INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH & DEVELOPMENT

ISSN 2278 – 0211 (Online)

A Two-Stage Model for Breast Cancer Detection and Classification from Mammograms

Dr. Adebanjo Adekiigbe
Chief Lecturer, Department of Computer Science,
The Federal Polytechnic, Ede, Nigeria
Dr. Temilola Morufat Adepoju
Senior Lecturer, Department of Computer Engineering Technology,
The Federal Polytechnic, Ede, Nigeria
Matthias Omotayo Oladele
Lecturer, Department of Computer Engineering Technology,
The Federal Polytechnic, Ede, Nigeria

Abstract:
Breast cancer is a type of disease which is common to women. It is an abnormal behaviour which hinders the functionality of the normal breast cells which leads to forming a tumor in the form of lump. Early detection of breast cancers tends to reduce death risk and provides a chance of a suitable and better treatment. A two-stage classification model was proposed to detect and classify breast cancer using two publicly available datasets namely MIAS and the DDSM dataset. The first classification stage uses Convolution Neural Network (CNN) to classify the mammogram image as either normal or abnormal while K-Nearest Neighbour (KNN) was used for the second classification stage to classify the abnormal image into benign or malignant. The features used for the second stage classification were extracted using Gray Level Co-occurrence Matrix (GLCM), which was later used for classification. The performance was evaluated based on accuracy, sensitivity, specificity, precision, False Positive Rate (FPR), F1 score and Matthews Correlation Coefficient (MCC). The results show that CNN obtained the highest accuracy, sensitivity, specificity, precision, FPR, F1 score and MCC of 99.03%, 0.9831, 1.0000, 1.0000, 0.0000, 0.9915 and 0.9804 respectively, while KNN has an accuracy, sensitivity, specificity, precision, FPR, F1 score and MCC of 76.27%, 0.7667, 0.7586, 0.7667, 0.2414, 0.7667 and 0.5253 respectively.

Keywords: *Breast Cancer, Convolution Neural Network, K-Nearest Neighbour, Mammogram, Benign, Malignant*

1. Introduction

Breast cancer is one of the major health challenges that is common to women. If detected early, it can reduce the death rate; according to the statistics, 96% of cancers can be cured when detected early. According to the prediction of the World Health Organization (WHO), breast cancer cases will hit 19.3 million by 2025 (Ragab, Sharkas, & Attallah, 2019). Breast cancer is defined as the abnormal growth of certain cells in the breast (Zhou *et al*., 2020). It appears in the form of tumors in the breast. Cancer tumor spreads across the surrounding tissues in the breast. Benign and malignant are the two classes of tumor (Saber, Sakr, Abo-Seida, Keshk, & Chen, 2021). A benign tumor is made up of non-cancerous cells that grow only in one location and do not spread throughout the body, whereas a malignant tumor is made up of cancerous cells that are capable of spreading faster and faster to various body parts and infecting the tissues. Breast cancer has a high mortality rate of around 17% (Tsochatzidis, Costaridou, & Pratikakis, 2019).

Mammography is a common method for detecting breast cancer. It helps in the early detection of breast cancer (Ragab *et al*., 2019). The mammogram is an x-ray image of the breast that is used for regular early identification and intervention of breast cancer in women (Viegas, Domingues, & Mendes, 2021). The basic views of mammogram image when scanned from various angles are the Medio-lateral Oblique view (MLO) and Craniocaudal view (CC). Masses, clusters of micro-calcifications (MCs), and architectural distortions are some of the signs needed to accurately detect breast cancer in mammographic images (Ragab *et al*., 2019). Mammogram images analysis is a difficult task for medical personnel including radiologists; hence there is a need to develop a Computer Aided Detection (CAD) method for mammogram image analysis.

Computer Aided Detection (CAD) is a machine learning method used to detect, diagnose and classify breast tumors using the mammogram images. It can be used to assist radiologists and medical experts in proper diagnosis and classification of tumors (Ragab et al., 2019). CAD can be used to achieve better diagnostic accuracy and decrease the

number of false positives and false negatives (Wang *et al*., 2020). The components of CAD include image preprocessing, image segmentation, feature extraction and classification. The CAD system helps to develop a genuine, reliable and accurate system to be used in the detection and classification of mammographic images into normal, benign and malignant.

Machine learning is now an important area of medical image research. The advancement has led to many intelligent systems in medical image analysis and computer aided diagnosis (Gardezi, Elazab, Lei, & Wang, 2019). Its focus is on building computer systems that can be enhanced automatically through knowledge (Jordan & Mitchell, 2015). Machine learning can be used to analyze data and extract the important aspects and areas that are related to the data by creating a computational model which describes the data (Maity & Das, 2017). Machine learning applications range from data mining, natural language processing, image processing, expert system, medical imaging analysis and so on.

Machine learning algorithms can be categorized into two groups according to how they learn, they are supervised and unsupervised (Yue, Wang, Chen, Payne, & Liu, 2018). In supervised learning, the process of learning is done from the training data. The training data must be labeled before the model can learn to classify the output based on the input data. Unsupervised learning uses an unlabeled training data. The K-Nearest Neighbour (KNN) is an example of the supervised machine learning. It is a non-parametric lazy learning algorithm which is used in classifying data or images using their nearest neighbours (Yue et al., 2018). KNN requires no training phase with the training data.

Deep learning algorithm is a technology which exhibited the performance better than the various conventional machine learning methods. Deep learning methods learn the proper feature extraction process from the input data in regarding the target output, as opposed to machine learning methods, which perform feature extraction by using feature extraction methods including principal component analysis (PCA), local binary pattern (LBP), and so on (Tsochatzidis *et al*., 2019). Classifications of deep learning are unsupervised, semi-supervised and supervised learning (Alzubaidi *et al*., 2021). Deep supervised learning uses labelled data and they are recurrent neural networks (RNNs), convolutional neural networks (CNNs), and deep neural networks (DNNs).

One of the most common models used in deep learning is RNN. It has a memory which is used to record past information. It uses the previous output to predict the next output making it suitable for sequential data i.e. time series. Different fields which include computer vision, speech processing, Face Recognition and so on have used CNNs significantly. CNNs were influenced by biological neurons and brains of animals, which have a similar structure to that of a traditional neural network. It has several convolution layers, sub-sampling or pooling layers and fully connected (FC) layers. The main benefit of this method is its ability in collecting data or generating data output from experience. The disadvantage is that the decision boundary might be overstretched when training set lacks samples required in a class. Generally, CNN is simple, when compared with other methods as it has a high performance while learning. Semi-supervised learning consists of both labeled and unlabeled data. Training is done with both the labeled and unlabeled data in semi-supervised learning which makes it more accurate than supervised learning (van Engelen & Hoos, 2019; Yang, Song, King, & Xu, 2021). In unsupervised learning, the datasets are not labeled. Learning process is through existing patterns which help to discover the unknown patterns in the input data, or through the important functions.

CNN is a feed-forward, deep neural network. The features are learned directly from the image dataset through a training process. It is made up of the input, hidden and output layers. The input layer is the image; hidden layer consists of the convolution, pooling, flattening and fully connected layers (Davoudi & Thulasiraman, 2021). The convolution layer converts the input image to a feature map by performing a linear operation known as convolution. Pooling layer performs feature extraction and reduces the number of features. The feature map matrix is converted into a single vector with the flattening layer and the converted feature vector serves as input to the fully connected layer. The fully connected hidden layer connects the input vector to the output layer. The output layer is the predicted output. Figure 1 shows the CNN model.

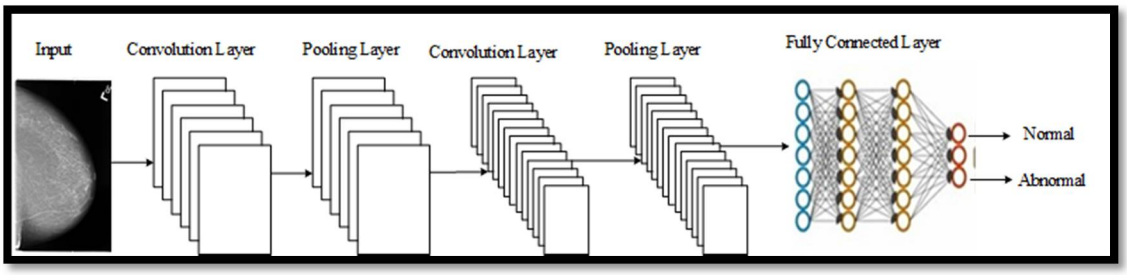

Figure 1: CNN Model

This research performs a two-stage classification using K-Nearest Neighbour (KNN) and Convolution Neural Network (CNN) to classify mammogram images first into Normal and Abnormal using KNN and classify the abnormal image into either benign or malignant using CNN. The contributions of this work are to improve the performance of two machine learning classifiers (CNN and KNN) to detect and classify breast cancer. The performance metrics used are accuracy, sensitivity, specificity, precision, FPR, F1 score and MCC.

This paper is organized as follows: Section 2 discusses the related works which highlight the proposed model for the detection and classification of breast cancer. Section 3 presents the materials and methods. Section 4 discusses the results. Finally, section 5 concludes the paper.

2. Related Works

The evolution of medical research has opened a way for various systems to detect and classify breast cancer. A lot of these algorithms produced good classification results but there is still a need to make the performance of new models better than that of the previous systems. Hence, this research aims at improving the accuracy of CAD system for breast cancer using two-stage classification system.

In the work of Azar & El-Metwally (2013), a decision support tool was used to detect breast cancer. Single decision tree (SDT), boosted decision tree (BDT) and decision tree forest (DTF) were used. Training and testing were done with the features from Wisconsin breast cancer dataset. The performance metrics used were accuracy, sensitivity, specificity, confusion matrix and receiver operating characteristic (ROC) curves. From the results obtained, SDT has a training accuracy of 97.07% with 429 samples classified correctly, while BDT produced a training accuracy of 98.83% given a total of 437 correct classifications. The ROC and Matthews correlation coefficient (MCC) for BDT was 0.99971 and 0.9746 in the training phase respectively, which was better than that of SDT. Considering DTF, an accuracy of 97.51% was obtained from the validation stage when compared with that of SDT and BDT which produced 95.75% and 97.07% respectively. ROC and MCC of 0.99382 and 0.9462 respectively were obtained for DTF.

A CAD system to detect and categorize breast tissue using BI-RADS was developed by Adepoju, Ojo, Omidiora, Olabiyisi & Bello (2015). A two-stage approach was used to categorize the tissue in the breast image into low dense i.e. fatty and high dense. The tumors in the fatty region were then segmented and later classified into normal, benign or malignant. An accuracy of 90.65% was obtained. Sakri, Rashid & Zain (2018), in their study, compared the accuracy of some researches on data mining algorithms to predict the recurrence of breast cancer. Feature selection used was the Particle Swarm Optimization (PSO) along with three classifiers which are Naïve Bayes (NB), K-Nearest Neighbour (KNN) and reduced error pruning (REP) tree. Classifications were done with and without feature selection on the Wisconsin Breast Cancer Dataset (WBCD) and the results were compared with each other. With feature selection, NB, KNN and REP produced an accuracy of 70%, 76.3%, and 66.3% respectively, while an accuracy of 81.3%, 80%, and 75% was recorded for NB, KNN and REP tree respectively when PSO feature selection was used.

In the work of Tosin *et al*., (2018), a curvelet Transform (CT)-Local Binary Pattern (LBP) feature extraction technique was proposed to detect and classify mass in digital mammogram. The MIAS dataset was used for the research and an accuracy of 94.17% was achieved. A comprehensive review on SVM, KNN, ANN and decision tree was done by Yue et al., (2018) to predict breast cancer using the Wisconsin Breast Cancer Diagnosis (WBCD) dataset. Deep Belief Networks was used with ANN (DBN-ANN) and an accuracy of 99.68% was obtained. SVM with a two-step clustering algorithm was used with an accuracy of 99.10%. Ensemble technique with SVM, Naïve Bayes and the voting method of J48 was also used and an accuracy of 97.13% was obtained.

Ed-daoudy & Maalmi (2020) proposed a breast cancer classification system with reduced feature set using association rules and support vector machine (SVM). The approach has two stages. The first stage eliminates the insignificant features using the Association Rules (AR), while the second stage uses some set of classifiers to distinguish the incoming tumours. The AR reduces the feature space dimension from nine to eight and four attributes. The Wisconsin Breast Cancer Diagnostic (WBCD) dataset was used to evaluate the performance of the model using a threefold cross-validation method. The results showed that Support Vector Machine (SVM) model with AR achieves the highest classification accuracy of 98.00% for eight attributes and 96.14% for four attributes.

A comparative study of breast cancer prediction was proposed by Islam *et al*., (2020). Five different supervised machine learning techniques were used; they are Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Random Forests, Artificial Neural Network (ANN) and Logistic Regression. The Wisconsin Breast Cancer dataset was and the performance was evaluated based on accuracy, sensitivity, specificity, precision, negative predictive value, false positive rate, false negative rate, F1 curve and Receiver Operating Characteristic (ROC) curve. The results obtained showed that ANN has the highest accuracy, precision and F1 score of 98.57%, 97.82% and 0.9890 respectively, while SVM produced an accuracy, precision and F1 score of 97.14%, 95.65% and 0.9777 respectively.

3. Materials and Methods

The flowchart of the developed model is shown in Figure 2. The mammogram image was acquired and later preprocessed. The processed image is then segmented such that the region of interest (ROI) was extracted. The segmented image was later fed into the CNN model for feature extraction and Classification. The CNN classifies the image into either normal or abnormal for the first stage classification. The second stage classification was done by the KNN model. This stage classifies the abnormal image into benign or malignant. Feature extraction was done by GLCM and the image was classified using KNN.

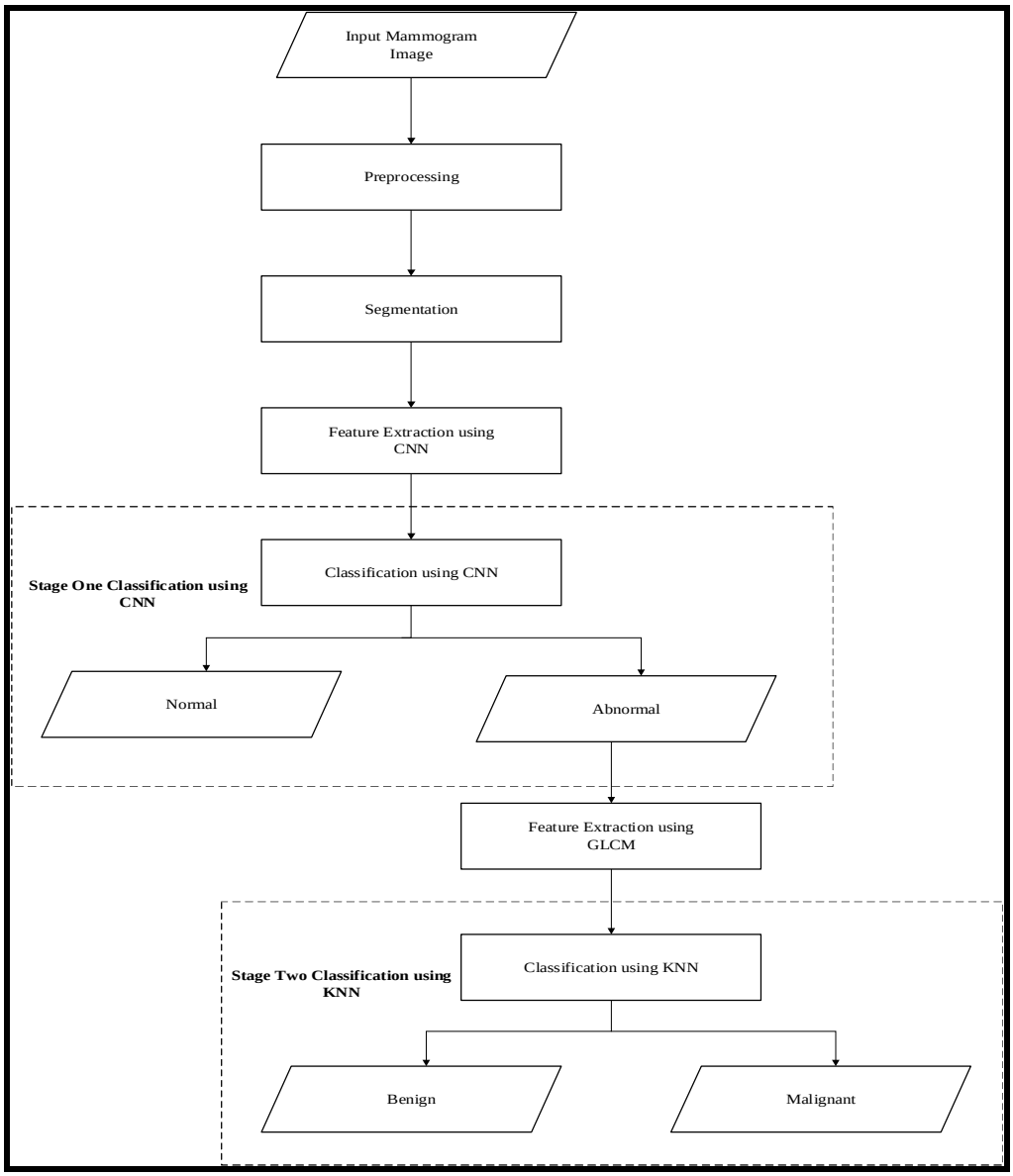

Figure 2: Flowchart of the Developed Model

3.1. Description of Dataset

Several databases for breast cancer datasets, such as, Mammography Image Analysis Society (MIAS), Wisconsin Breast Cancer Dataset (WBCD), Digital Database for Screening Mammography (DDSM), Breast Cancer Histopathology (BreakHis) and Breast Cancer Histology (BACH), are publicly available. Two of these datasets were used in this work i.e. the MIAS and DDSM datasets. The DDSM dataset was used to train the CNN model, while the MIAS dataset was used to test the developed model.

The MIAS database has a total of 322 (161 pairs) digitized MLO images at 50-micron resolution in "Portable Gray Map" (PGM) format. The dataset includes benign and malignant lesions and regular images. The MIAS dataset is labelled from column 1 to 7. The 1st column is the reference number of the MIAS database; the 2nd column is the background tissue character which could be F (Fatty), G (Fatty-glandular) and D (Dense-glandular); the 3rd column is the class of abnormality present i.e. CALC (Calcification), CIRC (Well-defined/circumscribed masses), SPIC (Speculated masses), MISC (Other, ill-defined masses), ARCH (Architectural distortion), ASYM (Asymmetry) and NORM (Normal).

The 4th column is the severity of abnormality which could be B (Benign) or M (Malignant). The 5th and 6th columns contain the *x, y* image-coordinates of centre of abnormality; the 7th column is the approximate radius (in pixels) of a circle enclosing the abnormality. The images are all 1024 by 1024 pixels in size. The example of the label is mdb001 G CIRC B 535 425 197.

The DDSM is the largest public dataset, which has 2,620 images containing two images from each breast, namely - the medio-lateral oblique (MLO) and craniocaudal (CC), and a total of 10,480 images containing all types of findings ranging from regular images to images containing benign and malignant lesions. The images have ROIs for calcifications and masses, as well as useful information for CADe and CADx algorithms, such as Breast Imaging Reporting and Data System (BI-RADS) attributes for mass shape, mass margin, calcification type, calcification distribution, and breast density. Figure 3 shows the sample images from the MIAS and DDSM dataset respectively for normal, benign and malignant.

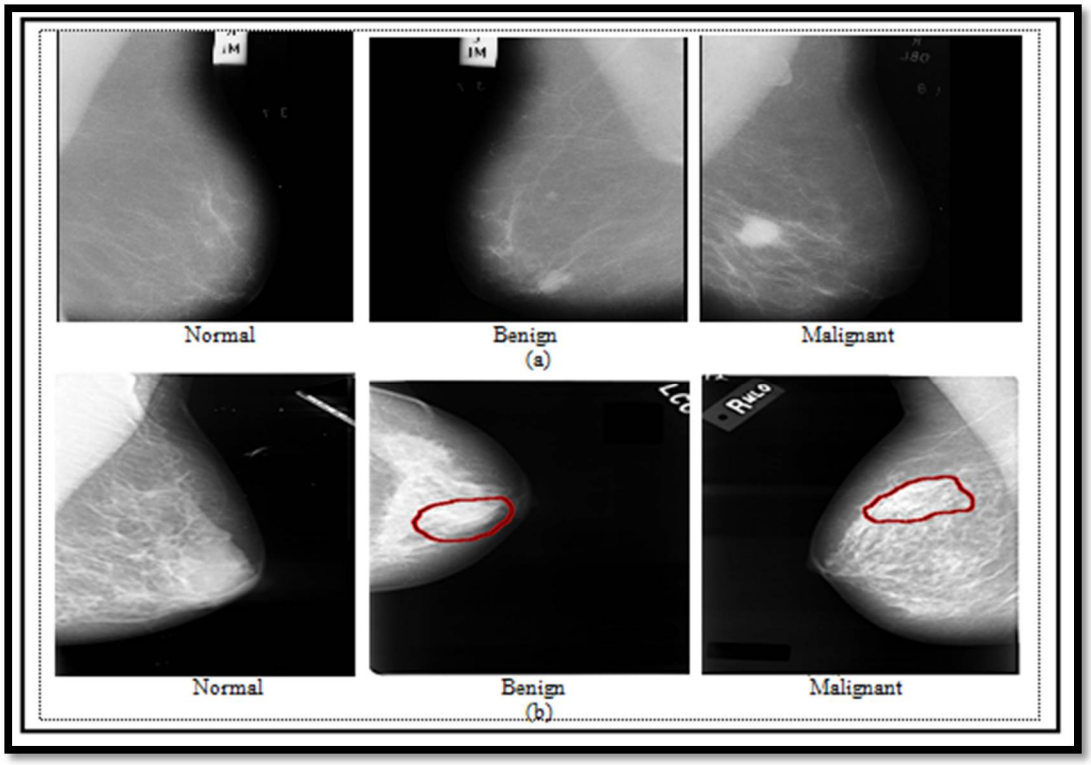

Figure 3: Normal, Benign and Malignant Samples from (A) MIAS and (B) DDSM Dataset

3.2. Data Acquisition

Breast images were acquired online from two sources, namely - the MIAS and the DDSM database. The MIAS dataset serves as the testing dataset, while the DDSM dataset serves as the training dataset. The DDSM was used as the training dataset because of the large amount of dataset required for training the convolution neural network (CNN). A total of 800 breast image samples were used for CNN i.e. 400 normal and 400 abnormal image samples. The abnormal image samples consist of both benign and malignant. A total of 400 images (200 samples for benign and 200 samples for malignant) were used for KNN. The testing set consists of 103 samples of breast images from the MIAS database which consists of benign, malignant and normal.

3.3. Image Preprocessing

Image preprocessing helps to prepare the image for feature extraction. It helps to improve the quality of the mammogram image by removing or reducing the irrelevant parts in the background and enhances the mammogram image features. The preprocessing methods employed in this research are grayscale conversion, median filtering and image enhancement. Grayscale conversion converts the input mammogram image into grayscale to reduce the dimensions from 3D array to 2D array in order to prepare the image for filtering. Grayscale conversion is done for the DDSM dataset only as the images in MIAS are already in pgm format which is already a 3D array. Median filtering helpd to remove unwanted noise from the acquired images. A 2D median filtering method with 3 × 3 neighbourhood window was used in this work. Image enhancement helps to increase the image contrast and reduce the noise. The histogram equalization method was used for image enhancement. Figure 4 shows the image preprocessing images at each stage.

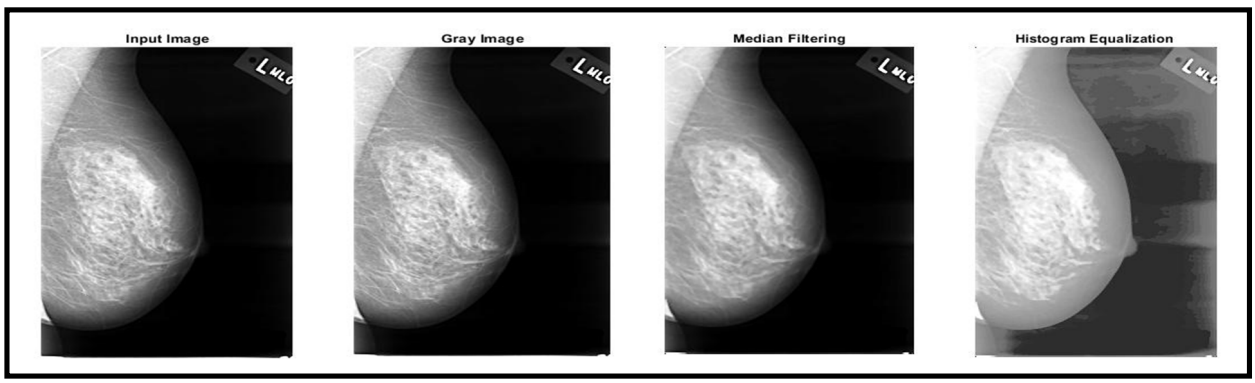

Figure 4: Image Preprocessing Stages

3.4. Image Segmentation

Segmentation helps to separate the image from the background. It helps to remove the pectoral muscles from the mammogram image. In image segmentation, only the Region of Interest (ROI) is considered and the rest of the image is eliminated. Segmentation partitions the image into multiple regions thereby extracting the ROI by identifying the masses from the mammogram image.

3.5. Feature Extraction

Feature extraction helps to extract the relevant and important characteristics of the mammogram images. For CNN, the feature extraction was done automatically using the pretrained model ResNet50, while KNN uses the Gray Level Co-occurrence Matrix (GLCM) to extract the statistical features and all these features were combined into a single vector for each image in the dataset. GLCM is a widely used statistical feature extraction method for extracting textural features from grayscale images such as mammograms. GLCM features take into account the spatial relationship between the pixel of interest and its neighboring pixels. Eleven (11) features were extracted from the mammogram image such as mean, contrast, correlation, energy, homogeneity, standard deviation, Root Mean Square (RMS), kurtosis, skewness, entropy, and smoothness. Table 1 shows the features, descriptions and their formula.

Feature	Description	Formula
Mean	Mean is used to find the value in the image that central clustering occurs	$Mean = \frac{1}{M}\sum_{i,j=1}^{M} P_{i,j}$ Where M is the number of image pixels $P_{i,j}$ is the pixel density
Contrast	Contrast is the difference between the highest and the lowest values of the adjacent set of pixels. The local variations of the image are being measured.	$Contrast = \sum_{i,j=1}^{M} P_{i,j}\left(i-j\right)^2$
Correlation	Correlation is the measure of linear dependencies between the gray tone of an image	$Correlation = \sum_{i=1}^{M}\sum_{j=1}^{M} \frac{\left(i-Mean\right)\left(j-Mean\right)P_{i,j}}{\sigma^2}$ Where σ is the standard deviation
Energy	Energy is the sum of squared elements in the GLCM. It ranges from 0 to 1	$Energy = \sum_{i,j=1}^{M}\left(P_{i,j}\right)^2$
Homogeneity	Homogeneity measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal. It ranges from 0 to 1. For diagonal GLCM, homogeneity is 1	$Homogeneity = \sum_{i,j=1}^{M} \frac{P_{i,j}}{1+\left(i-j\right)^2}$
Standard deviation (SD)	Standard deviation is used to estimate the mean distance between the pixel and the mean	$SD = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(P_{i,j}-Mean\right)^2}$
Root Mean Square (RMS)	RMS is used to calculate the RMS value of each row or column of input, as well as vectors of a specified dimension of input or the entire input	$RMS = \sqrt{\frac{1}{M}\sum_{i,j=1}^{M}\left\|Mean\right\|^2}$
Kurtosis	Kurtosis calculates the Peakness or flatness of a distribution relative to a normal distribution	$Kurtosis = \left\{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left[\frac{P_{i,j}-Mean}{SD}\right]^4\right\}-3$
Skewness	Skewness is the degree of asymmetry of a pixel distribution around its mean in the selected window	$Skewness = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(\frac{P_{i,j}-Mean}{SD}\right)^3$
Entropy	Entropy is a statistical measure of randomness that can be used to depict the texture of an image	$Entropy = \sum_{i=1}^{M}\sum_{j=1}^{N} P_{i,j}\log P_{i,j}$
Smoothness	Smoothness is a measure of grey level contrast that can be used to establish descriptors of relative smoothness	$Smoothness = 1-\frac{1}{1+SD^2}$

Table 1: GLCM Features, Descriptions and Their Formula

3.6. Classification

The classification was done using two classifiers which are the CNN and KNN. CNN was used for the first stage of classification, while KNN was used for the second stage of classification. CNN classifies the mammogram image into either normal or abnormal, while KNN classifies the abnormal image into either benign or malignant.

3.7. Performance Evaluation

The developed model was evaluated based on accuracy, sensitivity, specificity, precision, false positive rate (FPR), F1 score and Matthews correlation coefficient (MCC). Confusion Matrix was used to keep the correct and incorrect classification results using TP (True Positive), TN (True Negative), FP (False Positive) and FN (False Negative).

Accuracy: This is the ratio of correctly classified images to total number of tested images. The formula is shown in equation (1).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Sensitivity: This is the rate of the perceived positive images with the total number of positive cases. It is also known as recall. The formula is shown in equation (2).

$$Sensitivity = \frac{TP}{TP + FN} \quad (2)$$

Specificity: It is the rate of the perceived negative images with the total number of negative cases. The formula is shown in equation (3).

$$Specificity = \frac{TN}{TN + FP} \quad (3)$$

Precision: It is the ratio of the samples with actual positives to all samples that are predicted positive. The formula is shown in equation (4).

$$\Pr ecision = \frac{TP}{TP + FP} \quad (4)$$

False Positive Rate (FPR): FPR is calculated by dividing the number of false positive predictions by the total number of negatives. The value ranges between 0.0 and 1.0. Equation (5) shows the formula for FPR.

$$FPR = \frac{FP}{FP + TN} \quad (5)$$

F1 Score: This is defined as the harmonic mean of precision and sensitivity. The formula is shown in equation (6).

$$F1\ Score = \frac{2TP}{2TP + FP + FN} \quad (6)$$

Matthews Correlation Coefficient (MCC): MCC is used for binary classification. It ranges from +1 to -1. +1 indicates the best performance, while -1 indicates the worst performance. Equation (7) shows the formula of MCC.

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (7)$$

4. Results and Discussion

The developed breast cancer detection and classification model was tested using two classifiers, namely - CNN and KNN, on the MIAS dataset. The model was developed with MATLAB version 2018a and tested on an Intel core i5, 8GB RAM computer. CNN classified the mammogram images into either normal or abnormal, while KNN classified the predicted abnormal samples into benign or malignant. The testing on CNN was done on 103 mammogram image samples consisting of 44 normal, and 59 abnormal (benign and malignant), while the testing on KNN was done using 59 mammogram images consisting of 29 benign and 30 malignant.

CNN was used first, while KNN was used later. Table 2 shows the performance of the model using CNN and KNN. It was observed that CNN produces an accuracy of 99.03%, sensitivity of 0.9831, specificity of 1.0000, precision of 1.0000, FPR of 0.0000, F1 score of 0.9915 and MCC of 0.9804. The mammogram image classified as abnormal was later fed into the KNN classifier for further classification into either benign or malignant. The results of the KNN show an accuracy of 76.27%, sensitivity of 0.7667, specificity of 0.7586, precision of 0.7667, FPR of 0.2414, F1 score of 0.7667 and MCC of 0.5253.

	CNN	**KNN**
Accuracy	99.03%	76.27%
Sensitivity	0.9831	0.7667
Specificity	1.0000	0.7586
Precision	1.0000	0.7667
FPR	0.0000	0.2414
F1 Score	0.9915	0.7667
MCC	0.9804	0.5253

Table 2: Performance of the Model Using CNN and KNN

The confusion matrix for CNN and KNN was shown in Tables 3 and 4 respectively. From Table 3, the number of TP is 58, TN is 44, FP is 0 and FN is 1. TP is the number of abnormal images classified as abnormal, TN is the number of normal images classified as normal, FP is the number of normal images classified as abnormal and FN is the number of abnormal images classified as normal.

Table 4 shows the number of TP is 23, TN is 22, FP is 7 and FN is 7. TP is the number of malignant images classified as malignant, TN is the number of benign images classified as benign, FP is the number of benign images classified as malignant and FN is the number of malignant images classified as benign.

	Normal	**Abnormal**
Normal	44	0
Abnormal	1	58

Table 3: Confusion Matrix of the CNN Model

	Benign	**Malignant**
Benign	22	7
Malignant	7	23

Table 4: Confusion Matrix of the KNN Model

Figure 5 shows the graph of the performance metrics for the two models (CNN and KNN). It was observed that CNN outperformed KNN with an accuracy of 99.03% as against 76.27% of KNN and also for all other metrics.

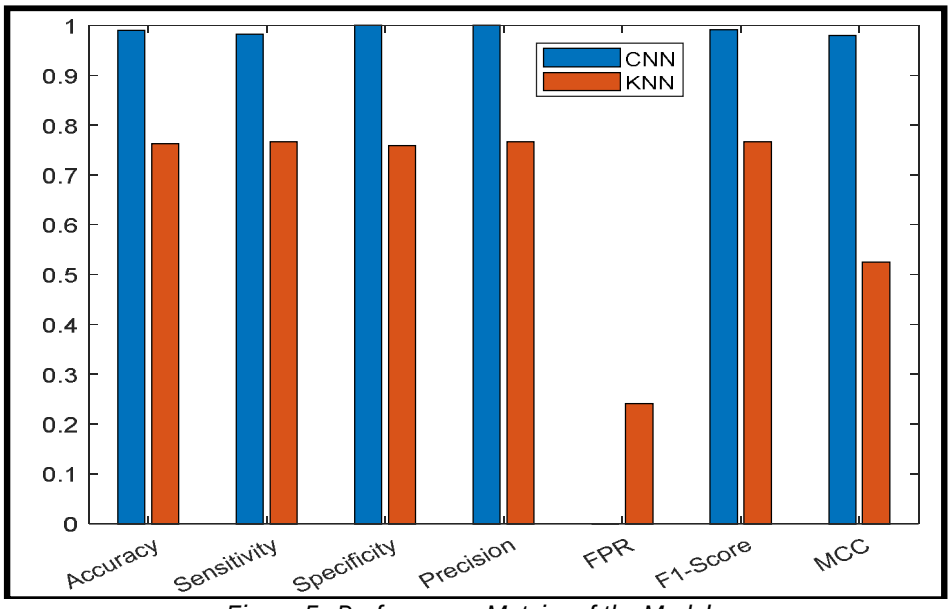

Figure 5: Performance Metrics of the Models

The features extracted using the GLCM and their values are shown in Table 5. The extracted features used for classification are mean, contrast, correlation, energy, homogeneity, standard deviation, RMS, kurtosis, skewness, entropy and smoothness.

S/N	**Extracted Features**	**Value**
1	Mean	17.0125
2	Contrast	3.6630
3	Correlation	0.2650
4	Energy	0.7191
5	Homogeneity	0.8927
6	Standard Deviation	55.5046
7	RMS	3.9505
8	Kurtosis	10.7740
9	Skewness	3.0754
10	Entropy	0.7921
11	Smoothness	1.0000

Table 5: GLCM Extracted Features and Their Values

5. Conclusion

A two-stage breast cancer detection and classification model was developed using CNN and KNN. Two different datasets, namely - MIAS and DDSM, were used. The DDSM dataset was used for training, while the MIAS dataset was used

to test the developed model. The first classification stage i.e., CNN classifies the input mammogram image into normal or abnormal, and an accuracy of 99.03%, sensitivity of 0.9831, specificity of 1.0000, precision of 1.0000, FPR of 0.0000, F1 score of 0.9915 and MCC of 0.9804 was achieved. The second classification stage classifies the abnormal image into benign and malignant, and achieved an accuracy of 76.27%, sensitivity of 0.7667, specificity of 0.7586, precision of 0.7667, FPR of 0.2414, F1 score of 0.7667 and MCC of 0.5253. The research showed that machine learning techniques can be used in the detection and prediction of the occurrence of breast cancer in women.

It was observed that the KNN prediction of benign and malignant showed a lower accuracy when compared with CNN, hence future work can be done by using another machine learning approach such as decision tree, support vector machine and so on or feature extraction method such as local binary pattern to compare the results with the results presented.

6. Acknowledgement

The authors would like to appreciate the financial support of Tertiary Education Trust Fund (TETFund) for sponsoring the project and the management of Federal Polytechnic Ede, Osun State, Nigeria for creating suitable environment for the execution of the research.

7. References

i. Adepoju, T., Ojo, J., Omidiora, E., Olabiyisi, O., & Bello, T. (2015). Detection of Tumour Based on Breast Tissue Categorization. British Journal of Applied Science & Technology, 11(5), 1-12. doi:10.9734/bjast/2015/20039
ii. Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data, 8(1), 53. doi:10.1186/s40537-021-00444-8
iii. Azar, A. T., & El-Metwally, S. M. (2013). Decision tree classifiers for automated medical diagnosis. Neural Computing and Applications, 23(7), 2387-2403. doi:10.1007/s00521-012-1196-7
iv. Davoudi, K., & Thulasiraman, P. (2021). Evolving convolutional neural network parameters through the genetic algorithm for the breast cancer classification problem. SIMULATION, 97(8), 511-527.
v. doi:10.1177/0037549721996031
vi. Ed-daoudy, A., & Maalmi, K. (2020). Breast cancer classification with reduced feature set using association rules and support vector machine. Network Modeling Analysis in Health Informatics and Bioinformatics, 9(1). doi:10.1007/s13721-020-00237-8
vii. Gardezi, S. J. S., Elazab, A., Lei, B., & Wang, T. (2019). Breast Cancer Detection and Diagnosis Using Mammographic Data: Systematic Review. J Med Internet Res, 21(7), e14464. doi:10.2196/14464
viii. Islam, M. M., Haque, M. R., Iqbal, H., Hasan, M. M., Hasan, M., & Kabir, M. N. (2020). Breast Cancer Prediction: A Comparative Study Using Machine Learning Techniques. SN Computer Science, 1(5). doi:10.1007/s42979-020-00305-w
ix. Jordan, M. I., & Mitchell, T. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349, 255 - 260.
x. Maity, N. G., & Das, S. (2017). Machine Learning for Improved Diagnosis and Prognosis in Healthcare. 2017 IEEE Aerospace Conference, 1-9.
xi. Ragab, D. A., Sharkas, M., & Attallah, O. (2019). Breast Cancer Diagnosis Using an Efficient CAD System Based on Multiple Classifiers. Diagnostics (Basel), 9(4). doi:10.3390/diagnostics9040165
xii. Saber, A., Sakr, M., Abo-Seida, O. M., Keshk, A., & Chen, H. (2021). A Novel Deep-Learning Model for Automatic Detection and Classification of Breast Cancer Using the Transfer-Learning Technique. IEEE Access, 9, 71194-71209. doi:10.1109/access.2021.3079204
xiii. Sakri, S. B., Rashid, N. B. A., & Zain, Z. M. (2018). Particle Swarm Optimization Feature Selection for Breast Cancer Recurrence Prediction. IEEE Access, 6, 29637-29647. doi:10.1109/ACCESS.2018.2843443
xiv. Tosin, A., Morufat, A., Omotayo, O., Bolanle, W., Olusayo, O., & Olatunde, O. (2018). Curvelet Transform-Local Binary Pattern Feature Extraction Technique for Mass Detection and Classification in Digital Mammogram. Current Journal of Applied Science and Technology, 28(3), 1-15. doi:10.9734/cjast/2018/42579
xv. Tsochatzidis, L., Costaridou, L., & Pratikakis, I. (2019). Deep Learning for Breast Cancer Diagnosis from Mammograms-A Comparative Study. J Imaging, 5(3). doi:10.3390/jimaging5030037
xvi. Van Engelen, J. E., & Hoos, H. H. (2019). A survey on semi-supervised learning. Machine Learning, 109(2), 373-440. doi:10.1007/s10994-019-05855-6
xvii. Viegas, L., Domingues, I., & Mendes, M. (2021). Study on Data Partition for Delimitation of Masses in Mammography. J Imaging, 7(9). doi:10.3390/jimaging7090174
xviii. Wang, Z., Luo, Y., Xin, J., Zhang, H., Qu, L., Wang, Z., . . . Wang, X. (2020). Computer-Aided Diagnosis Based on Extreme Learning Machine: A Review. IEEE Access, 8, 141657-141673. doi:10.1109/access.2020.3012093
xix. Yang, X., Song, Z., King, I., & Xu, Z. J. A. (2021). A Survey on Deep Semi-supervised Learning. abs/2103.00550.
xx. Yue, W., Wang, Z., Chen, H., Payne, A., & Liu, X. (2018). Machine Learning with Applications in Breast Cancer Diagnosis and Prognosis. Designs, 2, 13. doi:10.3390/designs2020013
xxi. Zhou, X., Li, C., Rahaman, M. M., Yao, Y., Ai, S., Sun, C., . . . Teng, Y. (2020). A Comprehensive Review for Breast Histopathology Image Analysis Using Classical and Deep Neural Networks. IEEE Access, 8, 90931 - 90956.