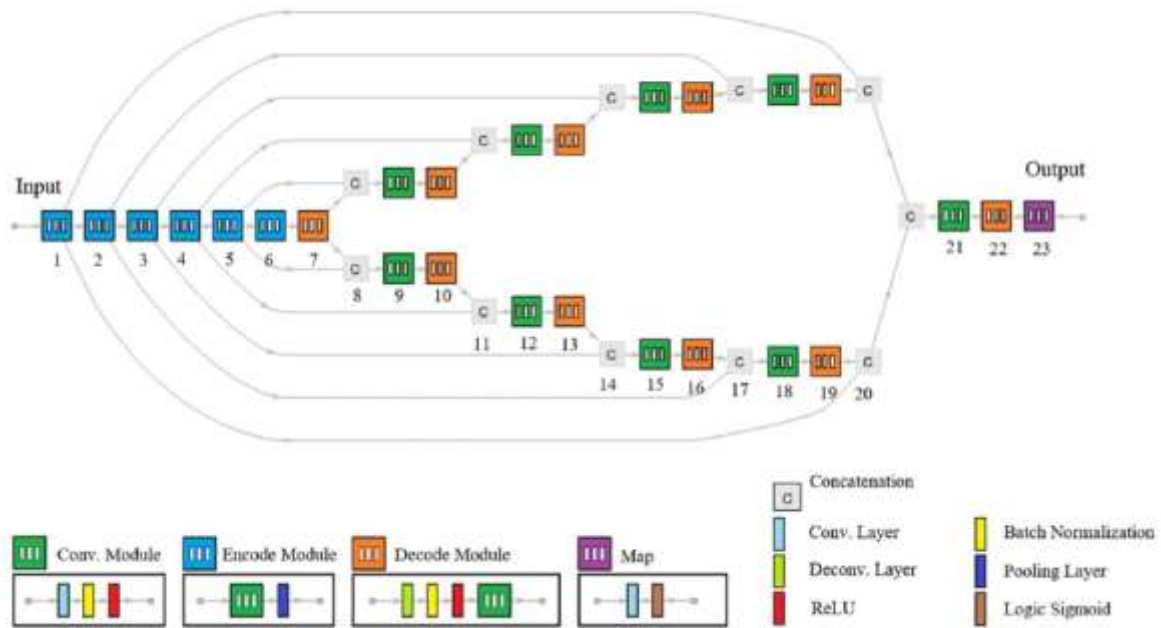


Figure 2: Architecture of the proposed System

Deployment of the model on Mobile Phone

### Results and discussions

**Implementation Details:** The model was developed and trained in MATLAB 2024Ra on a high-performance computing platform featuring a 3.60 GHz CPU, 64 GB RAM, and three nVidia GeForce GTX 1080 GPUs. All training computations were GPU-accelerated.

**Training Protocol:** The network was trained for up to 50 epochs using a batch size of 4 and the ADAM optimizer. To prevent overfitting and optimize training time, an early stopping

mechanism was implemented, which terminated training if the loss difference was less than  $10^{-3}$ . The input images were randomly shuffled prior to training.

Two distinct models were trained: one for microcalcification detection and another for mass detection. The complete training procedure for each network was completed in less than 100 minutes (pre-processing not included). Sample results are demonstrated in Figure 3.

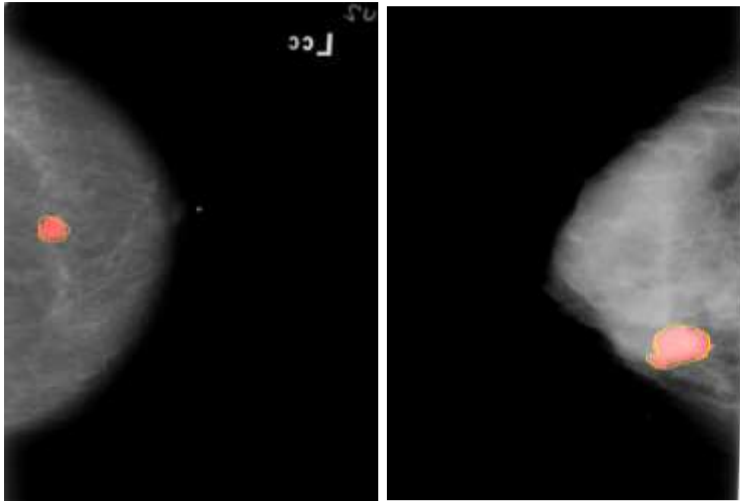


Figure 3: Abnormality detection results for microcalcification (left) and mass (right). Red labeled region is golden truth label and region identified with yellow line is the network labeled region.

The classification accuracy for microcalcification and masses are 94.31% and 95.01%. These results outperform reported results using well-known structures such as AlexNet, VGGNet and GoogleNet as reported in [9]. One potential reason for these results is due to the combination of two different convolution kernels at the encoder tracks.

### Conclusion

Our study demonstrates a convolutional neural network for detecting breast cancer in mammograms, highlighting the strength of deep learning for this application. The proposed system has the potential to improve the clinical performance of early breast cancer detection. However, the optimal network architecture for digital mammogram textures remains an open question, warranting further investigation into various design modifications and a comprehensive analysis of their sensitivity.

### Acknowledgement

The authors extend their acknowledgement to the sponsor of this research, the Tertiary Education Trust (tetfund) and Mai Idris Aloomaa Polytechnic for giving us this offer in its 2025 IBR intervention.

### Reference

*A Computer Vision Model for Early Detection and Classification of Breast Cancer Using Deep Learning Algorithms*

Agarap, A. F. M. (2018). *On Breast Cancer Detection: An Application of Machine Learning Algorithms on the Wisconsin Diagnostic Dataset*. 1, 5–9.

Al-salihy, N. K. (2017). Classifying Breast Cancer by Using Decision Tree Algorithms. *ACM*.

Alzubaidi et al. (2020). *Optimizing the Performance of Breast Cancer Classification by Employing the Same Domain Transfer Learning from Hybrid Deep Convolutional*.

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>

Bharat, A., Pooja, N., & Reddy, R. A. (2018). Using Machine Learning algorithms for breast cancer risk prediction and diagnosis. *2018 3rd International Conference on Circuits, Control, Communication and Computing (I4C)*, x, 1–4.

- Dai, X., & Chapman, S. (2019). *Machine Learning on Mobile: An On-device Inference App for Skin Cancer Detection*. 301–305.  
<https://doi.org/10.1109/FMEC.2019.8795362>
- 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>
- Golden, J. A. (2017). *Deep Learning Algorithms for Detection of Lymph Node Metastases From Breast Cancer Helping Artificial Intelligence Be Seen*. 12–14.
- Hamsagayathri, P. (2017). *Priority Based Decision Tree Classifier for Breast Cancer Detection*. 4–9.
- He, M., Lin, D., Gao, Z., & Fan, J. (2020). *Deep Learning Assisted Efficient AdaBoost Algorithm for Breast Cancer Detection and Early Diagnosis*. 8.
- Higa, A. (2018). *Diagnosis of Breast Cancer using Decision Tree and Artificial Neural Network Algorithms*. 7(01), 23–27.
- Kaji, S., & Kida, S. (2023). *Overview of image-to-image translation by use of deep neural networks* :
- Lee, S., Chen, T., Yu, L. U. N., & Lai, C. (2018). Image Classification Based on the Boost Convolutional Neural Network. *IEEE Access*, 6, 12755–12768.  
<https://doi.org/10.1109/ACCESS.2018.2796722>
- Maier, A., Syben, C., Lasser, T., & Riess, C. (2019). REVIEW A gentle introduction to deep learning in medical image processing. *Zeitschrift Fßr Medizinische Physik*, 29(2), 86–101.  
<https://doi.org/10.1016/j.zemedi.2018.12.003>
- 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.
- Wang, D., & Khosla, A. (2016). *Deep Learning for Identifying Metastatic Breast Cancer*. 1–6.
- 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>
- Zhang, W., Shi, H., Zeng, J., Sun, S., Peeters, W. H., Li, S., Liu, Z., & Zhao, H. (2019). *Classification of Breast Cancer using Fast Fuzzy Clustering based on Kernel*

*A Computer Vision Model for Early Detection and Classification of Breast Cancer Using Deep Learning Algorithms*

*Classification of Breast Cancer using Fast Fuzzy Clustering based on Kernel.*  
<https://doi.org/10.1088/1757-899X/546/5/052067>

Zheng, B. (2013). *Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms.*