

# Hybrid Approach with Bayesian Optimization for the Classification of Breast Cancer from Ultrasound Images

Monday Onoja  
Mathematics and Computer Science  
Department  
Federal University of Health Sciences,  
Otukpo, Nigeria.  
[monday.onoja@fuhso.edu.ng](mailto:monday.onoja@fuhso.edu.ng)

Joy Omeyi Omoha  
Department of Biostatistics  
Federal University of Health  
Sciences,  
Otukpo, Nigeria.  
[joy.omoha@fuhso.edu.ng](mailto:joy.omoha@fuhso.edu.ng)

Olawale Abimbola  
Machine Learning Engineer  
Creative Advanced Technologies  
Dubai, UAE  
[abimbolaolawale41@gmail.com](mailto:abimbolaolawale41@gmail.com)

Abayomi Jegede  
Computer Science Department  
University of Jos  
Jos, Nigeria  
[jegede@unijos.edu.ng](mailto:jegede@unijos.edu.ng)

Oga Ode  
Department of Biostatistics  
Federal University of Health  
Sciences,  
Otukpo, Nigeria.  
[oga.ode@fuhso.edu.ng](mailto:oga.ode@fuhso.edu.ng)

Temidayo Oluwatosin Omotehinwa  
Mathematics and Computer Science  
Federal University of Health Science,  
Otukpo, Nigeria  
[oluomotehinwa@gmail.com](mailto:oluomotehinwa@gmail.com)

**Abstract**—Breast ultrasound is one of the most common clinical practices for imaging abnormalities in the breast. Many modalities for identifying the type of cancer that exist in detected lesions have been proposed. However, the goal to provide a more accurate identification of whether cancer exists or not and how problematic they appear, is yet to be achieved. There's need to provide improved accuracy for the identification of breast cancer. Improved accuracy will enhance effective detection, leading to effective treatment. This paper integrates the efficiency of LightGBM algorithm and the effectiveness of XceptionCNN with Bayesian Optimization for the early detection of breast cancer from Ultrasound Images. The model was tested on 600 female patients between the ages of 25 and 75 years, with 780 images. Performance evaluation based on the breast Ultrasound images shows that our model produced an improved accuracy of 99.6%, which outperforms other state of the art models. This improved accuracy shows that integrating LightGBM and XceptionCNN with Bayesian Optimization greatly improves Breast cancer detection and thereby greatly impacts on its treatment.

**Keywords**— *LightGBM, XceptionCNN, Breast Cancer Detection, Bayesian Optimization, Ultrasound Images.*

## I. INTRODUCTION

Over the years, breast ultrasound has proven to be a method for imaging palpable abnormalities in the breast, which distinguishes cystic from solid masses [1]. Therefore, it stands as a potential alternative to mammography, which describes the procedure for breast cancer detection in good time. In low-resource countries, Ultrasound can be adopted because it is cheaper, cost-effective and has a broader scope of clinical use [2]. The goals of various imaging modalities are detection of breast cancer in good time and accurate assessment of lesions [3]. LightGBM applies a histogram-based methodology which partitions data into trunks using a histogram of the distribution.

It increases model efficiency and saves memory. It has proven to be effective for malware and breast cancer detection [4][5].

The study by [6] used data augmentation to add to the size of the original dataset for improved learning of CNN models. Their Machine Learning classifier achieved 99.1% accuracy on the Breast Ultrasound Image. In order to distinguish benign and malignant Breast Tumors in Ultrasound images, [7] used Deep Learning model to propose the Filer Network (NF-NET) model which achieved an accuracy of 73%, a precision of 69%, recall of 80% and 0.74 F1 score. A study by [8] used image fusion method with ensemble CNN (VGGNet, ResNet, DenseNet) architecture on Breast Ultrasound Images. The model obtained 94.62% accuracy, 90% precision and 91.14% F1 score. [9] developed a semi-supervised GAN model to augment the Breast Ultrasound Images. CNN was used to classify the synthesized Images. The model achieved 90.41% accuracy. Similarly, reference [10] combined Image decomposition and fusion techniques with adaptive spatial fusion technology to propose a reliable classification method for Breast Ultrasound Image tumor. The proposed model obtained 95.5% accuracy, 98% Precision, 94% recall and 96% F1-score. A study by [11] combined the effectiveness of LightGBM algorithm with efficiency of XceptionCNN model for efficient and effective malware detection using the maling dataset. It was compared to other novel models on malware detection, the model achieved the best classification and training accuracy. However, the model was only tested on malware classification domain.

In this paper, by taking advantage of the power of pre-trained CNN model for image feature extraction and Bayesian-Optimization for tuning of hyper-parameters, with an additional Gradient Boosting (LightGBM) framework, we introduce a faster and more accurate approach for classifying Breast Ultrasound Images.

We organize the rest of the paper as follows: we discuss the methodology/technique used in section Two, the results of our experiment and its implication in section Three and summary with conclusions made based on our research objectives in section Four.

## II. METHODOLOGY

### A. Dataset

We ran experiments on medical images of breast cancer using an Ultrasound Scan dataset. The dataset [12] examined in this study was collected in 2018. It's at baseline and is freely accessible. The number of patients is 600 females between the ages of 25 and 75 years and contains 780 PNG images with an average image size of 500x500 pixels. The dataset is imbalanced and organized into three classes namely; normal (133), benign (487) and malignant (210). However, each of these image classes also have the same number of segmented masked images associated with them. Fig. 1 shows the dataset distributions.

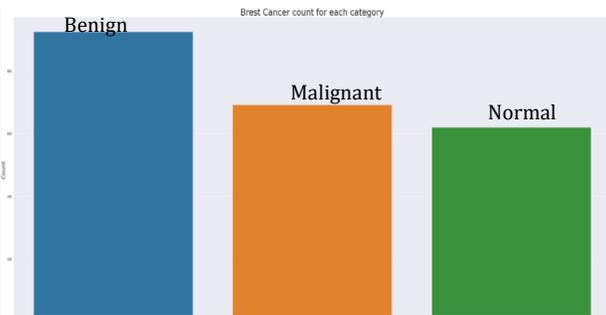


Fig. 1. Vertical bar plot of the dataset distributions

### B. Data Preprocessing

The data was organized into two folders i.e Original and Mask, with each folder classified into the three classes. As shown in fig. 1, the actual scans of the dataset do not clearly show the identified lesions. The mask images make them obvious for cases where they exist. As part of our preprocessing pipeline, we normalized the images in the data folder using histogram equalization and overlapped them with their respective masks to produce another (preprocessed) image containing information about the breast scan and the lesions as shown in fig. 2.

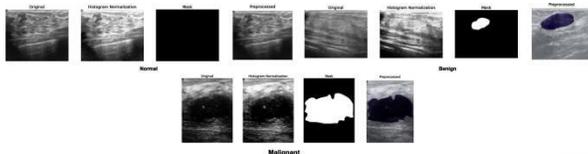


Fig. 2. Sample Data Transformation

### C. LightGBM

As proposed by Microsoft [13], LightGBM is a powerful Gradient Boosting Decision Tree (GBDT) framework. The technique serves in Machine Learning models such as ranking and classification [14]. The framework operates leaf-wise splitting of the tree by using the best fit method. Leaf-wise

splitting improves speed and decreases loss. When loss is reduced, improved accuracy can be obtained. In Fig. 3, we show the leaf-wise tree growth architecture of the LightGBM algorithm.

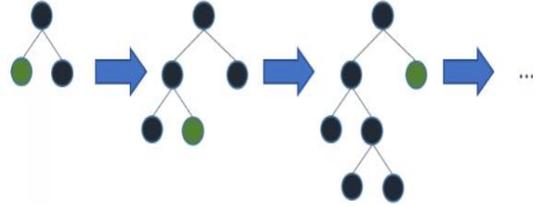


Fig. 3: LightGBM Architecture [21]

### D. XceptionCNN Architecture

The Xception model was created from an 'extreme' interpretation of Google's Inception model [15]. It is a linear stack of 36 independent convolution layers which makes use of depth-wise splitting technique and are connected together by residual connections. The network feature extraction is carried out by the layered stack. The 36 independent convolution layers are grouped into 14 modules, which are connected by linear residual connections, with the exception of the first and last modules [15].

### E. Hybrid Modelling

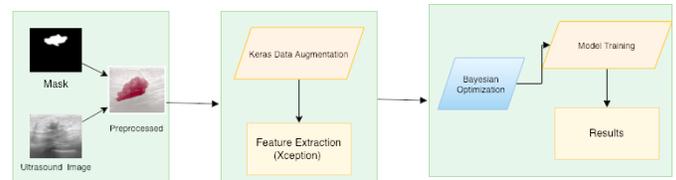


Fig. 4. Proposed model architecture

As presented in fig.4, we overcame the challenge posed by insufficient dataset by implementing keras data generator approach [16], which created additional images from the existing preprocessed images by first rescaling, then creating shear-range images, and finally, generating both vertical and horizontal flipped images. This process generated additional 500 images for each class. Secondly, we transformed the image dataset into numerical features by leveraging on the representational power of XceptionCNN pre-trained deep neural networks model [17]. The aim is to get the key information from the novel data and represent that information in a lower dimensionality space [18]. Finally, we adopted the LightGBM classifier for training because of how it speeds up training [19]. However, to achieve an accurate level of hyperparameters, we

made use of Bayesian Optimization, which is a more efficient hyperparameter tuning technique [20].

### III. RESULTS AND DISCUSSION

Data augmentation resulted to a new data size of 2237 images (normal: 620, benign: 925, malignant: 692). The data was further divided into 80% train set and 20% test set. Features of the trained data generated from Xception were fed into the LightGBM classification algorithm, producing a more accurate result across all classes. Table 1 shows the result of our experiment. The results from the confusion matrix of the LightGBM–XceptionCNN experiment revealed that the model achieved a very good true positive and true negative rates, with a low false positive and false negative rate as shown in Fig. 5. A total data point of 448 was obtained, which corresponds to the 20% (test set) of the total dataset.

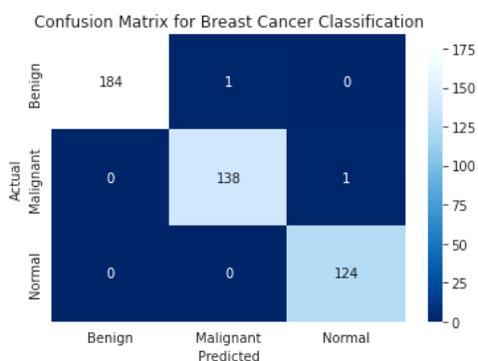


Fig. 5. Confusion matrix

The performance of the LightGBM – XceptionCNN model based on sensitivity (recall), F1 score, balanced accuracy, and Precision were obtained using Eqn. (1) to (4) and are presented in Table 1.

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions Made}}$$

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

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

$$\text{F1 Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

Where TN stands for True Negative, TP: True Positive, FN: True Negative, FN: False Negative.

TABLE 1: PERFORMANCE OF LIGHTGBM – XCEPTIONCNN MODEL

Metrics	Value
Accuracy	0.9955
Recall	0.9958
Precision	0.9955

F1_Score	0.9955
----------	--------

Our hybrid LightGBM-XceptionCNN model with Bayesian Optimization for breast cancer detection using Ultrasound Images was measured using equ. (1) to equ. (4). From the outcome of the confusion matrix shown in fig. 2, the model achieved an accuracy of 0.9955, Recall of 0.9958, Precision of 0.9955 and F1 Score of 0.9955. An Accuracy of 0.9955 implied that 99.6% instances of the model were correctly classified. Similarly, recall of 0.9958 implied that in 99.6% of actual malignant instances, the model was classified as malignant. A precision of 0.9955 shows that the model predicted 99.6% of malignant instances. Lastly, the F1-score of 0.9955 represents the model overall performance of 99.6%.

We compare the results in Table 1 to those obtained from recent and similar research with the same dataset.

TABLE 2. COMPARISON WITH OTHER NOVEL TECHNIQUES

Author	Year	Method	Accuracy (%)
Cao et al. [7]	2020	NF -- NET	73 . 0
Pang et al. [9]	2021	CNN	90.41
Moon et al. [8]	2020	CNN (VGGNet, ResNet, DenseNet)	94.62
Zhuang et al. [10]	2021	Fusion techniques	95.48
Kiran et al. [6]	2022	CNN	99.1
Proposed (ours)	2023	LightGBM – XceptionCNN	99.6

From the performance comparison presented in Table 2, our model has an improved accuracy of 99.6%, which outperformed the existing models for the detection of breast cancer using Ultrasound Images. This improvement may result in detecting breast cancer early, leading to early treatment. The strength of our approach was in combining the actual data with the masked image before pre-processing the data for the Machine Learning algorithm and adopting a bayesian Hyperparameter Optimizer which is known to be a powerful optimization approach. The overall improved performance accuracy will lead to early and accurate detection of breast cancer, which will reduce the number of early deaths. Breast Ultrasound images can produce great results when combined with machine learning.

### IV. CONCLUSION

We investigated the application of XceptionCNN and LightGBM algorithm with Bayesian optimization for early detection of breast cancer with Breast Ultrasound Images. We used XceptionCNN for feature extraction and LightGBM algorithm for training with Bayesian hyperparameter Optimizer for parameter tuning. Our model obtained an improved accuracy of 99.6% which significantly surpassed existing models for breast cancer detection as presented in Table 2. The model was tested on 600 female patients between the ages of 25 to 75 years

with about 780 images. The result shows that combining XceptionCNN with LightGBM algorithm produces a more accurate technique for the detection of breast cancer in good time from Ultrasound Images. It aids the detection of breast cancer accurately and in good time, thereby reducing the number of early mortalities. Breast Ultrasound images can produce great results when combined with Machine Learning.

#### REFERENCES

- [1] A. Athanasiou, A. Tardivon, L. Ollivier, F. Thibault, C. E. Houry, & S. Neuenschwander, "How to optimize breast ultrasound", *European Journal of Radiology*, Vol. 69 No. 1, pp. 6–13, 2009. <https://doi.org/10.1016/j.ejrad.2008.07.034>
- [2] R. Sood, A. F. Rositch, D. Shakoor, E. Ambinder, K. L. Pool, E. Pollack, D. J. Mollura, L. A. Mullen, & S. C. Harvey, Ultrasound for Breast Cancer Detection Globally: A Systematic Review and Meta-Analysis. *Journal of global oncology*, Vol. 5, pp. 1–17, 2019. <https://doi.org/10.1200/JGO.19.00127>
- [3] R. Guo, G. Lu, B. Qin, & B. Fei, "Ultrasound imaging technologies for Breast cancer detection and management: A review. *Ultrasound in medicine & biology*", Vol. 44, No. 1, pp. 37–70, 2018. <https://doi.org/10.1016/j.ultrasmedbio.2017.09.012>
- [4] M. Onoja, A. Jegede, M. Jesse, A. Gilbert, O. Ayodele, O. Kolawole, Exploring the Effectiveness and Efficiency of LightGBM Algorithm for Windows Malware Detection, 5th Information Technology for Education and Development (ITED 2022), pp. 1–6, Abuja, Nigeria. 1st - 3rd November, 2022, IEEE. <https://doi.org/10.1109/ITED56637.2022.10051488>
- [5] T. O. Omotehinwa, D. O. Oyewola, & E. G. Dada, A light gradient-boosting machine algorithm with tree-structured parzen estimator for breast cancer diagnosis. *Healthcare Analytics*, p. 100218, 2023. <https://doi.org/10.1016/j.health.2023.100218>
- [6] K. Jabeen, M. S. Khan, M. Alhaisoni, U. Tariq, Y. Zhang, A. Hamza, A. Mickus, & R. Damaševičius, Breast cancer classification from ultrasound images using probability-based optimal deep learning feature fusion. *Sensors*, Vol. 22, Issue 3, p. 807, 2022. <https://doi.org/10.3390/s22030807>
- [7] Z. Cao, G. Yang, Q. Chen, X. Chen, & F. Lv, Breast tumor classification through learning from noisy labeled ultrasound images. *Medical Physics*, Vol 47 Issue 3, pp. 1048–1057, 2019. <https://doi.org/10.1002/mp.13966>
- [8] W. K. Moon, Y. Lee, H. Ke, S. H. Lee, C. Huang, & R. Chang, "Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks", *Computer Methods and Programs in Biomedicine*, Vol. 190, p. 105361, 2020. <https://doi.org/10.1016/j.cmpb.2020.105361>
- [9] T. Pang, J. H. D. Wong, W. L. Ng, & C. S. Chan, "Semi-supervised GAN-based radiomics model for data augmentation in breast ultrasound mass classification". *Computer Methods and Programs in Biomedicine*, Vol 203, p. 106018, 2021. <https://doi.org/10.1016/j.cmpb.2021.106018>
- [10] Z. Zhuang, Z. Yang, A. N. J. Raj, C. Wei, P. Jin, & S. Zhuang, Breast ultrasound tumor image classification using image decomposition and fusion based on adaptive multi-model spatial feature fusion. *Computer Methods and Programs in Biomedicine*, Issue 208, p. 106221, 2021. <https://doi.org/10.1016/j.cmpb.2021.106221>
- [11] M. Onoja, A. Jegede, N. V. Blamah, A. V. Olawale, & T. O. Omotehinwa, "EEMDS: efficient and effective malware detection system with hybrid model based on XceptionCNN and lightGBM algorithm. *Journal of Computing and Social Informatics*, Vol. 1, Issue 2, pp. 42–57, 2022. <https://doi.org/10.33736/jcsi.4739.2022>
- [12] W. Al-Dhabyani, M. A. Gomaa, H. Khaled, & A. A. Fahmy, "Dataset of breast ultrasound images", *Data in Brief*, vol. 28, p. 104863, 2020. <https://doi.org/10.1016/j.dib.2019.104863>
- [13] G. Ke et al., "LightGBM: A highly efficient gradient boosting decision Tree," 31st International Conference on Neural Information Processing Systems, ACM DL, 2017, pp. 3149–3157
- [14] M. Abbadi, M. Al-Bustanji, and M. Al-Kasassbeh, "Robust Intelligent malware detection using LightGBM Algorithm," *International Journal of Innovative Technology and Engineering*, vol. 9, no. 6, pp. 1253-1263, April. 2020.
- [15] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions.," Google, Inc, 2017.
- [16] Keras\_ [https://www.tensorflow.org/api\\_docs/python/tf/keras/preprocessing/image/ImageDataGenerator](https://www.tensorflow.org/api_docs/python/tf/keras/preprocessing/image/ImageDataGenerator). [Accessed April 5th 2023]
- [17] F. Chollet, "Xception: Deep Learning with depthwise separable convolutions", 2017. <https://doi.org/10.1109/cvpr.2017.195>
- [18] G. Kumar, & P. Bhatia, A Detailed Review of Feature Extraction in Image Processing Systems, 2014 Fourth International Conference on Advanced Computing Communication Technologies, pp. 5–12, 2014. <https://doi.org/10.1109/acct.2014.74>
- [19] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, & T. Liu, "LightGBM: A Highly Efficient Gradient Boosting Decision Tree", NIPS, 2017.
- [20] Joy, T. T., Rana, S., Gupta, S., & Venkatesh, S. (2016). Hyperparameter tuning for big data using Bayesian optimisation. 2016 23rd International Conference on Pattern Recognition (ICPR), pages (pp. 2574–2579). <https://doi.org/10.1109/icpr.2016.7900023>
- [21] P. Khandelwal, "Which algorithm takes the crown: Light GBM vs XGBOOST?," 2017. [online]. Available: <https://www.analyticsvidhya.com/blog/2017/06/which-algorithm->. [Accessed Oct. 12, 2023].