International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Breast Cancer Detection with VGG16: A Deep Learning Approach with Thermographic Imaging

Ferdous Ahmed[1*], Sumaiya Akter Shukhy[2], Ishtiak Alam Rafi[3], Arif Mahmud Sisir[4], Md Mijanur Rahman[5]

Submitted:13/03/2024 **Revised**: 28/04/2024 **Accepted**: 05/05/2024

Abstract: Breast Cancer (BC) still stands as a major global health issue that demands state of the art diagnostic instruments for early detection and mortality reduction. It is a radiation-free, non-invasive method of detecting temperatures and might be used as an adjuvant next to the regular methods. This study establishes a deep learning (DL) model with the state-of-the-art method using transfer learning and pre-trained VGG16 convolutional neural network. We train and evaluate the model on thermal images taken from Database for Research in Mastology with Infrared Images (DMR-IR). They also use augmentation and normalization methods to improve model performance. The DL-based model obtained a promising detection rate of 99.4% for the prediction of BC lesions. It demonstrates a sensitivity of 100%, specificity of 97.5%, Recall of 99%, Precision of 98.9%, F1-Score of 99.8%, and AUC-ROC of 99.8%, surpassing previous models. Comparative analysis with alternative methods such as Support Vector Machines (SVM), K-Nearest Neighbor (KNN), Decision trees (DT), and Gradient Boosting (GB) underscores the superior performance of the DL approach. This research presents a groundbreaking approach to BC detection, leveraging deep learning and thermal imaging technology. The findings highlight the efficacy of the proposed DL model in enhancing diagnostic accuracy and supporting informed decision-making in clinical settings. Future research could explore broader applications and integration into healthcare practices to improve patient outcomes.

Keywords: *BC detection, Convolutional neural network, Deep learning, Data augmentation, Machine learning, Normalization, Thermal imaging, VGG16*

1. Introduction

BC is the most critical issue in public health, and it is the primary cause of cancer-related deaths among women worldwide [1]. Due to genetic mutations in breast cells, an uncontrolled increase in the number of cells occurs, leading to the development of BC [2]. Moreover, there are various types of breast cancer. Ductal Carcinoma In-Situ (DCIS) is a non-invasive form, while invasive ductal carcinoma, the most common type, is responsible for 70–80% of cases. Inflammatory breast cancer, obstructing lymph vessels, is rare (1–5% of cases). Triple-negative breast cancer, an aggressive form lacking estrogen, progesterone receptors, and HER2, constitutes approximately 15% of cases. Other less common types collectively make up about 1% of all breast cancer cases [3],[4]. Advancements in early disease diagnosis and treatment techniques have enhanced the survival prospects in cases of BC.

[1] *Student, Department of CSE, Southeast University, Tejgaon- 1208, Bangladesh,* ***Email: 2020100000011@seu.edu.bd*(Corresponding Author)***
ORCID ID: 0009-0002-7744-0998
[2] *Student, Department of CSE, Southeast University, Tejgaon- 1208, Bangladesh,* ***Email: 2018000000053@seu.edu.bd***
[3] *Student, Department of CSE, Southeast University, Tejgaon- 1208, Bangladesh,* ***Email: 2019200000013@seu.edu.bd***
[4] *Student, Department of CSE, Southeast University, Tejgaon- 1208, Bangladesh,* ***Email: 2019200000035@seu.edu.bd***
[5] *Assistant Professor, Department of CSE, Southeast University, Tejgaon- 1208, Bangladesh,* ***Email: mijanur.rahman@seu.edu.bd***

Different methods, like breast examinations, mammograms, ultrasounds, MRIs, and CTs, are utilized to detect BC, but their diagnostic performance and satisfaction are vastly limited. Mammography, effective in early detection and mortality reduction by 25%, has the drawback of radiation risk, limiting it to biannual screenings. Ultrasound, suitable for dense breast tissue, is quick but lacks early detection and has a higher false-positive rate. MRI is reliable but costly, requires a breastfeeding pause, and may miss early-stage cancer. Advanced CT techniques reduce radiation and imaging time but pose challenges like breath-holding and potential risks to pregnant patients [5],[6].

Nevertheless, the primary drawbacks of these methods include their impracticality for routine mass screening and, in some cases, limited accessibility, particularly for individuals residing in remote regions of a country. Enhancing accuracy in the thermal image dataset through advanced algorithms in deep learning shows promising outcomes in early disease detection and presents numerous benefits. Additionally, being painless, safe, non-invasive, and cost-effective, it significantly advances BC diagnostics and treatment.

Despite the advancements in breast cancer diagnosis, additional research is essential to explore diverse modalities and improve detection accuracy on extensive datasets through the integration of multiple datasets. Such research is particularly crucial when employing transfer learning alongside segmentation, augmentation, and normalization techniques on a broader scale. The primary objective of this study is to address these pressing needs. Identifying a dependable and practical approach to utilize pre-trained VGG16 with data enhancement and normalization for breast cancer detection using thermal imaging is crucial. Thermography, relying on the temperature of breast regions, indicates that they are warmer than the surrounding tissues [7],[8]. The aim is to overcome limitations associated with small datasets and mitigate concerns about sharing medical images, which can cause anxiety for some patients [9]. This research has the potential to significantly contribute to the development of a more comprehensive and reliable screening process for breast cancer tumors, ultimately benefiting both medical professionals and patients alike.

This research investigates the ability of the VGG16 convolutional neural network to detect breast cancer (BC). An augmentation technique generates a large dataset from a segmented dataset, effectively expanding available data without further annotation. Data normalization via labeling further enhances the model's performance. The evaluation metric will comprehensively assess accuracy, specificity, AUC score, and confusion matrix for malignant and benign tumors. Ultimately, comparing the results of the top-performing model against previous studies using the DMR-IR dataset will provide insights into the effectiveness of VGG16 for this specific application.

This research study cited the following contributions:

1. This study examined the performance of the VGG16 deep learning model for BC detection using the DMR-IR thermal dataset. It explored the impact of data augmentation and normalization techniques on VGG16's accuracy in classifying benign and malignant cases.
2. Diverse machine learning approaches like SVM, KNN, DT, GB, and Sequential were applied for accurate BC classification.
3. A comparative analysis is conducted between the proposed model and those presented in relevant research within the field.

This study unfolds as follows: Section 2 discusses the related work; Section 3 describes the methodology of the proposed model for BC detection, which encompasses data loading, augmentation, normalization, preprocessing, implementing imageNet weights, model building, and performance evaluation; Sections 4 and 5 presents the results and discussions, respectively; Section 6 addresses the conclusion and outlines directions for future research.

2. Literature Review

The following section provides an overview of past research studies on the prediction of BC screening, incorporating contemporary methodologies that harness machine learning and deep learning.

Alia Alshehri et al. [10] created VGG16 with and without self, soft, and hard deep attention mechanisms (AMs). They used the DMR-IR dataset to get outstanding results in telling the difference between normal and abnormal breasts in thermal images, with up to 99.8% accuracy compared to 99.18% without AMs. They utilized a stratified 10-fold cross-validation method, effectively mitigating overfitting-related generalization issues. However, the proposed model achieved better results without using AMs than what they achieved.

In their paper, Seifedine Kadry et al. [11] proposed enhancing transfer learning performance using thermal images from the Benchmark dataset. They employed VGG16 coupled with decision trees. Following their methodology, they divided the dataset into training (70%), testing (20%), and validation (10%). Upon evaluating the models, they found them to have a high accuracy of 95.5%. Besides accuracy, this approach also improved precision, sensitivity, specificity, and F1 scores. Moreover, the DT outcome is compared with other binary classifiers, achieving superior performance. However, the proposed study surpassed their accuracy.

Caroline B. Gonc et al. [12] operated convolutional neural networks such as VGG16, Densenet201, and ResNet50 to classify a thermography dataset. Transfer learning techniques were used to improve the process of categorizing static thermography images and differentiating between the "sick" and "healthy" categories. In particular, the Densenet model performed better than the others, with an F1-score of 0.92, accuracy of 91.67%, sensitivity of 100%, and specificity of 83.3%. This evaluation demonstrated the effectiveness of using deep learning models for thermography image classification, conducted on a dataset with 38 static photos for each class. Regardless, the suggested approach yielded superior outcomes.

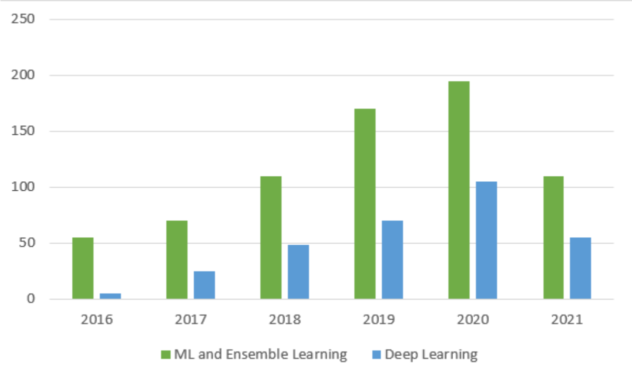
Dheeb Albashish et al. [13] used a CNN with a VGG16 architecture to pull out features from the BreaKHis benchmark histopathological image dataset. They got terrific results with VGG16 and Poly-SVM for binary classification and RBF-SVM for multiclass classification. They got a reading of 96.0 ± 2.20 for binary classification and 89.83 ± 0.005 for eight-class classification. Their approach outperformed previous works on the same dataset and surpassed recent classical machine learning algorithms. In contrast, the proposed work focused solely on binary classification, exceeding their outcomes.

In their paper, Devanshu Tiwari et al. [14] utilized deep transfer learning models to classify normal or abnormal breast thermal images actively, training a VGG16 model explicitly with a DMR-IR dataset featuring representations in multi and single-view configurations. Constructing a multi-view thermal image combining standard frontal, left, and right-view breast thermal images with infrared images enhances system accuracy by providing a more comprehensive thermal temperature reading. VGG16 demonstrated a promising test accuracy of 99% on a dynamic breast imaging test dataset using multi-view images. On the other hand, alternative deep transfer learning models such as VGG19, ResNet50V2, and InceptionV3 achieved test accuracies of 95%, 94%, and 89%, respectively. Despite their complexity and superior results in other medical imaging tasks, VGG16 outperformed these models. Moreover, recent studies employing thermal imaging for diagnosing BC have shown promising outcomes in contrast to other imaging methods. Furthermore, thermal imaging has proven effective when integrated with machine learning (ML) and deep learning (DL) techniques. In this study, notably, impressive results were achieved solely utilizing frontal views, which were reshaped through an augmented strategy, thereby mitigating data scarcity concerns.

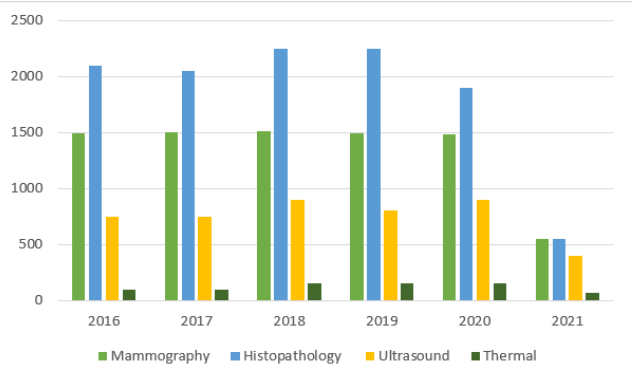
Subhrajit Dey et al. [15] introduced a pre-trained DenseNet121 model with two edge map detectors—Prewitt and Roberts—that converted input from thermal images. It achieves an impressive accuracy of 98.8% on the DMR-IR dataset, surpassing existing methods. This promising approach offers the potential for affordable and accurate early detection.

Madallah Alruwaili et al. [16] studied how to find BC using a transfer learning framework and different image enhancement techniques, leading to higher accuracy than previous research. On the MIAS dataset, ResNet50 (MOD-RES) achieved an accuracy of 89.5%, significantly outperforming the smaller Nasnet-Mobile network (70%). This observation suggests that pre-trained models excel in medical imaging, especially with limited data.

Abeer Saber et al. [17] created a deep learning model using transfer learning to find breast cancer automatically. Their study used 80-20 split and cross-validation on mammography images (MIAS dataset). The model employed a pre-trained CNN architecture, specifically the VGG16 algorithm, achieving high accuracy, sensitivity, specificity, precision, F-score, and AUC. However, their study faced limitations such as small datasets and unclear performance for specific cancer types. In contrast, the proposed research utilized a larger dataset (DMR-IR) and demonstrated more detailed performance. As shown in Figure 1, the search results from PubMed for the years 2016 to 2021 provide a comparative analysis of the number of research publications published on (a) BC detection and classification based on Machine Learning and Deep Learning and (b) Breast Mammography, Histopathology, Ultrasound, and Thermal techniques.



(a)



(b)

Fig 1. The search results from PubMed for the years 2016 to 2021 provide a comparative analysis of the number of research publications published on (a) BC detection and classification based on Machine learning and Deep learning and (b) Breast Mammography, Histopathology, Ultrasound, and Thermal techniques.

Table 1 presents the data obtained through extensive examination of numerous books. The first column displays reference numbers. The second column provides the respective periods, while the third lists the datasets. The study algorithms utilized are outlined in the fourth column. Lastly, the fifth column illustrates the evaluation results of the algorithms.

Table 1. Evaluation of Prediction Models Accessible to the Public.

Ref. No.	Period	Datasets	Algorithm	Evaluation Results (%)
Dennies Tsietso [18]	2023	DMR-IR Dataset	CADx, DNN, AlexNet	Acc = 90.48, Sn = 93.33, AUC = 0.94
Ahsan Rafiq [19]	2023	Kaggle data set (BC)	1-CNN,2-CNN,3-CNN	Acc = 90.10, Rec = 89.90, Prec = 89.80, F1-Score = 89.90
Noor Kamal Al-Qazzaz [20]	2023	Wisconsin University database	SVM, DT, KNN,	Acc = 60 to 80 and 93.33, respectively FA, t-SNE
N. Aidossov [21]	2023	DMR-IR Dataset	MobileNet BN+CNN,	Acc = 93.8
Manas Minnoor [22]	2023	Wisconsin BC Diagnostic dataset	Random Forest, SVM, DT, KNN, Multilayer Perceptron	Acc = 100 (initial dataset) and 99.30 (minimal dataset)
Esraa A. Mohamed [23]	2022	DMR-IR Dataset	U-Net, Two-class CNN	Acc = 99.33, Sn = 100 and Sp = 98.67
Mohammed Abdulla [5]	2020	DMR-IR Dataset	RBFN, KNN, PNN, SVM, ResNet50, SeResNet50, V Net, Bayes Net, CNN, C-DCNN, VGG-16, Hybrid (ResNet-50 and V-Net), ResNet101, DenseNet and InceptionV3	Acc = 80 to 100 (Not actually defined)
Mohamed Abdel-Nasser [24]	2019	DMR-IR Dataset	CAD (Computer-Aided Diagnostic) ML	Acc = 95.8, Rec = 97.1, Prec = 94.6, F1-Score = 95.4, AUC = 98.9

(Accuracy = Acc, Sensitivity = Sn, Specificity = Sp, Precision = Prec, Recall = Rec and Area Under Curve = AUC).

Deep learning models are often assessed based on accuracy in the existing literature. While some models show high accuracy, only some effectively diagnose conditions, performing well with small datasets but less with larger ones. The study addresses this by augmenting the dataset, normalizing to eliminate annotations, and balancing imbalances. The evaluation focuses on metrics like recall, precision, specificity, and AUC scores to enhance early-stage BC detection. Furthermore, incorporating these complementary metrics is imperative to safeguard patients from the severe risks of missed diagnoses.

This review examined recent research on breast cancer (BC) screening using machine learning (ML) and deep learning (DL) techniques applied to thermography. The studies explored in this paper demonstrate the promising potential of thermography, particularly when combined with advanced AI methods, for accurate and early BC detection. Several Deep learning architectures, including VGG16, DenseNet, ResNet, and U-Net, achieved high accuracy in classifying normal and abnormal breast tissue in thermal images. Notably, studies incorporating multi-view thermography, and data augmentation techniques yielded superior performance. These findings suggest that ML and DL have the potential to revolutionize BC screening by offering a non-invasive, potentially more affordable, and complementary tool for early detection alongside traditional screening methods. Further research with larger and more diverse datasets is crucial to refine these techniques and ensure their generalizability for real-world clinical applications.

3. Materials and Methods

This section presents an in-depth exposition of the proposed methodology for detecting breast cancer. An experimental study approach explored the following research topic: What strategies exist for handling substantial volumes of data in the context of breast cancer detection?

The recommended method uses the pre-trained VGG16 method to improve the accuracy of finding BC and classifying thermal images. It does this by pre-processing through segmentation, augmentation, and normalization, then proposing a pre-train VGG16 Model and binary classification. Figure 2 shows the steps in the recommended procedure. After that, it goes over these processes in more depth.

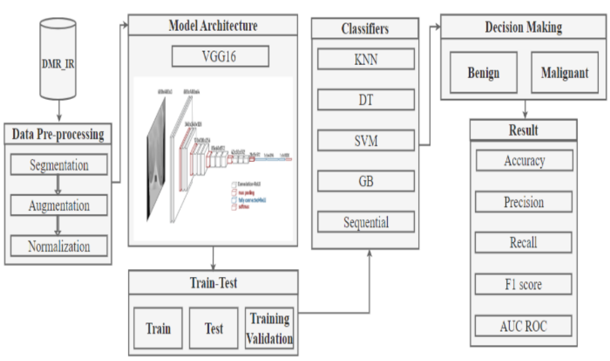


Fig 2. The Architecture of the Proposed Methodology.

3.1 Dataset Description

The DMR-IR dataset, vital for distinguishing between malignant and benign tumors in breast cancer (BC), was acquired from Kaggle, a renowned platform for data collection. This dataset incorporates features essential for diagnosis, explicitly sourced from the University Hospital Antonio Pedro (HUAP) at the Fluminense Federal University of Brazil. The DMR-IR dataset comprises 1246 thermal images captured using a FLIR SC-620 thermal camera, which features a resolution of 680 x 480 and a thermal sensitivity of 45 mK. The recommended examination of studies on the use of thermal images for BC diagnosis revealed that the DMR-IR dataset is the most frequently employed thermal image dataset. The imaging process engaged 282 volunteer women, and this research seamlessly integrated the data into the code directly from the directory and expanded the dataset through augmentation techniques to encompass 20,288 thermal images derived from 282 patients. This comprehensive dataset spans a spectrum of clinical conditions, including healthy and diseased scenarios, contributing to the thorough analysis [25].

3.2 Data Pre-processing

In this phase, the DMR-IR datasets included segmented grayscale images. Partitioning techniques were employed to automatically identify tumor regions before the learning process to reduce computation time. Image quality can be enhanced, and segmentation results can be more accurate using image pre-processing techniques; the study processed the images through augmentation. Finally, normalization is applied, where a large dataset is offered by augmenting all images with labeling for training.

3.2.1 Segmentation

Segmentation in deep learning, particularly using Convolutional Neural Networks (CNNs), is a powerful technique for dividing an image into distinct regions. The segmentation process is a pivotal stage that involves isolating the breast area from the background and highlighting the suspicious region, referred to as the region of interest (ROI), within the broader breast region. Breast image segmentation aims to minimize the background's influence and streamline the detection of abnormalities within the breast area [26]. As shown in Figure 3, segmentation with ROI is demonstrated.

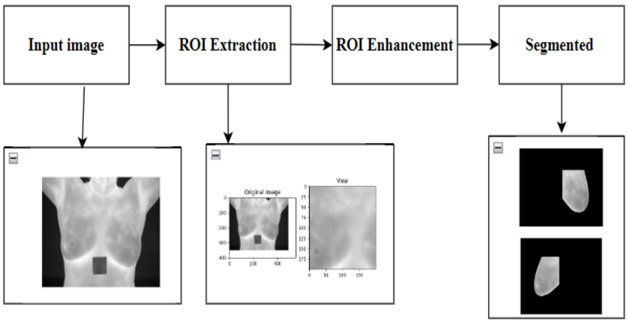


Fig 3. Segmentation with ROI.

3.2.2 Augmentation

Deep learning models demonstrate improved performance when trained on large datasets. Data augmentation is a widely adopted approach to enhance dataset size, which addresses the generalization challenges associated with deep learning. With various augmented versions of the training data, the model learns to recognize variations and generalize its insights to previously unseen images, effectively mitigating overfitting on smaller datasets. This strategy has proven effective in classifying medical images [27]. The study employed the ImageDataGenerator function in Keras to augment the training data. Leveraging this Keras function, various augmentation techniques are applied, such as tilting, zooming, rotating by 20 degrees, width shifting (20%), height shifting (20%), adjusting brightness within the range of 0.5 to 1.5, shear range (20%), flipping the images, and utilizing the features of jpg format. This comprehensive augmentation process contributes to a more robust and diverse training dataset, ultimately enhancing the performance of the deep learning model.

3.2.3 Normalization

Mapping is primarily a process that can automatically assign labels. It functions as a crucial part of the normalization process. Figure 4 shows the Normalization process below.

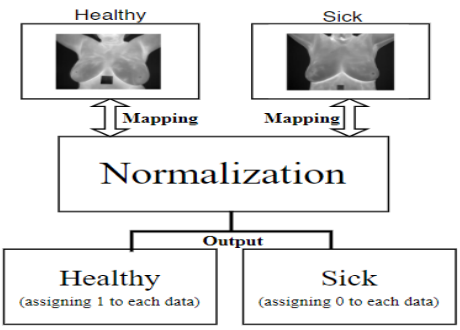


Fig 4. Breast Image Normalization.

3.3 Splitting the directory

The directory split in BC data processing categorizes images into "healthy" and "sick" directories, facilitating machine learning model training. This clear separation improves diagnostic accuracy in BC detection, a significant advancement in medical image processing. In this part, the study split the directory into 54% for healthy and 46% for sick.

3.4 Proposed Pre-train VGG16 model

VGGNet, a convolutional neural network (CNN), was collaboratively developed by the Visual Geometry Group at the University of Oxford and Google DeepMind. The VGG16 pre-trained model is considered an improved version of its predecessor, the AlexNet neural network. Proposed by Karen Simonyan and Andrew Zisserman, the VGG16 model achieved a precision of 92.7% through training on the ImageNet dataset, comprising over 1.4 Crore images [28]. It consists of sixteen layers separated by five max-pooling layers. Each block in the ImageNet weights serves a specific purpose. For the focus on thermal images, Block 5 holds particular significance, containing specific weights related to thermal images. The seized dataset is imperative to construct a customized model for training. The research used the built-in Sequential classifier and employed KNN, SVM, DT, and GB to define this model for classifying benign and malignant cases. The RMSProp optimizer is employed to address the intricacies of gradient descent. Adam needs to be revised to handle biases and correlations within the data, making RMSProp more suitable for focusing on thermal images. The study also chose Binary Cross Entropy for the binary classification task because of its suitability. Monitoring training and validation loss curves played a vital role in preventing models from overfitting or underfitting the data. Early stopping, a technique that halts training when the validation loss stops decreasing, was employed with a patience of 10 epochs. This ensured that training wouldn't continue indefinitely if the model wasn't effectively learning from the validation data. The model summary of the VGG16 pre-trained model with Block 5, intended for achieving high accuracy in breast cancer detection on an extensive dataset, is illustrated in Figure 5 and the model summary in Figure 6.

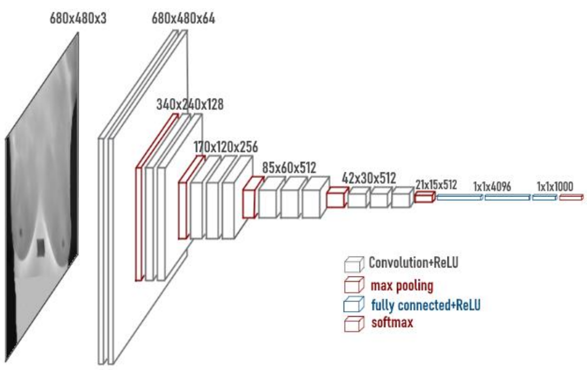


Fig 5. VGG16 Architecture Visualization.

```
Model: "vgg16"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 input_1 (InputLayer)        [(None, 680, 480, 3)]     0

 block1_conv1 (Conv2D)       (None, 680, 480, 64)      1792

 block1_conv2 (Conv2D)       (None, 680, 480, 64)      36928

 block1_pool (MaxPooling2D)  (None, 340, 240, 64)      0

 block2_conv1 (Conv2D)       (None, 340, 240, 128)     73856

 block2_conv2 (Conv2D)       (None, 340, 240, 128)     147584

 block2_pool (MaxPooling2D)  (None, 170, 120, 128)     0

 block3_conv1 (Conv2D)       (None, 170, 120, 256)     295168

 block3_conv2 (Conv2D)       (None, 170, 120, 256)     590080

 block3_conv3 (Conv2D)       (None, 170, 120, 256)     590080

 block3_pool (MaxPooling2D)  (None, 85, 60, 256)       0

 block4_conv1 (Conv2D)       (None, 85, 60, 512)       1180160

 block4_conv2 (Conv2D)       (None, 85, 60, 512)       2359808

 block4_conv3 (Conv2D)       (None, 85, 60, 512)       2359808

 block4_pool (MaxPooling2D)  (None, 42, 30, 512)       0

 block5_conv1 (Conv2D)       (None, 42, 30, 512)       2359808

 block5_conv2 (Conv2D)       (None, 42, 30, 512)       2359808

 block5_conv3 (Conv2D)       (None, 42, 30, 512)       2359808

 block5_pool (MaxPooling2D)  (None, 21, 15, 512)       0

=================================================================
Total params: 14714688 (56.13 MB)
Trainable params: 7079424 (27.01 MB)
Non-trainable params: 7635264 (29.13 MB)
```

Fig 6. Proposed Model Summary.

4. Result and Discussion

The research, using the thermal images from the DMR-IR dataset, used pre-trained VGG16 convolutional neural networks with augmentation and normalization strategies on BC diagnosis thermal images, enhancing data quality and accuracy. Moreover, VGG16 also used a different classifier.

4.1 Experimental Environment

This study employed a workstation with Windows 10 featuring an Intel Core i5-8500 @ 3.00 GHz CPU. The code employed Google Colab, utilizing a system RAM of 12GB, a T4 GPU with a size of 15GB, and a disk size of 78GB. Python Keras was instrumental in constructing the network architecture and conducting training. The Visual Lab website (http://visual.ic.uff.br/en/proeng) provided all the breast cancer thermal image database. Comprehensive details about the dataset are available in Table 2.

Table 2. Dataset Details.

Dataset	**Size**	**Healthy**	**Sick**
DMR-IR-Original	1246	380	740
DMR-IR-Augmented	19042	9140	9901
Total	20288	9520	10641

4.2 Classification Results

The proposed model combines machine learning algorithms and deep learning to pre-process the dataset. Five machine learning classification and regression algorithms, KNN, SVM, DT, GB, and Sequential, are applied to 50% of the training data. The remaining 50% forecasts cancer cell types, assessing the model's performance on unfamiliar data. To evaluate the model's superiority, metrics like the ROC curve are used, with a higher curve indicating better performance. The experiments used Sklearn, Numpy, Pandas, Matplotlib, and Seaborn.

Table 3. Model classification result.

Classifiers	**Acc (%)**	**Sn (%)**	**Sp(%)**	**Rec (%)**	**Prec (%)**	**F1-score (%)**	**AUC (%)**
KNN	82	82.5	93.75	80	80	87	92
SVM	88	100	75	92.9	90.9	95.23	100
DT	90	95	85	97.5	88.63	92.2	93
GB	50	70	80	90	90	87	93
SEQUENTIAL	99.4	100	97.5	99	98.9	98	99.8

(Accuracy = Acc, Sensitivity = Sn, Specificity = Sp, Precision = Prec and Area Under Curve = AUC).

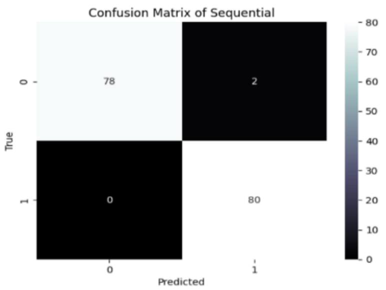
Table 3 compares VGG16 with several classifiers based on various evaluation metrics. The VGG16 model was built with a sequential classifier and had an accuracy(acc) of 99.4%, sensitivity(sn) of 100%, specificity(sp) of 97.5%, recall(rec) of 99%, precision(prec) of 98.9%, f1-score of 98%, and AUC score of 99.8%, which provide high performance than other classifiers.

Fig 7. Performance Measure Indices.

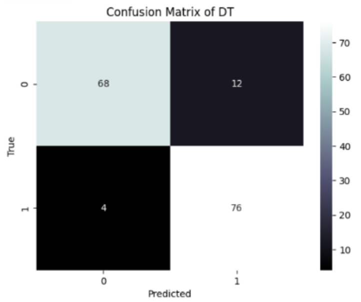
The graphical representation of performance metrics for multiple categories is shown in Figures 7, 8, 9, and 10, providing a comprehensive comparison of their proficiency and effectiveness in a visually informative manner.

Confusion Matrix

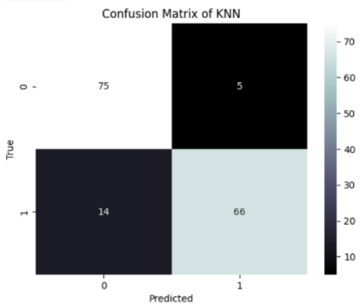
A confusion matrix is a table describing the performance of a classification model based on test data whose original valid values are known. Though it is simply stable, the other parameters could be more apparent. The sequential model achieved the highest accuracy rate of 78%, with only 2 instances of erroneous predictions, as depicted in the following Figure 8. Sequential model correctly predicted 78 images as 0-labeled, while falsely predicting two. For 1-labeled images, the same model correctly predicted 80 images and made no false predictions. DT model correctly predicted 68 images as 0-labeled but made 12 false predictions. It also correctly predicted 66 images as 1-labeled, with 14 false predictions. KNN algorithm correctly predicted 75 images as 0-labeled, with five false predictions, and 66 images as 1-labeled, with 14 false predictions. GB model yielded 64 correct predictions for 0-labeled images and 16 false predictions. For 1-labeled images, it correctly predicted 56 images and made 24 false predictions. Finally, SVM model correctly predicted 60 images as 0-labeled and made 20 false predictions, while correctly predicting all 80 images as 1-labeled.



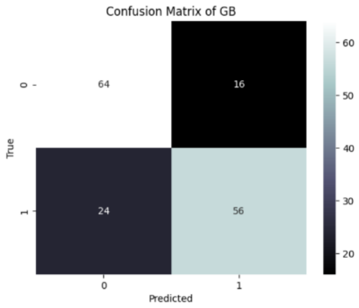
(a)



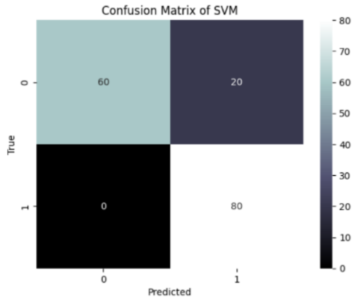
(b)



(c)



(d)



(e)

Fig 8. Confusion Matrices.

Different calculations for the model were performed. The calculations are given below:

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

$$\textbf{Prec} = \frac{TP}{TP+FP} \quad (2)$$

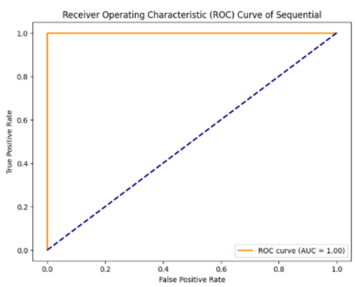
$$\textbf{Rec} = \frac{TP}{TP+FN} \quad (3)$$

$$\textbf{Sp} = \frac{TN}{TN+FP} \quad (4)$$

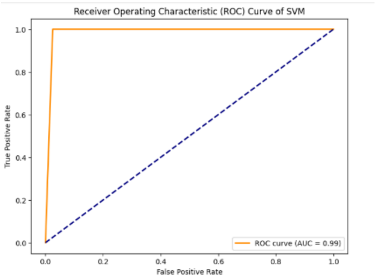
$$\textbf{F1 score} = 2 * \frac{Recall * Precision}{Recall+Precision} \quad (5)$$

(True positives = TP, True Negatives = TN, False positives = FP, False negatives = FN).

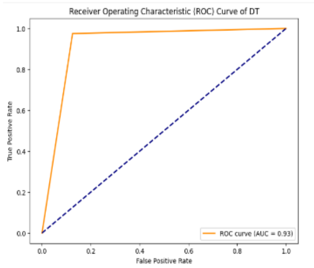
AUC-ROC: AUC primarily calculates the ranking of good prediction, and ROC is a graph illustrating the performance of all models. The Area Under the Receiver Operating Characteristic Curve (AUC-ROC) is determined by the False Positive Rate (FPR) and True Positive Rate (TPR). Utilizing the accurately predicted data alongside the true data, the AUC-ROC value can be approximated.



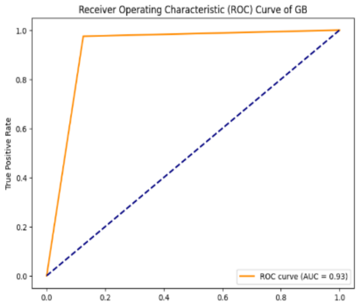
(a)



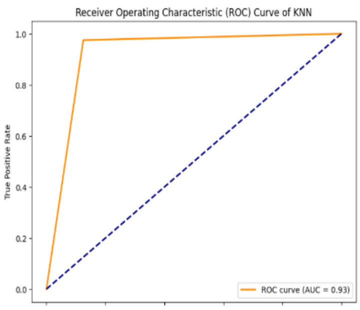
(b)



(c)

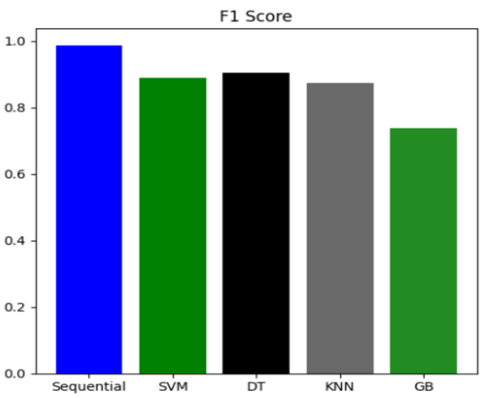


(d)

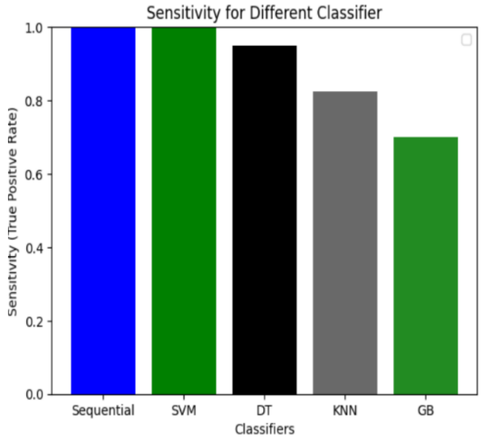


(e)

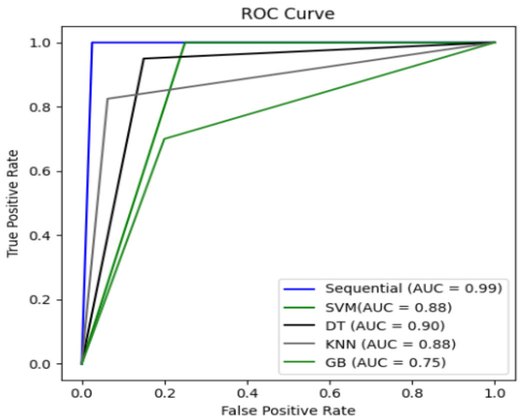
Fig 9. AUC-ROC of Classifiers.



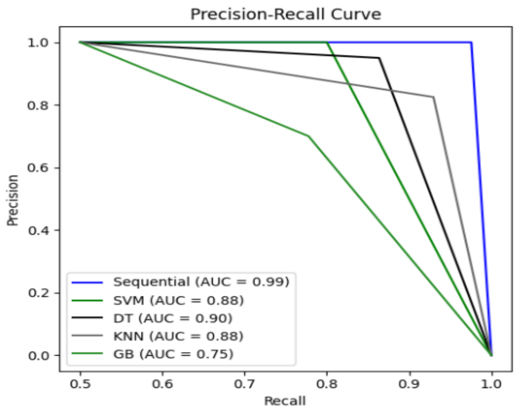
(a)



(b)



(c)



(d)

Fig 10. (a) F1 Score, (b) Sensitivity, (c) ROC Curve, (d) Precision-Recall Curve for evaluating classifiers.

5. Discussions

An interpretable integrated diagnostic approach has been developed and proven effective and accurate, even with limited datasets. Some transfer learning models have also been investigated and compared. The application of transfer learning models for this study also shows complementary results even when compared with the state-of-the-art transfer learning models. These approaches demonstrated high performance compared to the previously published works.

Table 4 presents a comparative analysis of the performance of this study compared to previous research utilizing thermal images. Alia Alshehri et al. introduced two variants of the VGG16 model - one with and one without a Three AM component - aiming at analyzing thermal imaging images for breast cancer diagnosis. Surprisingly, not using AMs on a large dataset achieved superior results compared to the outcomes reported [10].

Seifedine Kadry et al. [11] conducted a study employing VGG16 with the DT method for breast cancer classification, achieving a high accuracy of 95.5% on the Benchmark dataset. However, the proposed model exhibited 99.4% accuracy, indicating a significant improvement in performance over their study. Caroline B. Gonçalves et al. [12] employed three CNN architectures (Resnet, VGG, Densenet), with Densenet yielding the best results. They utilized two distinct learning rates (LR), specifically 0.001 and 0.0001, along with data augmentation in experiment setups 5 and 4, focusing exclusively on frontal images, resulting in outcomes of 84.04 and 87.23, respectively. In contrast, experiment setup 6 exclusively utilized frontal images while maintaining a balanced dataset to prevent data augmentation. In this scenario, using a learning rate of 0.001, combined with Densenet, demonstrated the highest performance at 91.67. This result emphasizes the effectiveness of deep learning models in classifying thermography images. Notably, this model achieved improved results by utilizing a single learning rate and incorporating data augmentation.

In [14], researchers classified breast tumors by utilizing VGG16 and removing the last fully connected layer, employing a by-layer feature fusion technique. They reported an accuracy of 98%. Conversely, the method involved data augmentation using ImageGenerator, in addition to identifying breast cancer tumors through VGG16 with a sequential classifier, resulting in an accuracy of 99.4%, a sensitivity of 100%, a specificity of 97.5%, a precision of 98.9%, an F-score of 98%, an AUC of 99.8, surpassing their results despite the limited data availability.

Subhrajit Dey et al. [15] utilized a pre-trained DenseNet121 with Prewitt and Roberts, achieving an accuracy of 98.8%, which is lower than the performance achieved in the study. In [29], they formerly utilized Dynamic Infrared Thermography (DIT) and Static Infrared Thermography (SIT) protocols, employing temperature time series to detect patient abnormalities and selecting texture features through a genetic algorithm. However, the research showed satisfactory results without using these protocols. Overall, the study affirms the approach's efficacy, showcasing its superiority over alternative methods and highlighting its potential for robust medical image analysis, particularly when handling substantial datasets.

Table 4. Comparison with other existing state-of-the-art models of Thermal images.

Reference	**Approach**	**Datasets**	**Acc (%)**	**Sn (%)**	**Sp (%)**	**Prec (%)**	**F1-score (%)**	**AUC (%)**
Alia Alshehri et al. [10]	VGG16+AMs (Deep Attention Mechanisms)	DMR-IR	99.80 (AMs), 99.18 (without AMs)					
Seifedine Kadry et al. [11]	VGG16+DT	Benchmark dataset	95.5	97.97	93.07	93.27	95.56	
Caroline B. Gonc et al. [12]	VGG-16, Densenet201, and Resnet502	Thermo-graphy	91.67	100	83.3		92	
Devanshu Tiwari et al. [14]	VGG16, VGG 19, ResNet50V2, InceptionV3	DMR-IR	98	98.48	100	100	99.01	

Subhrajit Dey et al. [15]	DenseNet121+ (Prewitt. and Roberts), VGG16, VGG19, DenseNet169	DMR-IR	98.8	98		99	98.5	
Roger Resmini et al. [29]	Fractals, Wavelet, 8 LTP, GLCM	DMR-IR	95	95		95		
Our proposed model	VGG16+ (KNN, SVM, DT, GB, SEQUENTIAL)	DMR-IR	99.4	100	97.5	98.9	98	99.8

6. Conclusion

Early detection of breast cancer (BC) is pivotal in saving lives, regardless of gender. Preventative screening is essential for the early diagnosis and treatment of BC, and some countries have effectively started screening programs that have reduced the disease burden by almost one-third [30]. Initiatives aimed at this endeavor are instrumental in empowering both patients and healthcare professionals with comprehensive information. The augmentation and normalization techniques applied across various metrics have showcased the superior performance of the VGG16 model. In addition to the VGG16 model, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and decision trees have proven valuable in aiding early BC detection, exhibiting commendable performance in certain evaluations. These methods have been incorporated into the study to enhance the accuracy and consistency of the VGG16 model. Notably, this model effectively addresses the limitations identified in prior research by leveraging an expanded dataset. Furthermore, a comparison with several related studies underscores the remarkable results achieved, boasting an impressive accuracy of 99.4%.

However, it's imperative to acknowledge the limitations encountered during the investigation, primarily stemming from the restricted availability of thermal breast imaging datasets and the prevalence of low-quality images. Despite these challenges, the study adeptly tackles dataset size constraints through the implementation of augmentation methods.

Moving forward, future endeavors aim to expand upon this study by conducting experiments with significantly larger datasets, amalgamating data from diverse sources, and exploring various augmentation techniques. Subsequent evaluations will encompass diverse biomedical imaging datasets to ascertain the robustness and generalizability of the extended study. Additionally, future research avenues include investigating alternative pre-trained deep learning models across numerous datasets and leveraging high-resolution cameras to capture cancerous tumors through colored thermal images.

Acknowledgements

I am grateful to all of those with whom I have had the pleasure to work during this and other related Research Work. Each of the members of my Dissertation Committee has provided me extensive personal and professional guidance and taught me a great deal about both scientific research and life in general

Conflicts of interest

The authors declare no conflicts of interest.

Funding Details

No funding was received to assist with the preparation of this manuscript.

References

[1] Rahman MA, Muniyandi Rc, Albashish D, Rahman MM, Usman OL. 2021. Artificial neural network with Taguchi method for robust classification model to improve classification accuracy of breast cancer. PeerJ Computer Science 7:e344 https://doi.org/10.7717/peerj-cs.344.

[2] Analysis and Detection of Breast Cancer Using Various Machine Learning Strategies. (2021,

December 17). IEEE Conference Publication | IEEE Xplore. https://doi.org/10.1109/ICCCA52192.2021.9666436.

[3] Khalid, A., Mehmood, A., Alabrah, A., Alkhamees, B. F., Amin, F., AlSalman, H., & Choi, G. S. (2023, October 2). Breast Cancer Detection and Prevention Using Machine Learning. Diagnostics. https://doi.org/10.3390/diagnostics13193113.

[4] Mahoro, E.; Akhloufi, M.A. Applying Deep Learning for Breast Cancer Detection in Radiology. Curr. Oncol. 2022, 29, 8767–8793, https://doi.org/10.3390/curroncol29110690.

[5] A Systematic Review of Breast Cancer Detection Using Thermography and Neural Networks. (2020). IEEE Journals & Magazine | IEEE Xplore. https://doi.org/10.1109/ACCESS.2020.3038817.

[6] Aidossov, N., Zarikas, V., Mashekova, A. S., Zhao, Y., Ng, E. Y. K., Midlenko, A., & Mukhmetov, O. (2023, January 1). Evaluation of Integrated CNN, Transfer Learning, and BN with Thermography for BC Detection. Applied Sciences https://doi.org/10.3390/app13010600.

[7] Abdel-Nasser, M., Moreno, A., & Puig, D. (2019, January 16). BC Detection in Thermal Infrared Images Using Representation Learning and Texture Analysis Methods. Electronics https://doi.org/10.3390/electronics8010100.

[8] Mammoottil, M. J., Kulangara, L. J., Cherian, A. S., Mohandas, P., Hasikin, K., & Mahmud, M. (2022, February 28). Detection of Breast Cancer from Five-View Thermal Images Using Convolutional Neural Networks. Journal of Healthcare Engineering. https://doi.org/10.1155/2022/4295221.

[9] Goceri, E. (2023). Medical image data augmentation: techniques, comparisons and interpretations. Artificial Intelligence Review, 56(11), 12561–12605. https://doi.org/10.1007/s10462-023-10453-z

[10] Alshehri, A., & AlSaeed, D. (2023, February 23). BC Diagnosis in Thermography Using Pre-Trained VGG16 with Deep Attention Mechanisms. Symmetry https://doi.org/10.3390/sym15030582.

[11] Kadry, S., Crespo, R. G., Herrera-Viedma, E., Krishnamoorthy, S., & Rajinikanth, V. (2023). Classification of Breast Thermal Images into Healthy/Cancer Group Using Pre-Trained Deep Learning Schemes. Procedia Computer Science, 218, 24–34. https://doi.org/10.1016/j.procs.2022.12.398

[12] C. B. Gonçalves, J. R. Souza and H. Fernandes, "Classification of static infrared images using pre-trained CNN for breast cancer detection," 2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS), Aveiro, Portugal, 2021, pp. 101-106, doi: 10.1109/CBMS52027.2021.00094.

[13] Deep CNN Model based on VGG16 for BC Classification. (2021, July 14). IEEE Conference Publication | IEEE Xplore https://doi.org/10.1109/ICIT52682.2021.9491631.

[14] Tiwari, D., Dixit, M., & Gupta, K. K. (2021, December 31). Deep Multi-View BC Detection: A Multi-View Concatenated Infrared Thermal Images Based BC Detection System Using Deep Transfer Learning. Traitement Du Signal https://doi.org/10.18280/ts.380613.

[15] Dey, S., Roychoudhury, R., Malakar, S., & Sarkar, R. (2022, January 8). BC screening from thermogram images by edge detection aided the deep transfer learning model. Multimedia Tools and Applications https://doi.org/10.1007/s11042-021-11477-9.

[16] Alruwaili, M., & Gouda, W. (2022, January 24). Automated BC Detection Models Based on Transfer Learning. Sensors https://doi.org/10.3390/s22030876.

[17] A Novel Deep-Learning Model for Automatic Detection and Classification of Breast Cancer Using the Transfer-Learning Technique. (2021). IEEE Journals & Magazine | IEEE Xplore https://doi.org/10.1109/ACCESS.2021.3079204.

[18] Multi-Input Deep Learning Approach for BC Screening Using Thermal Infrared Imaging and Clinical Data. (2023). IEEE Journals & Magazine | IEEE Xplore https://doi.org/10.1109/ACCESS.2023.3280422.

[19] Rafiq, A., Chursin, A., Alrefaei, W. A., Alsenani, T. R., Aldehim, G., Samee, N. A., & Menzli, L. J. (2023, May 11). Detection and Classification of Histopathological Breast Images Using a Fusion of CNN Frameworks. Diagnostics https://doi.org/10.3390/diagnostics13101700.

[20] Al-Qazzaz, N. K., Mohammed, I. K., Al-Qazzaz, H. K., Ali, S. H. M., & Ahmad, S. A. (2023, November 3). Comparison of the Effectiveness of Various Classifiers for BC Detection Using Data Mining Methods. Applied Sciences https://doi.org/10.3390/app132112012.

[21] Aidossov, N., Zarikas, V., Mashekova, A. S., Zhao, Y., Ng, E. Y. K., Midlenko, A., & Mukhmetov, O. (2023, January 1). Evaluation of Integrated CNN, Transfer Learning, and BN with Thermography for Breast Cancer Detection. Applied Sciences https://doi.org/10.3390/app13010600.

[22] Minnoor, M., & Baths, V. (2023). Diagnosis of Breast

Cancer Using Random Forests. Procedia Computer Science, 218, 429–437. https://doi.org/https://doi.org/10.1016/j.procs.2023.01.025

[23] Mohamed, E., Rashed, E., Gaber, T., & Karam, O. (2022). Deep learning model for fully automated breast cancer detection system from thermograms. PLOS ONE, 17, e0262349. https://doi.org/10.1371/journal.pone.0262349

[24] Abdel-Nasser, M., Moreno, A., & Puig, D. (2019, January 16). BC Detection in Thermal Infrared Images Using Representation Learning and Texture Analysis Methods. Electronics https://doi.org/10.3390/electronics8010100.

[25] Tsietso, D., Yahya, A., Ravi, S., Tariq, M. U., Babar, M., Qureshi, B., & Koubaa, A. (2023). Multi-Input Deep Learning Approach for Breast Cancer Screening Using Thermal Infrared Imaging and Clinical Data. IEEE Access, PP. https://doi.org/10.1109/ACCESS.2023.3280422

[26] Ben Ammar, M., Ayachi, F. L., Cardoso de Paiva, A., Ksantini, R., & Mahjoubi, H. (2024). Harnessing Deep Learning for Early Breast Cancer Diagnosis: A Review of Datasets, Methods, Challenges, and Future Directions. International Journal of Computing and Digital Systems, 15(1), 1–13. https://journal.uob.edu.bh:443/handle/123456789/5337

[27] Goceri, E. (2023). Medical image data augmentation: techniques, comparisons and interpretations. Artificial Intelligence Review, 56(11), 12561–12605. https://doi.org/10.1007/s10462-023-10453-z

[28] Tiwari, D., Dixit, M., & Gupta, K. K. (2021, December 31). Deep Multi-View Breast Cancer Detection: A Multi-View Concatenated Infrared Thermal Images Based Breast Cancer Detection System Using Deep Transfer Learning. Traitement Du Signal. https://doi.org/10.18280/ts.380613

[29] Resmini, R., Faria da Silva, L., Medeiros, P. R., Araujo, A. S., Muchaluat-Saade, D. C., & Conci, A. (2021). A hybrid methodology for breast screening and cancer diagnosis using thermography. Computers in Biology and Medicine, 135, 104553, https://doi.org/10.1016/j.compbiomed.2021.104553.

[30] Abdar, M., Zomorodi-Moghadam, M., Zhou, X., Gururajan, R., Tao, X., Barua, P. D., & Gururajan, R. (2020). A new nested ensemble technique for automated diagnosis of breast cancer. Pattern Recognition Letters, 132, 123–131. https://doi.org/https://doi.org/10.1016/j.patrec.2018.11.004