

Breast Cancer Prediction and Control Using BiLSTM and Two-Dimensional Convolutional Neural Network

Moses A. Agana, University of Calabar, Nigeria*

Chukwuemeka Odi Agwu, Ebonyi State University, Abakaliki, Nigeria

Nsinem A. Ukpoho, University of Calabar, Nigeria

ABSTRACT

Breast cancer has a devastating effect on women. Different strategies of breast cancer classification exist with minimal work done on the prediction of the occurrence of the disease in potential carriers. In this study, a breast cancer predictive system has been developed using bidirectional long short-term memory (BiLSTM) for feature extraction and learning while the two-dimensional convolutional neural network (CNN) was used for breast cancer classification. Histopathological images were used for cancer prediction. Python was used as the programming language for implementing the system. The model was tested using datasets from The Cancer Imaging Archive (TCIA) repository. An accuracy level of 98.8% (higher than the most recent existing model) was achieved for the prediction of the future occurrence of breast cancer based on the tests on the dataset. The application of the model using live data from women can help in the prediction and control of the occurrence of breast cancer amongst women.

KEYWORDS

Breast, Cancer, Classification, Convolutional, Network, Neural, Prediction

INTRODUCTION

Breast cancer is a form of invasive cancer and one of the most common health problems for women that is globally responsible for a large number of deaths (Ali, 2019). According to the World Health Organisation's projection, every year, the estimated number of breast cancer diagnosis amongst women is up to 1.5 million (WHO, 2014). Factors contributing to breast cancer mortality have been a topic of intense research and discussion in the scientific world. There is, however, a dearth of information on the incidence of breast cancer mortality in most resource-poor countries. Available data from most African workers on breast cancer focused on incidence, risk factors, and complications rather than mortality (Mohammed et. al., 2017). Breast cancer is believed to affect one in eight women during

DOI: 10.4018/IJSI.316169

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

their lives. It is the commonest site specific malignancy affecting women and the most common cause of cancer mortality in women worldwide. Cancer is a disease in which abnormal cells grow in an uncontrolled way (Siegel et. al., 2016). Breast cancer is a malignant (cancerous) growth that begins in the tissues of the breast. It is the most common cancer in women, but it can also appear in men. Breast cancer is now an epidemic, posing a serious threat to the health of women of all races globally. In Nigeria, for instance, breast cancer is the leading cause of cancer related deaths among women. This is not due to a reduction in cervical cancer but an increase in the incidence of breast cancer. Breast cancer is commonly seen in four stages that represents its progression (Lee et al., 2010).

In stage-1, the disease is confined entirely to the breast. The cancer usually starts as a very tiny growth that cannot yet be felt but can be detected with imaging tests such as mammography and ultrasound. At this first stage, treatment is usually curative and more than 95% of those so detected will survive the disease beyond 5 years (Egenti, 2016). Stage II is a cancer that has involved lymph nodes in the armpit of the same side of the breast, while stage III disease is one that involves the muscles under the breast. Stages II and III therefore require very aggressive treatment using different modalities to contain the spread of the disease. It is however difficult to cure a patient in stage IV because the disease has spread and may have involved other organs in the body such as the lungs, liver, bones, the brain or the spine (Lorena et al., 2011; Egenti, 2016). The five-year survival rate for breast cancer patients in the United States exceeds 85%, in Nigeria it is a dismal 10% and is responsible for about 16% of all cancer related deaths (Mohammed et al, 2017).

To reduce death rate resulting from the disease, early detection and diagnosis are critical. Better diagnostic tools and method can minimize the fatality rate. Breast cancer diagnosis allows identifying the cancer cells; if the diagnosis tools become more efficient, then the detection and prediction can be more effective (Giu & Jyh-Cheng 2015; Karabatak, 2015).

Machine Learning (ML) is a subfield of Artificial Intelligence (AI) which allows machines to learn with or without the intervention of a human. Machine Learning has multiple potential applications in medicine and has been applied to a wide variety of oncology tasks, such as predicting disease susceptibility, survival rates and treatments. In the field of AI, ML is one of the most popular models which has been implemented rapidly to train machines and develop predictive models for effective decision making. In classification and prediction problems, Machine Learning methods are the leading methods for obtaining a better outcome. In cancer research, the ML methods could be used for identification and prediction of cancer. These ML methods could predict whether cancer is malignant or benign (Kononenko, 2001).

Many researchers have devoted efforts to developing a high-performance and reliable computer-aided system to help medical staff diagnose breast cancer using histopathological microscopic image and improve diagnostic efficiency. The diagnostic process of breast cancer is not only time consuming and expensive but also largely dependent on the consistency of the pathologist' existing knowledge and pathological reports (Yao et. al., 2019). The increased volume of medical data to be interpreted and filtered for diagnostic and therapeutic purposes calls for the adoption of deep learning models in medical diagnosis (Ursuleanu et al., 2021).

The standard procedure to diagnose breast cancer by pathologists usually requires extensive microscopic assessments. Therefore, having an automated solution like a Computer Aided Diagnosis (CAD) system not only contributes to an easier diagnostic process, but also reduces the subjectivity in diagnosis. Image classification which can be defined as the task of categorizing images into one of several predefined classes, is a fundamental problem in computer vision. It forms the basis for other computer vision tasks such as localization, detection and segmentation (Karpathy, 2016).

With the advanced development of artificial intelligence, many machine learning techniques have been applied to CAD systems. This technique can potentially outperform humans and learn more efficiently with time, therefore integrating machine learning in diagnosis can supply useful knowledge to assist pathologist in evaluating and analyzing enormous amounts of medical data (Ciresan et. al.,

2011). It could also speed up the process due to the capability to process large data much faster than manual diagnosis by a pathologist.

Bias and imbalance class problem among datasets can lead to undesired classification for the diagnosis result using CAD. Class imbalance occurs in machine learning classifiers when predicted class probabilities are geared toward the majority class ignoring the significance of minority classes. This imbalance causes the outcome to be biased toward a certain class of the dataset. When a CAD system is built upon a dataset with imbalanced classes, the result will be more likely to be biased and therefore produce wrong diagnosis. When a trained model is biased to a specific class due to the imbalanced dataset it destroys the reliability of a CAD system because it will increase the rate of wrong classification (He & Garcia, 2009). A breast cancer predictive system using machine learning algorithms can be a panacea for the aforementioned problems.

Classification and data mining methods are an effective way to classify data. Especially in medical field where those methods are widely used in diagnosis and analysis to make decisions. Classification is one of the most important and essential task in machine and data mining. About a lot of research has been conducted to apply data mining and machine learning on different medical datasets to classify Breast cancer. Many of them show good classification accuracy (Hiba et. al., 2016).

Convolutional Neural Network (CNN) has attracted much attention for the analysis of histopathological images, because of its steadily improving performance that is nearly as accurate as or better than human experts (Goodfellow et. al., 2016).

Neural Networks are models loosely based on the structure of the brain. These models are made up of layers of neurons that relate to each other via weighted connections. These weights are adjusted during training through a process called back propagation. Neural network can handle noisy data and model complex non-linear functions well (Lorena et.al., 2011).

The foundation of most modern deep learning models is artificial neural networks. The term “deep learning” was introduced to refer to the use of many layers of neural networks to progressively generate features from the original input data. Networks with more layers can learn more complex functions, thus explaining the power of deep learning. They are composed of multiple inter-connected layers, each of which consists of separate computational units called neurons (Wang & Gao, 2019). The input information flows through the network as follows: each layer receives input data for each of its neurons, each neuron then executes a simple user-defined function, and then the output of the neuron is transmitted as input to neurons in the next layer. The connections are weighted, reflecting the contribution to the prediction. The learning process of a neural network is the updating of these connection weights, based on prediction errors made with training data. By composing the numerous simple functions executed by each neuron in a network structure, complex relationships between input and their relevance to the output can be learned. The most successful architectures of neural networks are convolutional neural networks (CNNs) and RNN, which are now cornerstone of all leading methods in image classification and natural language processing, respectively (Wang & Gao, 2019).

The CNN has a powerful feature extraction ability. As the depth of the CNN structure increases the problem, of gradient disappearance becomes more and more obvious (Huang et. al, 2017). CNN contains convolutional layers, dropout layers and an output layer, hierarchically positioned that each learn specific characteristics in the image (Vizcarra, et al., 2019). The CNN has a powerful feature extraction ability. As the depth of the CNN structure increases, the problem of gradient disappearance becomes more and more obvious (Huang et. al., 2017). CNN contains convolutional layers, grouping layers, dropout layers and an output layers, hierarchically positioned that each learn specific characteristics in the image (Vizcarra et al., 2019).

In summary, the introduction has x-rayed the concept of breast cancer, the health implications of having breast cancer, the challenges of predicting and diagnosing breast cancer, as well as the need to apply a combination of machine learning techniques to predict and diagnose breast cancer.

RELATED WORKS

In a study by Patrizia, *et. al.*, (2019), Machine Learning was used to develop a prognostic classification model that can be used to predict outcomes in individual cancer patients. Machine learning based Decision Support System (DSS), combined with Random Optimization (RO), was used to extract prognostic information from routinely called demographic, clinical and biochemical data of breast cancer (BC) patients. The DSS model was developed in a training set (n= 318), whose performance analysis in the testing set (n=136) resulted in a C-index for progression free survival of 0.84 with an accuracy of 186.

In a research carried out by Saria and Huda (2019), they reviewed the role of machine learning and data mining techniques in breast cancer detection and diagnosis. They reviewed a total of 46 researches and came to the conclusion that many of the studies they reviewed focused mainly on application of classification techniques to breast cancer prediction rather than studying various home data cleaning and pruning techniques that can prepare and make a dataset suitable for mining. They observed that a good dataset provides better accuracy. Selection of appropriate algorithms with good home dataset will lead to development of prediction systems. These systems can assist in proper treatment methods for a patient diagnosed with breast cancer (Saria & Huda, 2019).

In a research carried out by Wembin *et al.*, (2018), they stated that the application of classification technique is widely used as a machine learning technique to identify people with Breast Cancer to distinguish benign from malignant tumors and to predict prognosis. They observed that although they tried to find the best algorithm to achieve the most accurate classification result, data of variable quality influenced the classification result.

Bogdan *et al.*, (2019) conducted a study on breast cancer classification on histopathological images affected by data imbalance using Active Learning and Deep Convolutional Neural Network. In the work, they proposed an algorithm for training deep neural networks for classification of breast cancer in histopathological images affected by data unbalance with the support of Active Learning. They used the output of the neural network on unlabeled samples to calculate weighted information entropy. It was utilized as uncertainty score to automatically select both samples with High (H) and Low (L) confidence. A number of low confidence samples that was selected in each iteration was manually labeled by the pathologist. A threshold that decays over iteration number was used to decide which high confidence samples should be concatenated with manually labeled samples and then used in fine-tuning the convolutional neural network. The neural network could be optionally trained using weighted cross-entropy loss to better cope with bias towards the majority class.

Schat *et al* (2020) proposed the Data Representativeness Criterion (DRC) to ascertain how representative a training data set is of a new unseen data set. They presented a proof of principle, to see whether the DRC could quantify how similar the data sets were and whether the DRC relates to the performance of a supervised classification algorithm. They compared a number of magnetic resonance imaging (MRI) data sets with varying severity parameters. The results indicated that, based on the similarity of data sets, the DRC is able to indicate when the performance of a supervised classifier decreases.

Oluwashola (2021) researched on prediction of breast cancer images classification using bidirectional long short term memory and two-dimensional convolutional neural network. He optionally trained the neural network using weighted cross-entropy loss to take care of bias towards the majority class. Upon comparing the developed model with an existing model, an accuracy level of 98,3% was realized, against the 93.97% in the existing model.

Aruna, Rajagopalam & Nandakishore (2011) compared the performance of C4.5, Naïve Bayes, Support Vector Machine (SVM) and K- Nearest Neighbor (K-NN) to find the best classifier in WBC. SVM proved to be the most accurate classifier with accuracy of 96.99%.

Chaurasia & Pal (2014) compared the performance criterion of supervised learning classifiers as Naïve Bayes, SVM-RBF neural networks, Decision trees (J48) and simple CART, to find the best

classifier in breast cancer datasets. The experimental results showed that SVM-RBF kernel is more accurate than other classifiers; it scores accuracy of 96.84% in Wisconsin Breast (original datasets).

Djebbari et al (2008) considered the effect of ensemble machine learning techniques to predict the survival time in breast cancer. Their technique showed better accuracy on their breast cancer dataset compared to previous results. In a related development, Aruna, Rajagopalam & Nandankishore (2011) compared the performance of 4.5, Naïve Bayes, Support Vector Machine (SVM) and K-Nearest Neighbor (K-NN) to find the best classifier in NBC. SVM proved to be the most accurate classifier with accuracy of 96.99%. Similarly, Christobel and Sivaprakasam (2011) achieved accuracy of 69.23% using decision tree classifier in breast cancer datasets.

Sameer, Yudhveer & Basant (2020) conducted a study on the use of boundary detection to segment the pectoral muscle from digital mammograms images. They stated that radiologists recognize the sign of breast cancer by performing X-ray called screening mammography and the biggest problem during such analysis of mammography arises due to the presence of pectoral muscle which is the mass of tissue on which the breast rests. As a result of the confusion this generates in recognizing tumour cells, they reviewed different segmentation techniques for pectoral muscle removal in mammograms through digital images. They concluded that there is no specific technique that proffers absolute solution to the problem of pectoral muscle segmentation and that in most situations, the solution given concentrates more on a particular collection of information or a particular issue in hand and that the findings obtained from accessible study articles are very hard to quantify.

In summary of the related works include contributions by various authors relating the application of various machine learning techniques and a combination of most algorithms for the prediction and diagnosis of breast cancer and related ailments. The review has shown that various researchers have made tremendous efforts apply machine learning to the diagnosis of breast cancer, but much work still needs to be done in the area of predicting the occurrence of the disease in potential careers.

METHODOLOGY

In this study, the combination of Bi-LSTM and Conv2D Neural Network algorithms has been adopted because the Bi-LSTM can actively learn the dataset and perform feature selection from the dataset so as to reduce feature dimensionality and the encoded features from the Bi-LSTM will then be fed into the Conv2D to eventually make the classification. This will reduce the learning curve of the model and allow for better prediction and classification.

Data for the study was gathered from TCIA repository using Bidirectional Long Short Term Memory and Two-Dimensional Convolutional Neural Machine Learning algorithms to classify and train the data sets for breast cancer prediction and control. To achieve this:

- i. Data from expert medical practitioners on early signs that can be used to predict the later occurrence of breast cancer in patients was collected.
- ii. Python programming language was used to simulate and implement the system.

An analysis of the prevalent situation was made. It was observed that in the study area, there is no breast cancer predictive and control system. What is prevalent is a diagnostic process employed to discern between benign and malignant cancerous patterns. Mammography is one of the conventional approaches for breast cancer diagnosis along with sonography and MRI (Magnetic Resonance Imaging). The prevalent approach for diagnosing breast cancer is the BI-RADS classification.

The training was achieved with the support of Active Learning (AL). Instead of random selection, AL methods typically actively select samples with lowest confidence as the most valuable samples to add them to the query and finally train the model incrementally. Randomly selecting samples instead of actively choosing samples establishes a lower bound. In this method, both samples with high and low confidence were included in the query; and to achieve this, the Bi-LSTM algorithm was used.

The prediction is aimed at improving the life expectancy of breast cancer patients and reducing the overall cost of treatment. It can also serve as a reference to physicians and researchers who are interested in investigating ways of predicting and controlling breast cancer in women.

The conceptual diagram of the system is shown in figure 1.

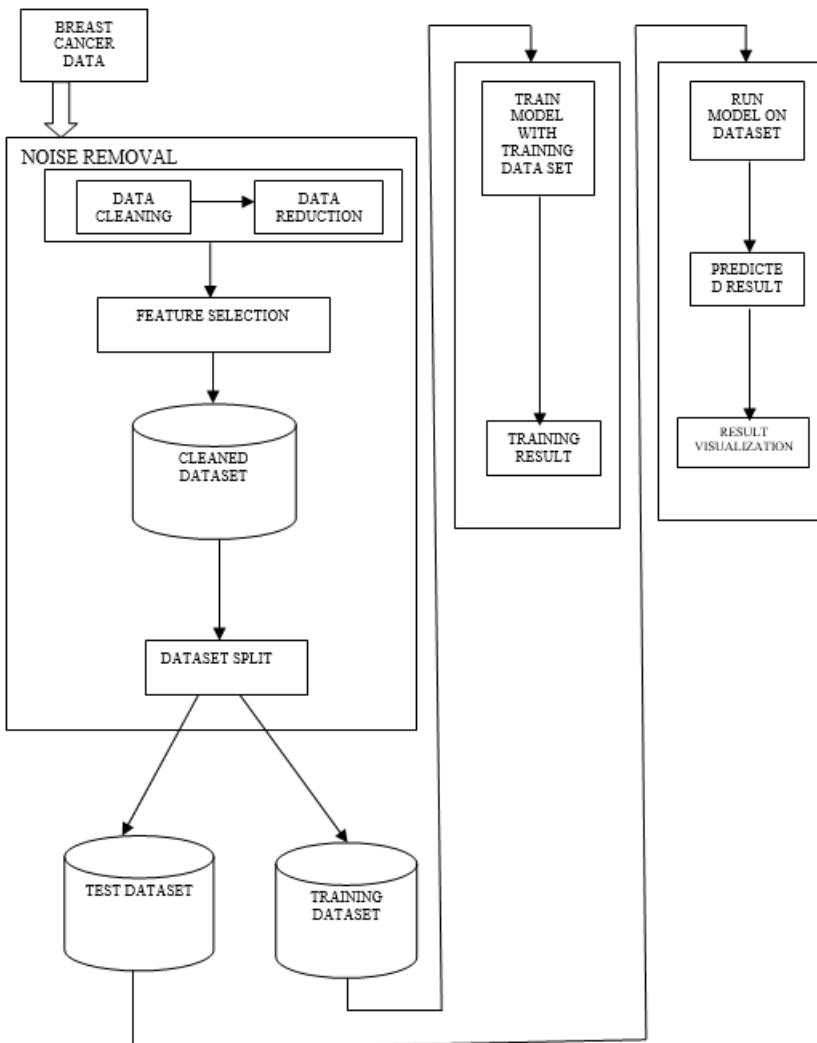
The components of the system are as follows:

Breast Cancer Data: this part represents the breast cancer dataset gotten from the TCIA data repository. An analysis of the dataset will be presented in later chapters.

Noise Removal: this part handles the removal of noise from the collected data so as to make the data good for classification and analysis.

Data discretization module: This module handles the discretization of the collected data preparing them for training and learning by the machine learning algorithms.

Figure 1. Conceptual Diagram of the System



Data discretization and Training module: This module handles the preparation of the classified data for the machine learning algorithm to analyze.

Visualization and Results: After the data has parsed to the machine learning algorithm, the final step will be to visualize them. This is a very important step in the process, as data visualization is essential for helping us gain some insight into the investigation of the problem domain.

The input into the system is a breast cancer dataset from TCIA repository denoted by 'breast cancer data' in the conceptual diagram of the system (figure 1). The Cancer Imaging Archive is a service which de-identified and hosts a large archive of medical images of cancer accessible for public download. The data are organized as "collections", typically patients imaging related by a common disease (e.g. lung cancer), image modality or type (MRI, CT, digital histopathology, etc.) or research focus. DICOM is the primary file format used by TCIA for radiology imaging (TCIA, 2021). Figure 2 shows the input stage in the system where data processing takes place.

Figure 3 shows a sample of dataset of the system.

The input to the system is an array of TCIA breast cancer dataset images with the following attributes/features for the image, returning a 30 real-valuated vector. (TCIA Repository, 2021).

Attribute Information:

ID number

Diagnosis (M = malignant, B = benign)

Thirteen real-valued features are computed for each cell nucleus:

Radius (mean of distances from center to points on the perimeter)

Figure 2. A Structural Background of the System showing the Input Design

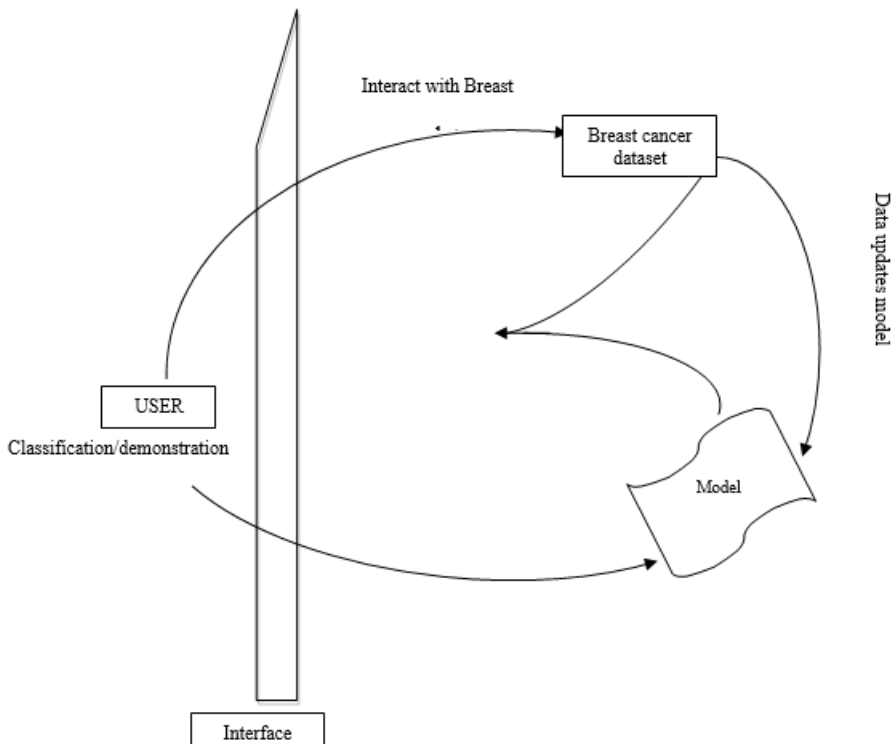
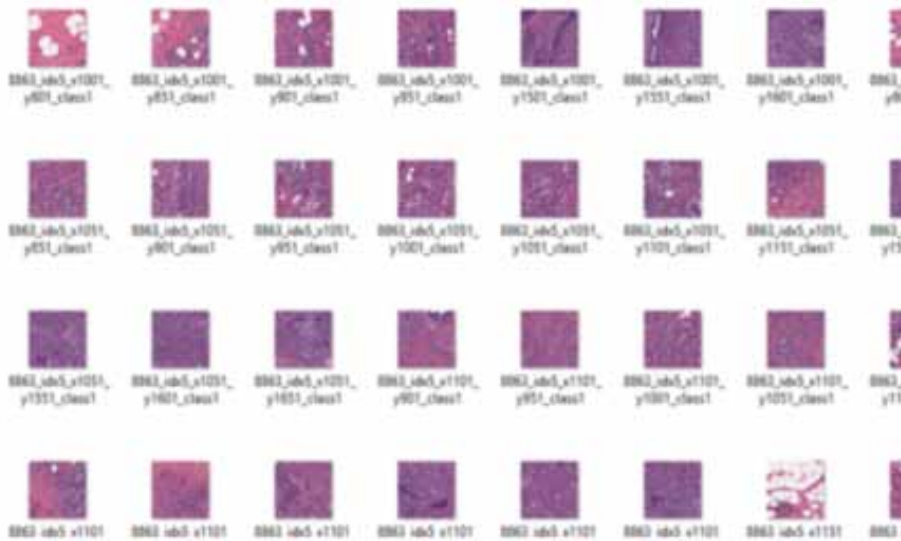


Figure 3. A Sample of the Breast Cancer Dataset (TCIA, 2021)



- Texture (standard deviation of gray-scale values)
- Perimeter (of the mass)
- Area (of the mass)
- Smoothness (local variation in radius lengths)
- Compactness ($\text{perimeter}^2 / \text{area} - 1.0$)
- Concavity (severity of concave portions of the contour)
- Concave points (number of concave portions of the contour)
- Symmetry
- Fractal dimension (“coastline approximation” — 1)
- Age of the woman
- History of breast cancer in the family
- Estrogen exposure

The mean, standard error and “worst” or largest (mean of the three largest values) of these features were computed for each image, resulting in 30 features. For instance, field 3 is Mean Radius, field 13 is Radius SE, field 23 is Worst Radius.

This analysis aims to observe which features are most helpful in predicting malignant or benign cancer and to see general trends that may aid us in model selection and hyper parameter selection. The goal is to classify whether the breast cancer is benign or malignant. To achieve this, a model is developed that is capable of fitting a function on the input value to generate an output value that is dependent of the activation function.

DATA PREPROCESSING

After identification, extraction, and cleaning the data needed for the use case, the next step is to have an understanding of that content. The use of Active Learning helped in preprocessing the dataset to a quality that is usable by the deep learning machine. This is very important because, a dataset that is clean and non-redundant is needed both in the train and test sets to allow generation.

MODEL DESIGN, TRAINING AND VALIDATION

After selecting and cleaning the dataset, the next phase in the methodology is the design of the model. The machine learning algorithm used is the Bi-LSTM and 2dConv. They are algorithms that are used on supervised data. The model is discussed in the section that follows.

AUTOENCODERS

An autoencoder is a type of artificial neural network used to learn efficient data coding in an unsupervised manner. The aim of an autoencoder was used to learn a representation (encoding) for a set of data, typically for dimensionality reduction, by training the network to ignore signal “noise.” Along with the reduction side, a reconstructing side is learnt, where the autoencoder tries to generate from the reduced encoding a representation as loose as possible to its original input, hence its name.

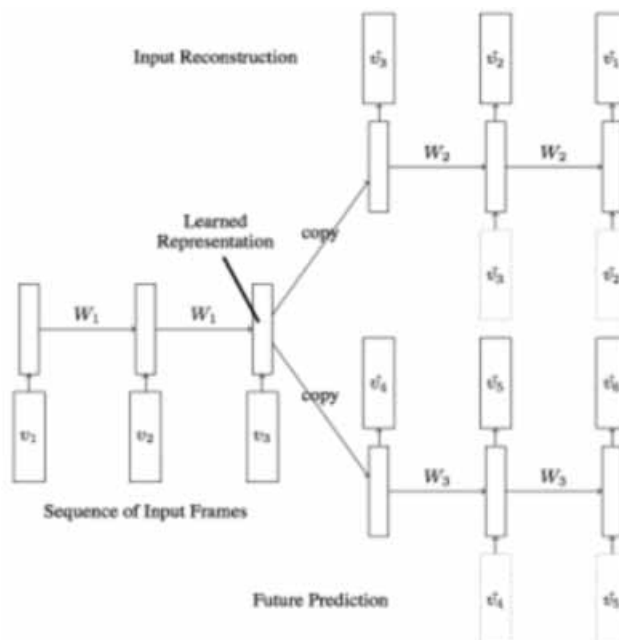
The autoencoder was used to encode and decode the data gotten from the dataset. Before the data is fed into the bidirectional long short term memory, the auto encoder takes in the data from the dataset and encodes it after which it is reconstructed and fed to the bidirectional long short term memory, as demonstrated in figure 4. It serves as a means of compressing the data without losing information and learning from the data.

It was employed in training the model. Given the amount of parameter in the dataset, this prevents overfitting the model with unnecessary data.

From figure 4, W stands for weight matrices, V stands for input vectors and \hat{V} stands for reconstructed feature space.

The autoencoder is composed of three layers: the input layer, a hidden layer using the sigmoid activation function, and the output layer. The autoencoder is trained so that the out-put layer attempts to be as similar as possible to the input layer. This way, the hidden layer results in a non-linear compact representation of the input layer. The rationale behind this transformation is that data will be more

Figure 4. Autoencoder Reconstruction (Adopted from Srivastava et.al., 2016)



compact (i.e., less prone to over rating) and hopefully some interesting non-linear relationships that improve the explanation of the output variable would be discovered.

The hidden layer of the autoencoder, the non-linear compact representation of the original input, is directly connected to a multilayer perceptron i.e the Bi-LSTM directly connected to the convolutional neural network. This convolutional network is the one responsible for making predictions in our problem, by taking the new problem representation as an input.

The dataset used contained 281562 samples each falling within two main classes: benign or malignant. In this experiment, we used 70% of the samples for training and 20% of the samples for testing and 10% for validation. To generalize the classification task to perform successfully when testing new patients, we ensure that the patients selected for training are not used during testing. Per the experimental design by Idowu, *et al*, (2021), the average accuracy was reported after successfully completing five trials.

In each dataset, different data augmentation was applied techniques including: sequential rotation by 40 degrees, width shift with factor of 0.2, height shift with factor of 0.2, shear with a factor of 0.2, zooming with a range 0.2, horizontal flipping, and vertical flipping. From Figure 4, it can be observed that noise has been added in some parts of the images. Therefore, we have also evaluated our method using only the center patch of the augmented samples. The down sampled and center patches are shown for two different input samples. In the first experiment, we trained with the proposed model using the stochastic gradient descent (SGD) optimization function. We set the momentum to 0.9 and decay is calculated based on the initial learning rate and number epochs of the respective trial. We have experimented for three trials where 5 epochs are used in each trial. After 5 epochs, the learning rate is decreased by the factor of 10.

The most commonly reported metric for evaluating metric is the accuracy. The metric can be misleading when the data are imbalanced. In such cases, other evaluation metrics should be considered in addition to accuracy. (Akosa, 2017). However, when only 2% of your dataset is of one class (malignant) and 98% some other class (benign), misclassification scores don't really make sense. It can be 98% accurate and still catch none of the malignant cases which could make a terrible classifier. To achieve an accurate prediction, the dataset will be augmented with patches to balance it for a better prediction.

RECURRENT NEURAL NETWORK (RNN)

Recurrent neural networks (RNNs) are able to process input sequences of arbitrary length via the recursive application of a transition function on a hidden state vector h_t . At each time step t , the hidden state h_t is a function of the input vector x_t that the network receives at time t and its previous hidden state h_{t-1} . For example, the input vector x_t could be a vector representation of the t -th word in body of text. The hidden state $h_t \in R^d$ can be interpreted as a d - dimensional distributed representation of the sequence of tokens observed up to time t , where R_d is D dimensional feature space. commonly, the RNN transition function is an affine transformation followed by a point-wise nonlinearity such as the hyperbolic tangent function:

$$h_t = \tanh(Wx_t + Uh_{t-1} + b) \quad (1)$$

Unfortunately, a problem with RNNs with transition functions of this form is that during training, components of the gradient vector can grow or decay exponentially over long sequences. This problem with exploding or vanishing gradients makes it difficult for the RNN model to learn long-distance correlations in a sequence. The LSTM architecture addresses this problem of learning long-term dependencies by introducing a memory cell that is able to preserve state over long periods of time.

LONG-SHORT TERM MEMORY (LSTM)

The LSTM unit is now defined at each time step t to be a collection of vectors in R^d : an input gate i_t , a forget gate f_t , an output gate o_t , a memory cell c_t and a hidden state h_t . The entries of the gating vectors i_t , f_t and o_t are in $[0,1]$. We refer to d as the memory dimension of the LSTM. The LSTM transition equations are the following

$$\begin{aligned}
 i_t &= \sigma \left(W^{(i)} x_t + U^{(i)} h_{t-1} + b^{(i)} \right) \\
 f_t &= \sigma \left(W^{(f)} x_t + U^{(f)} h_{t-1} + b^{(f)} \right) \\
 o_t &= \sigma \left(W^{(o)} x_t + U^{(o)} h_{t-1} + b^{(o)} \right) \\
 u_t &= \tanh \left(W^{(u)} x_t + U^{(u)} h_{t-1} + b^{(u)} \right) \\
 c_t &= (i_t \circ u_t) + (f_t \circ c_{t-1}) \\
 h_t &= o_t \circ \tanh(c_t)
 \end{aligned} \tag{2}$$

Where: x_t is the input at the current time step, σ denotes the logistic sigmoid function and \circ denotes element-wise multiplication.

Intuitively, the forget gate controls the extent to which the previous memory cell is forgotten, the input gate controls how much each unit is updated, and the output gate controls the exposure of the internal memory state. The hidden state vector in an LSTM unit is therefore a gated, partial view of the state of the unit's internal memory cell. Since the value of the gating variables vary for each vector element, the model can learn to represent information over multiple time scales.

A limitation of the LSTM architectures described above using equation (2) is that they only allow for strictly sequential information propagation which will limit its power in the sentiment analysis of the problem domain we are investigating. Hence, the Bi-LSTM-CRF was used for the sentiment analysis of the problem domain.

The processing task in the system reads in the dataset and pre-processes it. The pre-processed dataset is then mined for features which are extracted. The extracted features were then split into a training and test set. This was then used to build a model which was trained with the training set and finally tested and validated with the test set. The algorithm and component discussion are shown below:

In the Bi-LSTM model, we have the observation variable X whose values are observed, random variables Y whose values the task requires the model to predict, and an undirected graph G where Y are connected by undirected edges indicating dependencies. CRF defines the conditional probability of a set of output values $y \hat{Y}$ given a set of input values $x \hat{X}$ to be proportional to the product of potential functions on cliques of the graph as illustrated in Equation (3).

$$p(y|x) = \frac{1}{Z_x} \prod_{s \in S(y,x)} \varphi_s(y_s, x_s) \tag{3}$$

Where Z_x is a normalization factor overall output values, $S(y, x)$ is the set of cliques of G , $\varphi_s(y_s, x_s)$ is the clique potential on cliques.

Afterwards, in the Bi-LSTM-CRF model, a softmax over all possible tag sequences yields a probability for the sequence y . The prediction of the output sequence is computed as follows:

$$y^* = \operatorname{argmax}_{y_i} \sigma(X, y)$$

Where $\sigma(X, y)$ is the score function defined as follows:

$$\sigma(X, y) = \sum_{i=0}^n A_{y_i, y_{i+1}} + \sum_{i=0}^n P_{i, y_i} \tag{4}$$

Where A is a matrix of transition scores, $A_{y_i, y_{i+1}}$ represents the score of a transition from the tag y_i to y_{i+1} , n is the length of a sentence, P is the matrix of scores output by the Bi-LSTM network, P_{i, y_i} is the score of the y_i tag of the i th word in a sentence.

Figure 5 is a diagrammatic representation of the Bidirectional Long Short Term Memory (Bi-LSTM)

The Two Dimensional Convolved Neural Networks (Conv2D) takes as input the fallout from the Bi-LSTM of varying lengths and produces fixed-length vectors as outputs. If we view the states as the hidden representations of moving objects, the combination of the Bi-LSTM and Conv2D with a larger transitional kernel should be able to capture faster motions while one with a smaller kernel can capture slower motions. Also, if we adopt this view, the inputs, cell outputs and hidden states of the traditional FC-LSTM may also be seen as 3D tensors with the last two dimensions being 1. In this sense, FC-LSTM is actually a special case of Bi-LSTM with all features standing on a single cell. To ensure that the states have the same number of rows and same number of columns as the inputs, padding is needed before applying the convolution operation.

CNN connectionism model was used for classification, focusing on learning from environmental stimuli and storing this information in a form of connections between neurons (figure 6). The weights in a neural network are adjusted according to the training data by some learning algorithm. That is, the greater the difference in the training data, the more difficult for the learning algorithm to adapt the training data, and the worse classification results it will produce.

The major drawbacks of Bi-LSTM in handling spatio temporal data is its usage of full connections in input-to-state and state-to-state transitions in which no spatial information is encoded. To overcome this problem, a distinguishing feature of our design is that all the inputs X_1, \dots, X_t , cell outputs C_1, \dots, C_t , hidden states H_1, \dots, H_t , and gates i_t, f_t, o_t of the Bi-LSTM and Conv2D are 3D tensors whose last

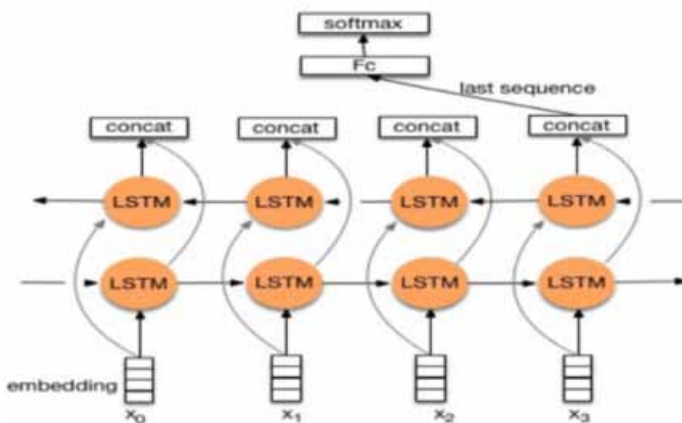
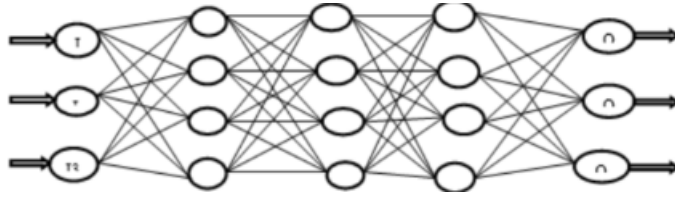


Figure 5. An illustration of Bi-LSTM-CRF. Bi-LSTM layer incorporates a forward LSTM layer and a backward LSTM layer (Adopted from Lee et al., 2010)

Figure 6. Convolutional Neural Networks



two dimensions are spatial dimensions (rows and columns). To get a better picture of the inputs and states, we may imagine them as vectors standing on a spatial grid. The Bi-LSTM and Conv2D determine the future state of a certain cell in the grid by the inputs and past states of its local neighbors. This can easily be achieved by using a convolution operator in the state-to-state and input-to-state transitions. The key equations of the Bi-LSTM and Conv2D are shown below as Equation (5):

$$\begin{aligned}
 &= \sigma \left(W^{(i)} x_t + U^{(i)} h_{t-1} + b^{(i)} \right) \\
 f_t &= \sigma \left(W^{(f)} x_t + U^{(f)} h_{t-1} + b^{(f)} \right) \\
 o_t &= \sigma \left(W^{(o)} x_t + U^{(o)} h_{t-1} + b^{(o)} \right) \\
 u_t &= \tanh \left(W^{(u)} x_t + U^{(u)} h_{t-1} + b^{(u)} \right) \\
 c_t &= (i_t \circ u_t) + (f_t \circ c_{t-1}) \\
 h_t &= o_t \circ \tanh(c_t)
 \end{aligned} \tag{5}$$

With the above equation in conjunction with the Conv2D, we formulate breast cancer prediction as a linearly separated parametric function problem that can be solved under the general sequence-to-sequence learning framework. In order to model well the spatiotemporal relationships, we extend the idea of FC-LSTM to Bi-LSTM and Conv2D which has convolutional structures in both the input-to-state and state-to-state transitions. By stacking multiple Bi-LSTM and Conv2D layers and forming an encoding-forecasting structure, we build an end-to-end trainable model for breast cancer prediction

This will allow the two dimensional convolutional neural network to analyze and classify the data to make a prediction. For evaluation. The Cancer Imaging Archive Repository dataset is used which can facilitate further research especially on devising machine learning algorithms for the problem.

To create a Bi-LSTM network for sequence-to-label classification:

First create a layer array containing a sequence input layer, a Bi-LSTM layer, a fully connected layer, a softmax layer, and a classification output layer, then specify the size of the sequence input layer to be the number of features of the input data. Lastly, specify the size of the fully connected layer to be the number of classes. We do not need to specify the sequence length. And,

For the Bi-LSTM layer, we specify the number of hidden units and the output mode 'last'. To train the one dimensional neural network to classify sequence data, we use the output from the Bi-LSTM network. The Bi-LSTM network enables us to input sequence data into the two dimensional neural network, and make predictions based on the individual time steps of the sequence data. Next, we load the training data from the output of the Bi-LSTM. Let's assume an array variable called XTrain. XTrain is a cell array containing say N sequences of dimension 32. Y is a categorical vector of labels, which correspond to the 32 features. The entries in XTrain are matrices with 32 rows (one row for each feature) and varying number of columns (one column for each time step).

$$\begin{bmatrix} Y_{11}, Y_{12} & \dots & Y_{1n} \\ \vdots & \ddots & \vdots \\ Y_{321}, Y_{322} & \dots & Y_{32n} \end{bmatrix}$$

The flowchart of the model is given in figure 7

The system algorithm is given as follows:

```

ALGORITHM: Breast_Cancer_Prediction
1. Get dataset from source; // split data = pdata
2. Set pdata ← Pre_processed_dataset
3. Test dataset
4. Train dataset
5. Set Trained_dataset ← pdata
6. Model = Build(Bi-LSTM, Conv2D)
7. Model_Train; //(Train_data)
8. Set Pred ← Model_Predict; //(Test_data)
9. Evaluate Pred
10. Generate report
11. End
    
```

After the data analysis using the Bidirectional-Long Short Term Memory algorithm and the Two-dimensional Convolutional Neural Network, the next step is to present the aggregated analyzed breast cancer prediction in visual components like tables and graphs where necessary. This part of the proposed solution is very important as it allows the researcher compare results of this research work with other results in this problem domain.

This module is responsible for the extraction and aggregation of the disease predicted by the Bi-LSTM-CRF and 2d-CNN algorithm. The predicted breast cancer extractor extracts predicted breast cancer made by the Bi-LSTM-CRF and 2d-CNN algorithm while the predicted breast cancer aggregator aggregates the predicted precipitation extracted by the predicted breast cancer extractor module. This way, the aggregated predicted breast cancer that has been extracted can be presented in a coherent manner by the result presentation module. The collection of these three (3) modules ensures that the predictions made by the Bi-LSTM-CRF and 2d-CNN algorithm are presented in a human readable form.

Figure 7. The Flowchart of the System



RESULTS

A screenshot of the program running on Python is shown in figure 8.

The dataset used contains 281562 samples each falling within two main classes: benign and malignant. Figures 9 (a and b) show the screen shot of the confusion matrix of the model. Confusion Matrix is a very important metric when analyzing misclassification. Each row of the matrix represents the instances in a predicted class while each column represents the instances in an actual class. The diagonals represent the classes that have been correctly classified. This helps to show which classes are being misclassified but also what they are being misclassified as.

The result of the Precision, F-Score and Recall is given in table 1 below:

As shown in the Table 1, we pick the best results for each of the execution of the model. This includes First run, Second run Third run, Fourth run and Final run.

EVALUATION OF THE SOLUTION

Before beginning the evaluation of the result, the research work already done on this problem domain with respect to the solution is to be considered. This is discussed below:

Oluwashola (2021) conducted a similar study and obtained an accuracy of 98.3% after running 500 epochs on 909 samples. In this study, we ran 600 epochs on samples and got an accuracy of 98.8%, an improvement of 0.3% on a sample less than the previous by 9. The epoch refers to the number of test instances. Accuracy or reliability of results ordinarily increases with the number of tests done. The implication of the findings of this study is that more accuracy has been recorded in a lesser time and with fewer instances than that recorded in the method adopted by Oluwashola (2021). This approach will thus improve the performance of breast cancer prediction in a shorter time interval that existing systems.

This implies the accuracy of the model depends greatly on the number of epochs run. This means that the accuracy of the prediction increases when the learns over a longer duration, even if the number of samples is slightly reduced.

Figure 8. The PyCharm Program Interface

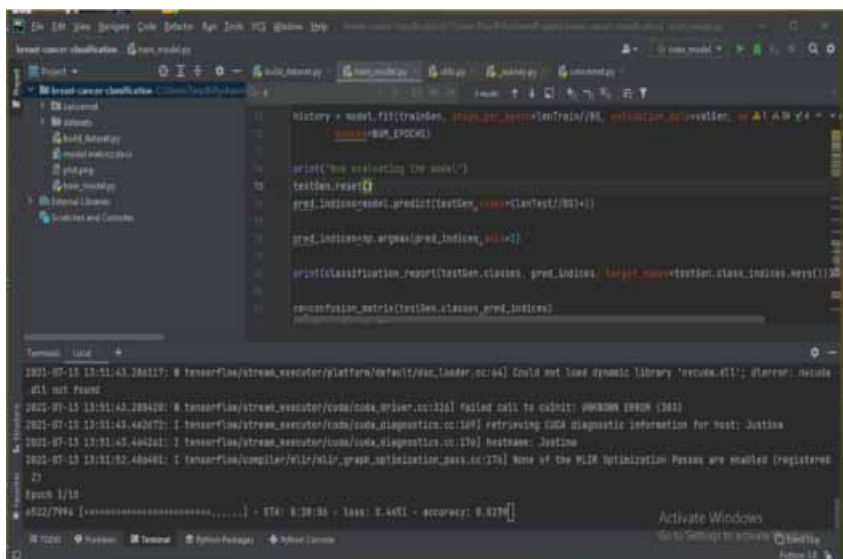


Figure 9. Confusion matrix for the predictions (a) benign (b) malignant

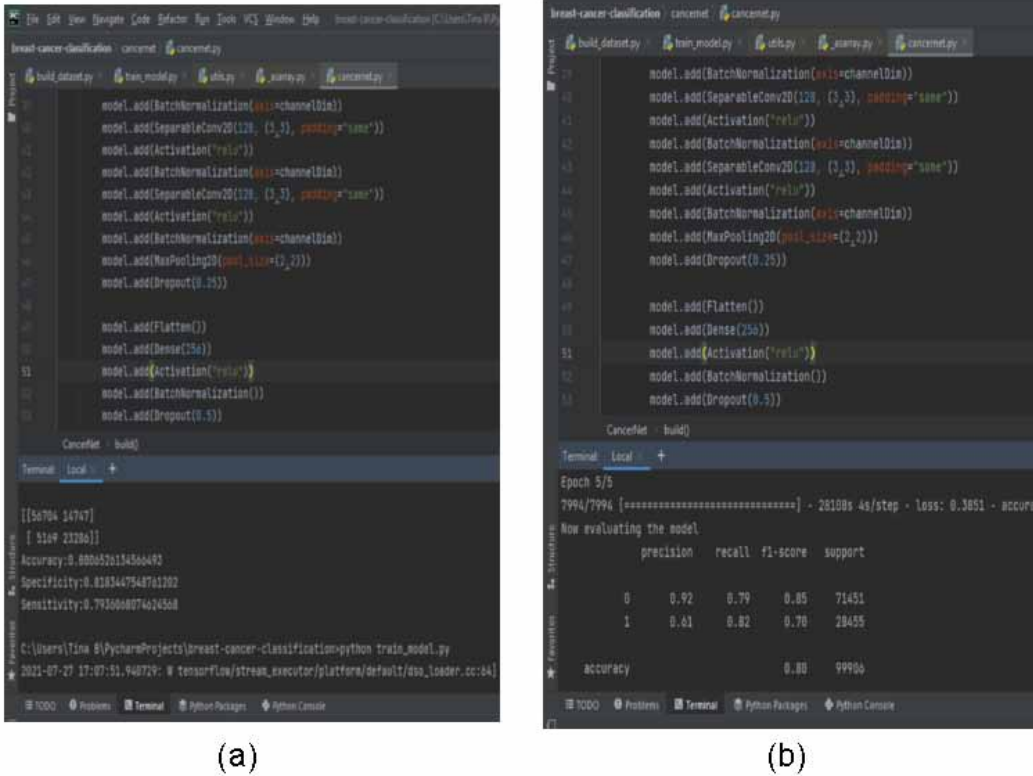


Table 1. Result of Precision, F - Score and Recall for the solution

Model execution number	Accuracy	Precision	Recall	F1 score	ROC-AUC
1	98.8	0.61	0.82	0.77	0.79
2	98.6	0.58	0.76	0.70	0.63
3	98.5	0.55	0.73	0.68	0.61
4	98.4	0.53	0.71	0.63	0.58

The result may have been influenced by a number of reasons such as the dataset and the number of epochs the system was programmed with, the preprocessing method, and the model used for analysis. But it is pertinent to present the fact that the use of bidirectional long short term memory to actively learn the dataset played an important role in the outcome of the results.

CONCLUSION

In this study, an Active Learning (AL) method on breast cancer dataset was designed using a bidirectional long short term memory and a two dimensional convolutional neural network model. The experiments were conducted using the model on the TCIA breast cancer dataset, and performance was evaluated using different performance metrics. The performance of the method was evaluated via image level, patient level; images based, and patch based analysis. The criteria considered were

different criteria (features) such as magnification factor, resized sample inputs, augmented patches and samples, radius (mean of distances from center to points on the perimeter), texture (standard deviation of gray-scale values), perimeter, area, smoothness (local variation in radius lengths), compactness ($\text{perimeter}^2 / \text{area} - 1.0$), concavity (severity of concave portions of the contour), concave points (number of concave portions of the contour), symmetry and fractal dimension (“coastline approximation” — 1) in this implementation.

This model had a good performance with an accuracy of 98.8%, specificity of 0.82 and a sensitivity of 0.79 after running on 600 epochs with 194,330 benign breast cancer images and 84,232 malignant breast cancer images.

This model will improve the performance of breast cancer prediction assuring the breast cancer patients get the timely medical treatment they need thereby reducing mortality rate. It is also recommended that this model be applied on a dataset obtained from breast cancer patients in real time.

REFERENCES

- Ali, A. B. (2019). A Comparative Analysis of Nonlinear Machine Learning Algorithms for Breast Cancer Detection. *International Journal of Machine Learning and Computing*, 3(9), 248–254. <http://www.ijmlc.org/vol9/794-L0201.pdf>
- Aruna, S., Rajagopalam, S., & Nandakishore, L. (2011). *CCSEA 2011. Knowledge Based Analysis of Various Statistical Tools in Detecting Breast Cancer*. In D. C. Wyld (Eds.), *Computer Science & Information Technology CS & IT 02* (pp. 37–45). doi:10.5121/csit.2011.1205
- Bogdan, K., Michał, K., Andrzej, B., Zbigniew, A., Bogusław, O., Paweł, W., Jakub, S., & Bogusław, C. (2019). Breast Cancer Classification on Histopathological Images Affected by Data Imbalance Using Active Learning and Deep Convolutional Neural Network. *ACM Digital Library: Proceedings of the Artificial Neural Networks and Machine Learning – ICANN 2019: Workshop and Special Sessions: 28th International Conference on Artificial Neural Networks*, 299–312. doi:10.1007/978-3-030-30493-5_31
- Chaurasia, V. & Pal, S. (2014). Data mining Techniques: To Predict and Resolve Breast Cancer Survivability. *International Journal of Computer Science and Mobile Computing*, 3(1), 10-22.
- Christobel, A., & Sivaprakasam, P. (2011). An Empirical Comparison of Data Mining Classification Methods. *International Journal of Engineering and Advanced*, 3(2), 24–28.
- Ciresan, D., Mierer, U., Masci, J., Gambardella, M., & Schmidhuber, J. (2011). Flexible High Performance Convolutional Neural Networks for Image Classification. *Artificial Intelligence*, 1, 1237–1242.
- Egenti, E. (2016). *Breast Cancer in Nigeria: Diagnosis, Management and Challenges*. Author House.
- Giu, C., & Jyh-Cheng, Y. (2015). American College of Radiology Breast Imaging Reporting and Data System Ultrasound: An Overview of the second edition. *Journal of Medical Ultrasound*, 23, 117-119.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- He, H., & Garcia, E. (2009). Learning from Imbalanced Data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263–1284. doi:10.1109/TKDE.2008.239
- Hiba, A., Hajar, M., Hassan, M., & Thomas, N. (2016). Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis. *Procedia Computer Science*, (83), 1064–1069.
- Huang, G., Lui, Z., Van der Maaten, I., & Weigberger, K. (2017). Densely Connected Neural Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4700-4708.
- Karabatak, M. (2015). A new classifier for breast cancer detection based on Naive Bayesian. *Measurement*, 72, 32–36. doi:10.1016/j.measurement.2015.04.028
- Karpathy, A. (2016). *CS231n Convolutional neural networks for Visual recognition. Course Materials and Notes for Stanford Class in CS231n*. <http://cs231n.stanford.edu/2016/>
- Lee, C. H., Dershaw, D. D., Kopans, D., Evans, P., Monsees, B., Monticciolo, D., Brenner, R. J., Bassett, L., Berg, W., Feig, S., Hendrick, E., Mendelson, E., D’Orsi, C., Sickles, E., & Burhenne, L. W. (2010). Breast cancer screening with Imaging: Recommendations from the Society of Breast Imaging and the ACR on the use of Mammography, Breast MRI, Breast Ultrasound, and other Technologies for the Detection of Clinically Occult Breast Cancer. *Journal of the American College of Radiology*, 7(1), 18–27. doi:10.1016/j.jacr.2009.09.022 PMID:20129267
- Lorena, C. A., Jacinto, F. O., Siqueira, F. M., DeGiovanni, R., Lohmann, L. D., de Carvalho, C. P. L. F., & Yamamoto, M. (2011). Comparing Machine Learning Classifiers in Potential Distribution Modeling. *Expert Systems with Applications*, 38(5), 5268–5275. doi:10.1016/j.eswa.2010.10.031
- Oluwashola, D. A. (2021). Prediction of Breast Cancer Images Classification Using Bidirectional Long Short Term Memory and Two-Dimensional Convolutional Neural Network. *Transactions on Networks and Communications*, 9(4), 29–38. doi:10.14738/tnc.94.10663
- Patrizia, F., Fabio, M. Z., Scarpato, N., Guadagni, F., & Roselli, M. (2019). Breast Cancer Prognosis Using a Machine Learning Approach. *Cancers*, 11(3), 328. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6468737/>

- Sameer, S., Yudhveer, S., & Basant, A. (2020). Boundary Detection to Segment the Pectoral Muscle from Digital Mammograms Images. *International Journal of Engineering and Advanced Technology*, 9(3), 1593–1603. doi:10.35940/ijeat.C5423.029320
- Saria, E., & Huda, K. (2019). Breast cancer diagnosis and prediction using Machine learning and Data Mining Techniques. *IOSR Journal of Dental and Medical Sciences*, 18(4), 85–94.
- Schat, E., van de Schoot, R., Kouw, W. M., Veen, D., & Mendrik, A. M. (2020). The Data Representativeness Criterion: Predicting the performance of supervised classification based on data set similarity. *PLoS One*, 15(8), e0237009. doi:10.1371/journal.pone.0237009 PMID:32780738
- Siegel, R., Miller, D. K., & Jemal, A. (2016). Cancer statistics, 2016. *CA: a Cancer Journal for Clinicians*, 66(1), 7–30. doi:10.3322/caac.21332 PMID:26742998
- The Cancer Imaging Archive (CIA). (2021). *Cancer Imaging Archive*. <https://www.cancerimagingarchive.net/about-thecancer-imaging.archive.tcia/>
- Ursuleanu, T. F., Luca, R. A., Gheorghe, L., Grigorovici, R., Iancu, S., Hlusneal, M., Preda, C., & Grigorovici, A. (2021). Deep learning Application for Analyzing of Constituents and their Correlations in the Interpretations of Medieval Images. *Diagnostics (Basel)*, 11(8), 1373. doi:10.3390/diagnostics11081373 PMID:34441307
- Vizcarra, J., Place, R., Tong, L., Gutman, D., & Wang, M. (2019). Fusion in Breast Cancer Histology Classification. *BCB 19: Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, 485-493. doi:10.1145/3307339.3342166
- Wang, W., & Gao, X. (2019). Deep Learning in Bioinformatics. *Methods. PubMed*, 166, 1–3. doi:10.1016/j.ymeth.2019.06.006 PMID:31181259
- Wembin, Y., Zidong, W., Hongwei, C., Annette, P., & Xiaohui, L. (2018). Machine Learning and Applications in Breast Cancer Diagnosis and Prognosis. *Designs*, 2(2), 13. doi:10.3390/designs2020013
- World health Organisation (WHO). (2014). *WHO position paper on mammography Screening*. World Health Organisation.
- Yao, H., Zhang, X., Zhou, X., & Liu, S. (2019). Parallel Structure Deep Neural Network Using CNN and RNN with an Attention Mechanism for Breast Cancer Histology Image Classification. *Cancers (Basel)*, 11(12), 1901. doi:10.3390/cancers11121901 PMID:31795390

Moses Adah Agana holds a Bachelor of Science Degree in Mathematics/Computer Science, a Master's Degree in Computer Science, and a Ph.D. in Computer Science (Cyber Security). He is an astute scholar who has been teaching, mentoring, and supervising Computer Science students at various tertiary institutions since 1998, with several publications in both local and international journals. He has taught in the then Polytechnic Calabar (1998-1999), Federal College of Education Obudu (1999-2012), Federal University Lafia (2012), Federal University of Agriculture Makurdi (2012-2017), University of the Free State, South Africa, as a Postdoctoral Research Fellow (2017-2018), and in the University of Calabar (2017 till now). He has been a Head of Department at several instances and is the current Deputy Director, ICT in the University of Calabar. His research interest is predominantly in cyber security, Intelligent Systems and Internet of Things (IoT).