

Quantitative InfraRed Thermography Journal

ISSN: 1768-6733 (Print) 2116-7176 (Online) Journal homepage: www.tandfonline.com/journals/tqrt20

Breast cancer classification on thermograms using deep CNN and transformers

Ella Mahoro & Moulay A. Akhloufi

To cite this article: Ella Mahoro & Moulay A. Akhloufi (2024) Breast cancer classification on thermograms using deep CNN and transformers, Quantitative InfraRed Thermography Journal, 21:1, 30-49, DOI: [10.1080/17686733.2022.2129135](https://doi.org/10.1080/17686733.2022.2129135)

To link to this article: <https://doi.org/10.1080/17686733.2022.2129135>



Published online: 28 Sep 2022.



Submit your article to this journal



Article views: 924



View related articles



View Crossmark data



Citing articles: 35 [View citing articles](#)



Breast cancer classification on thermograms using deep CNN and transformers

Ella Mahoro  and Moulay A. Akhloufi 

Perception, Robotics and Intelligent Machines Research Group (PRIME), Department of Computer Science, Université de Moncton, 18 Antonine-Maillet Ave, Moncton, NB E1A 3E9, Canada

ABSTRACT

Breast thermography is a screening approach for breast cancer detection by measuring the breast skin temperature. Breast cancer is the most common cancer among women and can affect either women or men. Its early diagnosis and treatment reduce deaths and increase survival chances. The use of deep learning algorithms and techniques has made it easier to detect breast cancer in its early stages, but some challenges remain. In this work, we propose a breast cancer detection system. In the first step, TransUNet, a vision-based Transformer, is used to segment the breast region and separate it from the rest of the body. In the second step, four different models such as EfficientNet-B7, ResNet-50, VGG-16 and DenseNet-201 are used to classify the dataset into three types: healthy, sick, and unknown. The best accuracy achieved is 97.26%, sensitivity of 97.26% and specificity of each class healthy, sick, unknown is 100%, 96.94% and 99.72% respectively with the ResNet-50 model.

ARTICLE HISTORY

Received 21 April 2022

Accepted 22 September 2022

KEYWORDS

Breast cancer; thermogram; convolutional neural network; segmentation; classification

1. Introduction

Breast cancer is a cancer that grows in the cells of the breasts and can occur in both men and women, but it is more common in women [1].

Studies have found mutations in two genes known as breast cancer (BRCA) genes, BRCA1 and BRCA2, are associated with 5% to 10% of breast cancers [2]. Gene mutations and the absence of alleles in the BRCA1 locus have been demonstrated to lead to early breast cancer [3]. However, these genes can be damaged during adulthood by toxins, radiation, and other chemicals such as free radicals [4]. Radiologists' main role is to make accurate differential diagnoses based on the medical images of their patients and a wide range of applications use image classification, from identifying disease to determining malignancy type [5].

Mammography is the most common method used to detect breast cancer [6] but it also has some drawbacks. It presses on the breast to spread the breast tissue and increase the clarity of the x-ray image [7] but however this can be uncomfortable and painful for the patient. Besides mammography, there are many other methods used to screen breast

CONTACT Ella Mahoro  eem2154@umoncton.ca  Perception, Robotics and Intelligent Machines Research Group (PRIME), Department of Computer Science, Université de Moncton, Moncton, Canada

This article has been corrected with minor changes. These changes do not impact the academic content of the article.

© 2022 Informa UK Limited, trading as Taylor & Francis Group

cancer such as thermography, ultrasonography, computed tomography, magnetic resonance imaging (MRI) [8].

The breast thermography, also called as infrared imaging of the breast, is an illustrated representation of the infrared energy of the breasts [9]. It observes the temperature of the breast skin of a patient over breast cancer. The images produced by thermography are called thermograms. In addition to being painless, thermography requires no radiation exposure, so many choose it over mammography [10]. Breast cancer can be detected through thermography 8 to 10 years before it can be detected by mammography [11]. However, its prevention is still a challenge, so efforts to increase early detection of the disease are the main focus in the fight against it [12].

Therefore, to ensure a high patient survival rate, early detection of breast cancer is essential. According to this study [13], the survival rate is lower for a delay of 3–6 months. It cannot be explained by lead-time bias. The delay in treatment may turn the cancer into a more advanced stage and cause fewer chances for survival. Keeping delays to a minimum should be the goal of patients.

There is growing interest in deep learning, which has been used for a number of applications and offers potential in the field of medical research as well [14]. By combining layers of neural networks, deep learning algorithms can identify patterns and improve the accuracy of imaging-based classifiers in various medical applications. There is still a challenge in correctly identifying breast cancer despite continuous research in automated breast cancer application [15]. Further, large training data for deep learning are not easily available in the medical field and there is also a problem of imbalanced class distribution [16]. In this study, we propose a system to perform the image classification of dynamic thermograms by:

- (1) Training the thermograms and their mask with TransUNet using a pre-trained model ResNet-50
- (2) Applying the full database on the pre-trained TransUNet model and removing the background
- (3) Classifying the database using deep convolutional neural networks

The structure of this paper is as follows: [section 2](#) provides a detailed review of the state of the art, [section 3](#) describes the dataset and explains the proposed method, [section 4](#) contains the experimental results and the paper is discussed in [section 5](#) and concludes in [section 6](#).

2. Related work

Research has shown that early detection and treatment of breast cancer prevents the spread of malignant cells throughout the body [17]. An individual can conduct a breast self-exam by pressing on the breast and check if there are any changes [18]. unfortunately, it is not a very reliable method of detecting cancer. To detect breast cancer, low-energy x-rays called mammography have been used since 1960 [19]. However, a high proportion of false positives result from relying only on mammograms, which can be a risk on patient health and result in unnecessary biopsies and surgeries [20]. A number of computer-aided systems have been developed to detect cancer at its early stage. There have been many

applications built using machine learning algorithms like random forest, decision tree, KNN, SVM, naïve Bayes etc. Additionally, deep learning is now being used to classify breast cancer cases.

2.1. Classification

The aim of the study [21] was to classify thermograms into benign, malignant and cyst categories based on the behaviour of different classification methods. The classification was performed utilising a variety of methods, including Artificial Neural Networks (ANN), Multi-Layer Perceptron (MLP), Extreme Learning Machines (ELM), Decision Trees (DT) and Bayesian classifiers. The MLP classifier provided highest accuracy of 76.01%, sensitivity of 78% and specificity of 88%. Using automated analysis methods, Torres-Galván *et al.* [22] evaluated deep neural networks capabilities to classify thermograms. The system was based on VGG-16 convolutional neural network and using a dataset of 173 thermograms, they obtained an accuracy of 91.18%, sensitivity of 100%, specificity of 82.35% and an AUC of 84.52%. Ekici *et al.* [23] presented a new algorithm for feature extraction based on bio-data, image analysis and image statistics. Using a CNN optimised by the Bayesian algorithm, the thermograms were classified into normal or suspected. According to their proposed algorithm, they achieved 98.95% accuracy.

Krawczyk *et al.* [24] proposed a method for managing imbalanced datasets by using a multiple classifier system based on a set of asymmetry characteristics using 150 thermograms. Different base classifiers of the ensemble were trained on different object subspaces and the best result showed an accuracy of 90.03%, a sensitivity of 80.35% and a specificity of 90.15%. Husaini *et al.* [25] developed classifiers based on Inception V3, Inception V4 and a modified version of Inception MV4 models. Thermograms were classified into healthy and sick using the proposed method on the DMR database. Four effects were applied to the thermograms to ensure accuracy: Blur, Shaken, Tilted and Flipping. In greyscale images, Inception V3 outperforms V4 and MV4. As result, they found that detection accuracy improved by 0.0002% in blurred images and shook images and decreased by nearly 11% for tilted images. Baffa *et al.* [26] proposed a methodology based on convolutional neural networks (CNN) to classify thermograms into healthy or unhealthy. In the Database for Mastology Research, two different protocols are used to acquire images, the static protocol and the dynamic protocol. The images acquired from two different protocols are used to evaluate the proposed method. Their proposed model achieved 98% of accuracy for the static protocol and 95% for the dynamic protocol. In the case of the dynamic protocol, the sensitivity, specificity, and precision obtained are 97%, 93%, and 97% respectively using 2740 thermograms. Chebbah *et al.* [27] proposed a computer-aided diagnosis system based on automatic segmentation and classification of thermograms. A total of 170 thermograms were segmented using the U-Net model to extract the region of interest and an Intersection over Union (IoU) result of 89.03% was obtained. Various models are used to classify normal from abnormal thermograms, such as K-Nearest Neighbour (KNN), AdaBoost, Random Forest, and Support Vector Machine (SVM). With an accuracy of 94.4%, a precision of 96.2%, a sensitivity of 86.7%, and an F1-score of 91.2%, SVM achieved the best results. Torres-Galván *et al.* [28] proposed an approach based on the automatic classification of thermograms into normal and abnormal using a deep convolutional neural network (CNN) with transfer learning. A total of 311

patients were analysed using two different approaches having respectively balanced and unbalanced class distributions. With an unbalanced distribution, the ResNet-101 model obtained a sensitivity of 84.6%, a specificity of 65.3% and a balanced accuracy of 74.9%, while with a balanced distribution, these values were 92.3% sensitivity and 53.8% specificity and a balanced accuracy of 73.1%.

2.2. Segmentation and classification

Pramanik *et al.* [29] proposed a system that segments the breast region of 306 thermograms in three steps: background removal, inframammary fold detection and axilla detection. The background of the image is removed by using otsu's thresholding and grey level reconstruction techniques. To find out that lower limit in breast by removing the other part of the body, the inframammary fold region of the lower part of the breast is hypothesised to be where the maximum temperature occurs. The curvature point of the top-half breast image is used to determine the upper limit of the breast. They employed a feature extraction technique based on multi-resolution analysis on breast region and these features are sent to the feed-forward neural network classifier to classify the thermograms into 2 classes: healthy and unhealthy. In their proposed system, the result showed an accuracy of 90.48%, sensitivity and specificity are 87.6% and 89.73% respectively. Ma *et al.* [30] developed a breast cancer detection system based on smart phone with infrared camera. A total of 1520 patients are included in the dataset, of which 760 are normal and 760 are patients with breast cancer. The proposed portable detection system includes different modules such as data acquisition to capture a thermogram, greyscale processing to convert the image to greyscale, segmentation and flipping to isolate the ROIs of the left and right breasts, greyscale compression to reduce the grey level of the original image, GLCM-based feature extraction and K-Nearest Neighbour (KNN) classifier for breast cancer detection. Using the KNN classifier, the proposed system achieved an accuracy, sensitivity, and specificity of 99.21%, 98.4% and 100% respectively. In [31], Milosevic *et al.* proposed a system with a feature extraction, classification, and segmentation. There are 40 thermograms in the dataset used, of which 20 GLCM features were extracted. Support Vector Machine classifier, Naive Bayes classifier and K-Nearest Neighbour classifier were used to differentiate abnormal from normal tissues. In their study, receiver operating characteristic analysis and five-fold cross validation were used to evaluate the classification performance. The K-Nearest Neighbour classifier was the most accurate with 92.5% accuracy. The thermograms are segmented after classification. Results showed that the proposed method was capable of providing accurate tumour shapes. Using a Genetic Algorithm (GA) and Support Vector Machine (SVM) classifier, Resmini *et al.* [32] proposed an ensemble method for selecting models and features for breast cancer diagnosis. The proposed model was evaluated on two datasets, the DMR-IR database and the private thermogram database of the Federal University of Pernambuco (UFPE), each having 80 and 98 examinations respectively. Four - experiments were constructed by segmenting these thermograms using different approaches (with and without armpits). These experiments evaluated the behaviour of proposed methods for detecting breast cancer in different thermograms. This approach yielded an AUC, accuracy, sensitivity, and specificity of 94.79%, 97.18%, 97.18% and 94.79% for detection of breast cancer respectively. Mohamed *et al.* [33] presented a fully automatic system for

breast cancer detection. U-Net network is used in the proposed method for extracting, isolating the breast area of 1000 thermograms and classifying them into normal or abnormal breast tissues using CNN model. The proposed system achieved an accuracy, sensitivity, and specificity of 99.33%, 100% and 98.67% respectively.

Following [Table 1](#) summarises the state-of-the-art methods and results.

3. Materials and methods

3.1. Various algorithms and techniques are outlined in this section

Architecture of the proposed solution The proposed system integrates segmentation and classification. A Transformers-based U-Net framework called TransUNet was used to segment breast areas, which make strong encoders for medical image segmentation tasks [34]. The breast area is automatically extracted and isolated from other parts of the body. Furthermore, we compare the classification performance of the proposed system with four different deep learning models. The novelty of the proposed system resides in using TransUNet network for automating the segmentation process and classifying the output of TransUNet using deep learning models. [Figure 1](#) is a diagram illustrating various aspects of the proposed breast cancer classification system.

3.2. Dataset

The data used in this research is collected from DMR-Database for Mastology Research-Visual Lab, UFF, Niteroi, Brazil [1]. This dataset was taken by FLIR SC- 620 thermal camera with a spatial resolution of 480×640 pixels and are available in jpg format. At the moment the image was captured, the body temperature of the patient was recorded. There are two categories of body heat transfer behaviour that was taken into account for the image acquisition: static and dynamic. Static images are recorded after 10 to 15 minutes of thermal stabilisation during the patient resting [35]. On the other hand, the dynamic one consists of a series of thermograms captured every 15 seconds during five minutes. Therefore, the database was splitted into 2 categories dynamic and static

Table 1. Some of the state-of-the-art methods using thermograms.

Ref.	Methodology	Database (thermograms)	Accuracy (%)	Sensitivity (%)	Specificity (%)
[33]	U-Net + classification using CNN	1000	99.33	100.00	98.67
[30]	K-Nearest Neighbour classifier	1520	99.21	98.40	100.00
[23]	CNN optimised by the Bayesian algorithm	3895	98.95	-	-
[32]	Support Vector Machine classifier	178	97.18	97.18	94.79
[26]	Classification using CNN	2740	95.00	97.00	93.00
[27]	U-Net + Support Vector Machine classifier	170	94.40	86.70	-
[31]	K-Nearest Neighbour classifier	40	92.50	-	-
[22]	Automatic classification	173	91.18	100.00	82.35
[29]	Wavelet based thermogram analysis	306	90.48	87.60	89.73
[24]	Classifier ensembles system	150	90.03	80.35	90.15
[21]	Multi-Layer Perceptron classifier	825	76.01	78.00	88.00
[28]	Classification using ResNet-101	311	73.10	92.30	53.80

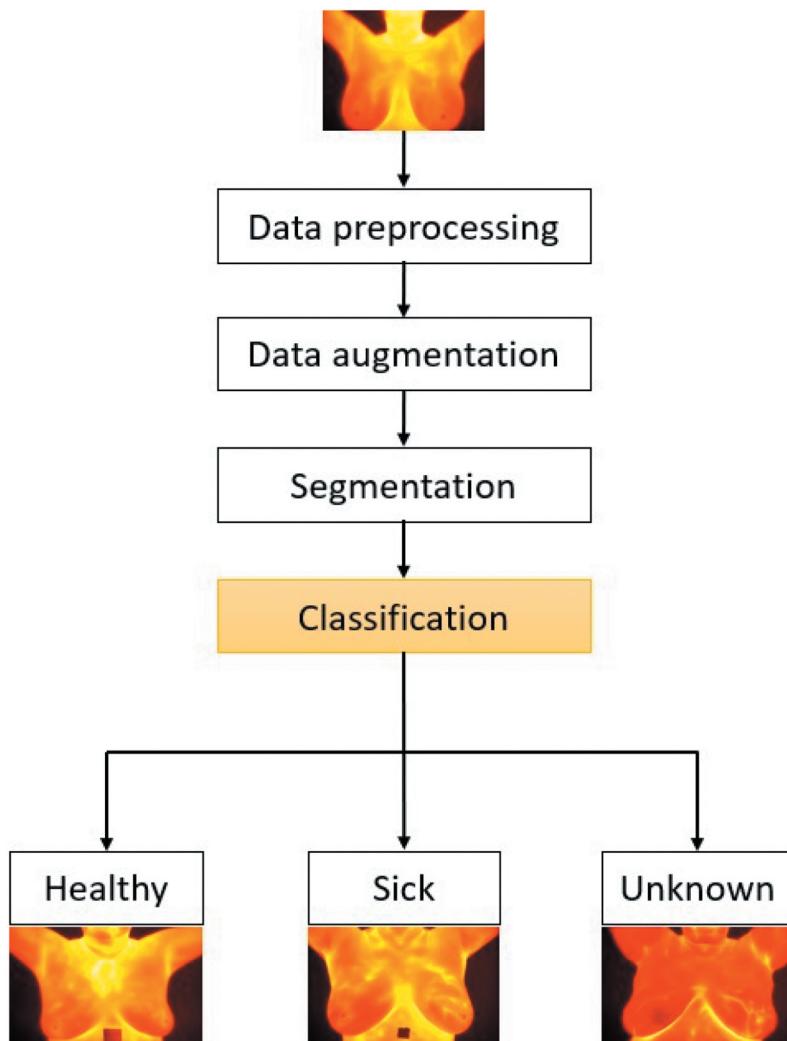


Figure 1. Architecture of the proposed model.

thermograms. In our research, dynamic thermograms were used. The dynamic protocol generates the front and lateral positions of sequential images for each patient.

An example of a healthy, sick, or unknown thermograms is shown in [Figure 2](#).

3.3. Data preprocessing

Within the same class, all the thermograms of each patient are merged with those of other patients. The data are splitted into 3 parts train, validation, and test sets, each having 75%, 22% and 3% of the dataset respectively. The total numbers of thermograms in the train, validation and test sets are: 2891, 732 and 364 respectively. Inside each folder, there are 3 classes healthy, sick, and unknown. The thermograms are also resized into 224 × 224 pixels. The [Table 2](#) shows the number of thermograms into each class. This paper uses

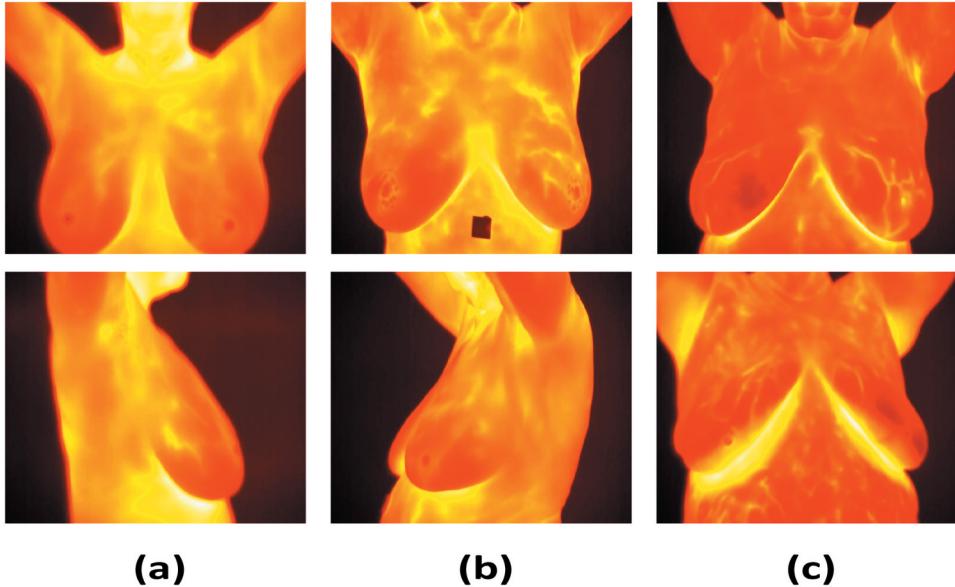


Figure 2. Sample thermograms from the breast cancer dataset (a): healthy; (b): sick; (c): unknown.

Table 2. Number and percentage of thermograms based on classes.

Class	No. of thermograms	Percentage
Healthy	3089	77.44
Sick	856	21.46
Unknown	44	1.10
Total	3989	100

a variety of techniques to handle the imbalanced data by augmenting the data, pre-trained weights from ImageNet and applying the appropriate deep learning models and evaluation metrics.

3.4. Data augmentation

Data augmentation is used to improve the performance of our model. It creates out of our dataset many different examples for training. By using keras preprocessing layers to augment the images, we rotate the images by a 10 degree angle, shift the image vertically and flipped the image horizontally. This step will prevent the overfitting and increase the generalisation of our model on the test data.

3.5. Segmentation

Initially, we manually created masks for 446 thermograms selected at random. These thermograms and their mask will be used to train a TransUNet model. The thermograms are divided into two folders, train and test, each containing 390 and 10 thermograms.

3.5.1. TransUnet

In medical image segmentation, TransUNet, a Transformers-based U-Net framework, is comprised of both the transformers and U-Net network [34]. By using a CNN to extract local features and Transformers to encode global context of the features, the model provided the best performance. Their performance in natural language processing and image recognition is exceptional. They are designed for sequence-to-sequence prediction and rely only on attention mechanisms, but their insufficient low-level details can limit their ability to provide localisation. Previous results obtained using Transformer-based models shows that they perform better than CNN-based models in general. The Transformer encoder consists of 12 Transformer layers, each layer includes Multihead Self-Attention (MSA), a normalisation layer and Multi-Layer Perceptron (MLP). CNN units extract features from input images based on different depth levels and layers and the Transformer unit receive the linear projections of extracted features and provide the global information. The skip-connections from the CNN unit, which act as an encoder and the output features of the Transformer unit are feeding the decoder. The decoder consists of 3-skip-connections and ReLU activations and its output is fed for segmentation. [Figure 3](#) presents the TransUNet architecture. As a result, we obtained a F1-score value of 90.92%. [Figure 4](#) presents the mask predicted by this model.

3.5.2. Extraction

The pretrained TransUNet is employed on the full database and provides a mask predicted for each thermogram. In this section, our main objective is to completely remove the background, by subtracting the thermograms from their mask. We remain only with the breast region. This new database will be mixed with the DMR-IR database. [Figure 5](#) presents the region of interest extracted from the dataset.

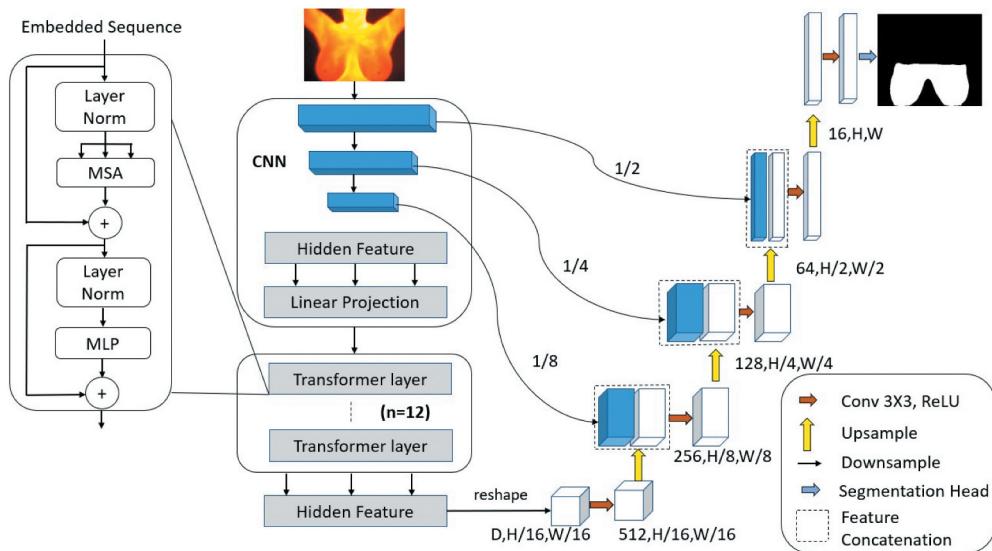


Figure 3. Architecture of TransUnet [34]. .

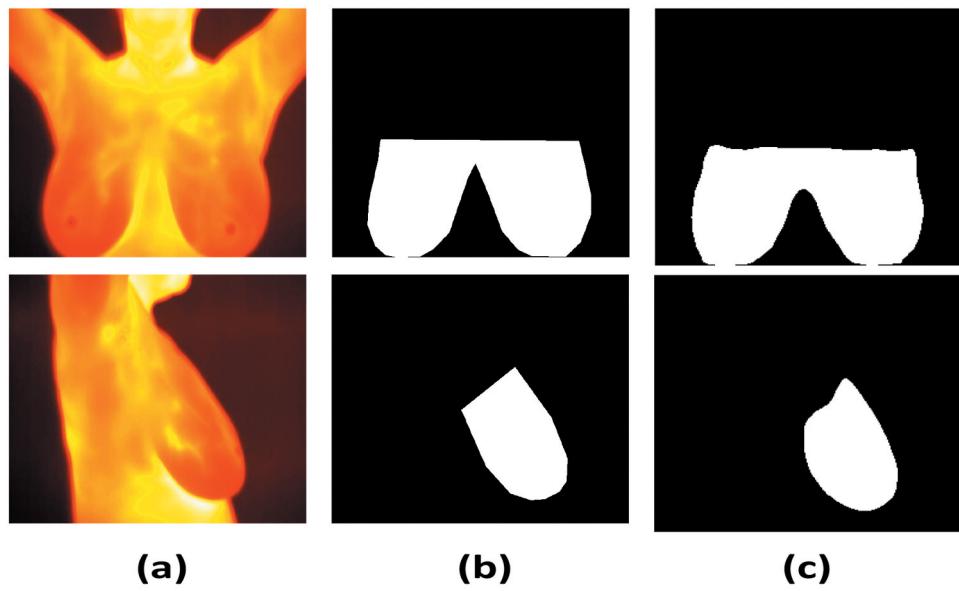


Figure 4. TransUnet mask prediction from the breast cancer dataset (a): thermogram; (b): ground truth; (c): mask predicted.

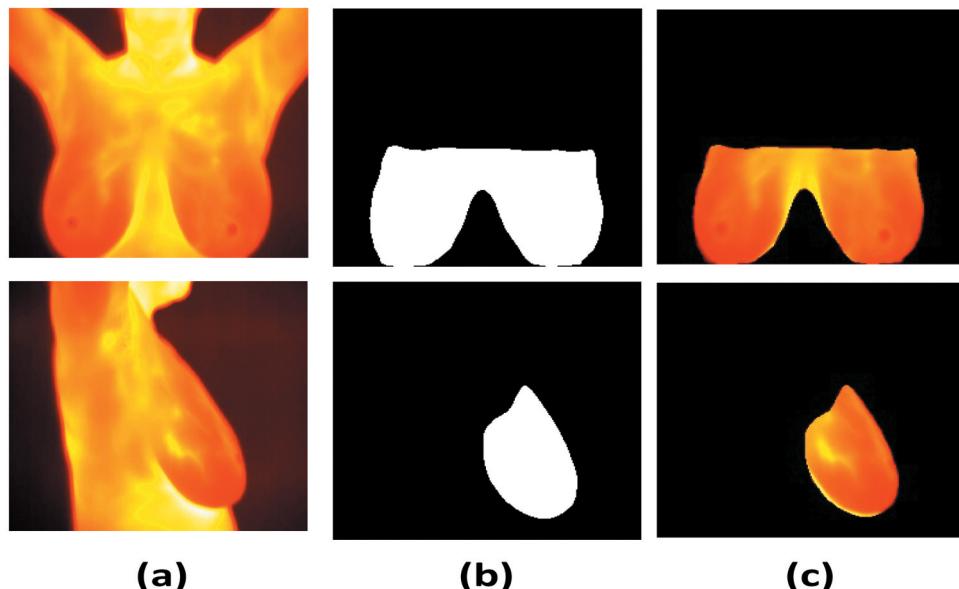


Figure 5. After the semantic segmentation: (a) thermogram; (b) mask predicted; (c) thermogram extracted.

3.6. Classification

Different models are selected for breast cancer classification based on their performance.

3.6.1. EfficientNet-B7

The EfficientNet is an architecture that scales all depth, width, and resolution dimensions uniformly across the convolutional neural networks. Tan *et al.* [36] propose a way to balance all the 3 dimensions of network by scaling each of them. A compound scaling was introduced in order to increase the accuracy and efficiency. In our architecture, the EfficientNet-B7 consists of different blocks, a 3×3 convolution layer followed by 3×3 and 5×5 inverted residual blocks layers such as MBConv1, MBConv6 and a softmax layer activation function. Figure 6 presents the EfficientNet-B7 architecture [37].

3.6.2. ResNet-50

In computer vision, ResNet, or residual networks, is used as the backbone in many algorithms. The fundamental achievement with ResNet was its ability to train highly complex neural networks with more than 150 layers. In this research, we will use ResNet-50. ResNet-50 model is a variant of ResNet model which consists of 50 layers of blocks with short-cuts connections. It has one convolutional layer, a batch normalization layer, 16 residual modules in between two pooling layers and perform spatial convolution using 3×3 filters [38]. As presented in Figure 7, shortcut connections are used to perform identity mappings [39]. The short-cut connection and the output layer are dimensioned by adding x to a linear projection $F(x)$.

3.6.3. VGG-16

VGG-16 is convolutional neural network model proposed by K. Simonyan and A. Zisserman [40]. It consists of 13 convolutional layers and 5 pooling layers. The input of our VGG-16 architecture is an image of size 224×224 . The input is followed

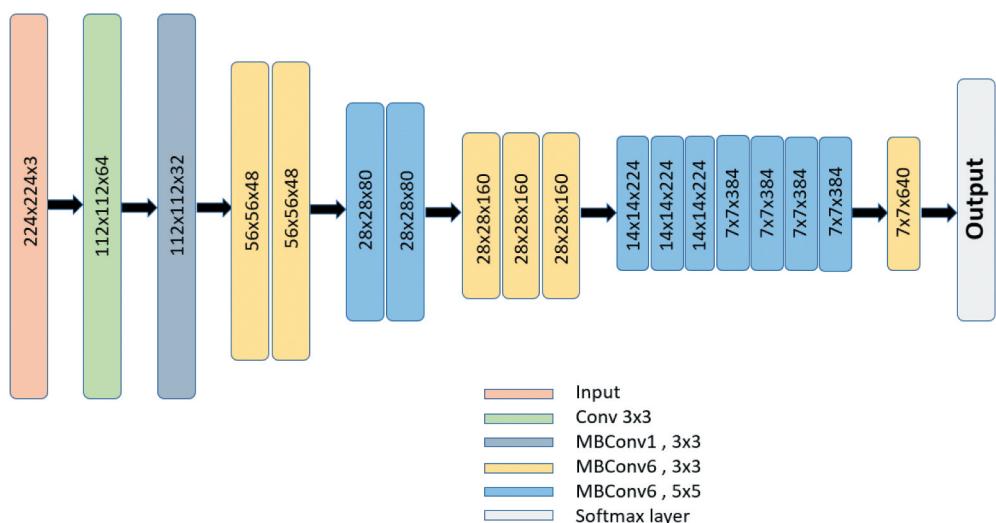


Figure 6. Architecture of EfficientNet-B7.

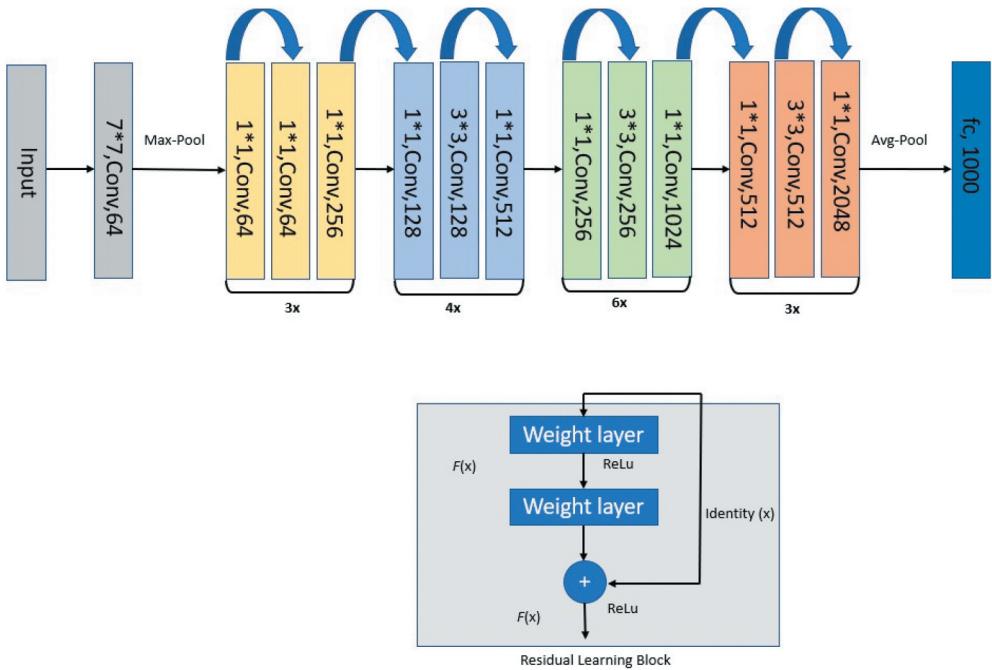


Figure 7. Architecture of ResNet-50.

by different blocks, 2 convolutional layers with 64 kernels, 2 convolutional layers with 128 kernels, 3 convolutional layers with 256 kernels, 3 convolutional layers with 512 kernels and 3 convolutional layers with 512 kernels. Each block is followed by a max pooling using 2×2 filters. The next block is a fully connected layers with 4096 neurons followed by a softmax activation function. Figure 8 presents the VGG-16 architecture [39].

3.6.4. DenseNet-201

DenseNet-201 is convolutional neural network model. It consists of 201 layers. Layers in DenseNet get additional inputs from previous layers and pass their respective feature maps on to the following layers [41]. In Figure 9, the input of our DenseNet-201 architecture is an image of size 224×224 [42].

4. Experiments and results

4.1. Implementation details and performance metrics

Four models are used to train both the database: DMR-IR database and DMR-IR+ segmented database. The training is performed on the augmented data with each thermogram resized to 224×224 pixels. Additionally, we employ an adaptive learning rate, which will be reduced during the learning process if there is no improvement in validation accuracy on 5 consecutive epochs. The batch size and epochs are initially set to 10 and 100 respectively, but an early stopping mechanism is introduced to stop the training if no improvement is observed for over 10 consecutive epochs.

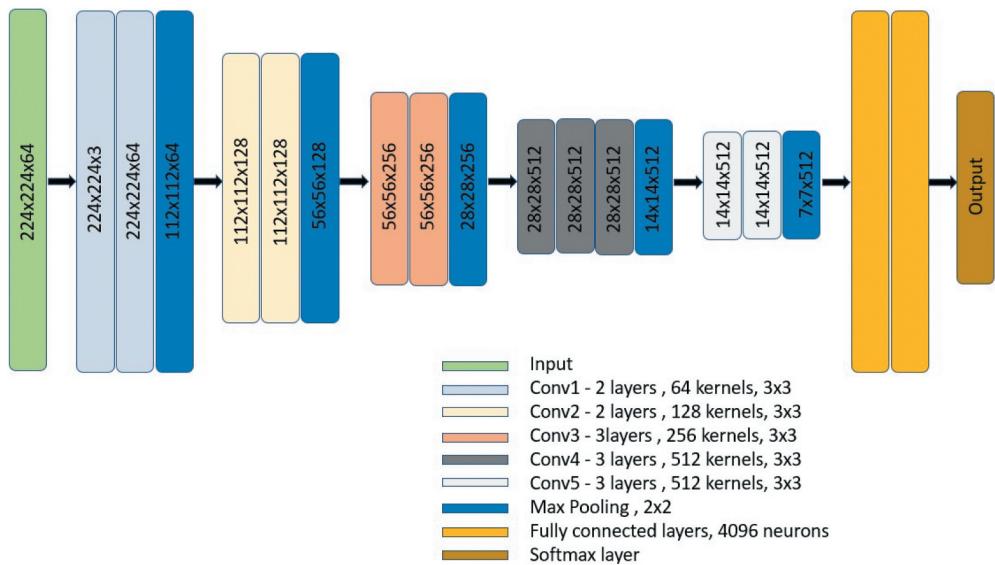


Figure 8. Architecture of VGG-16.

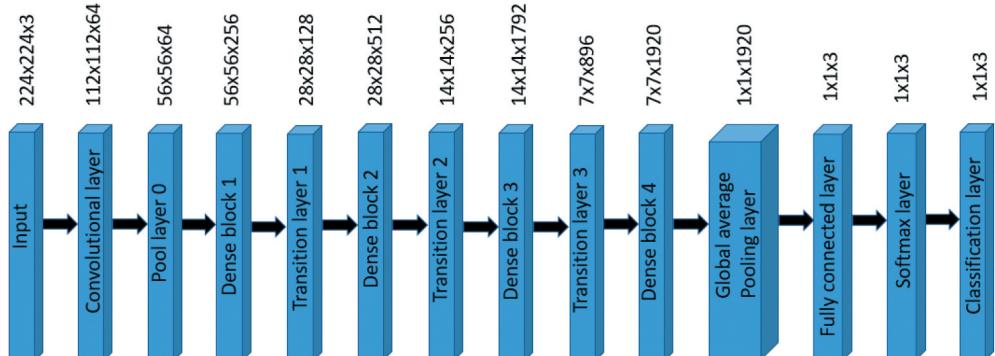


Figure 9. Architecture of DenseNet-201.

Pre-trained weights from ImageNet were applied to the 4 models for training and different optimisations were used such as Adam, Nadam, RMSprop, Adagrad with a variation of learning rates and different batch size depending on the available memory. The best result obtained was with parameters such as Adam optimiser and a learning rate of 0.00003. These parameters were used for the final result. We evaluated the performance of the four models using accuracy, area under the ROC curve (AUC), sensitivity, specificity, and precision. Following is the formula of the metrics that will be used. For class healthy, the true positive (TP) represents healthy patients that are correctly recognised as healthy, false positive (FP) represents the infected patients that are incorrectly recognised as healthy, true negative (TN) represents the infected patients that are correctly recognised as sick and the false negative (FN) represents the healthy patients that are incorrectly recognised as infected.

- Accuracy: The total samples that are correctly classified.

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

- Precision: The fraction of predictions as a positive class were actually positive.

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

- Recall or Sensitivity: The fraction of all positive samples were correctly predicted as positive

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

- Specificity: The fraction of all negative samples that are correctly predicted as negative.

$$\text{Specificity} = \frac{TN}{TN + FP}$$

- AUC: Area under the ROC curve is used to summarise the ROC curve and measure how well a classifier can distinguish between classes.

4.2. Results

The performance metrics of CNN models are shown in Tables 3 and 4. For DMR-IR database, the accuracy for breast cancer classification on dynamic thermograms is 98.36% by EfficientNet-B7, 97.25% by ResNet-50, 77.53% by VGG-16 and 97.81% by DenseNet-201. For DMR-IR + segmented database, the accuracy for breast cancer

Table 3. Comparison of performance metrics of CNN models using DMR-IR database.

Model	Precision (%)	AUC (%)	Sensitivity (%)	Specificity (%)			
				Healthy	Sick	Unknown	Accuracy (%)
EfficientNet-B7	98.36	99.38	98.36	98.71	98.95	99.72	98.36
ResNet-50	97.25	99.79	96.99	91.46	97.90	99.72	97.26
VGG-16	77.53	91.80	77.53	0.00	78.63	98.90	77.53
DenseNet-201	97.81	99.77	97.81	94.93	97.57	99.72	97.81

Table 4. Comparison of performance metrics of CNN models using DMR-IR + segmented database.

Model	Precision (%)	AUC (%)	Sensitivity (%)	Specificity (%)			
				Healthy	Sick	Unknown	Accuracy (%)
EfficientNet-B7	96.44	99.68	96.44	94.87	97.58	99.72	96.44
ResNet-50	97.26	99.51	97.26	100.00	96.94	99.72	97.26
VGG-16	77.53	90.36	77.53	0.00	78.63	98.90	77.53
DenseNet-201	96.44	99.49	96.44	100.00	96.30	99.72	96.44

Table 5. Comparison of the performance metrics of our best CNN model with those of the literature using the same DMR-IR database.

Article	Model	Classes	No. of thermograms - Views	Accuracy (%)	Sensitivity (%)
Ours	ResNet-50	multi-class	3989 - front and lateral	97.26	97.26
[33]	U-Net + CNN	binary	1000 - front	99.33	100.00
[30]	KNN	binary	1520 - front and lateral	99.21	98.40
[23]	CNN	binary	3895 - front and lateral	98.95	-
[32]	SVM	binary	178 - front and lateral	97.18	97.18
[26]	CNN	binary	2740 - front and lateral	95.00	97.00

classification on dynamic thermograms is 96.44% by EfficientNet-B7, 97.26% by ResNet-50, 77.53% by VGG-16 and 96.44% by DenseNet-201. A comparison between the evaluation metrics of the proposed model with those of the literature is shown in Table 5.

4.3. Confusion matrix

Figures 10 and 11 displays a heatmap of confusion matrix for each CNN model that distinguishes healthy, sick and unknown thermograms. A testing dataset containing 364 thermograms was used to generate the confusion matrix.

4.4. Grad-CAM visualisation

Grad-CAM, known as Gradient-weighted Class Activation Mapping, is used to visualize the highlighted region of interest of the model while making prediction [43]. The input of a Grad-CAM is an image and a prediction is made based on the trained model followed by the last convolutional layer applied to Grad-CAM. Figures 12 and 13 shows the results obtained using Grad-CAM visualisation. The evaluation of the model using the Grad-CAM reveals that, with the DMR-IR database, it also takes into account other objects in the thermogram besides the breast region. We have therefore used the DMR-IR + segmented database, which contains both the DMR-IR database and segmented thermograms, in order to improve the Grad-CAM model's ability to focus on the area of interest more precisely. Grad-CAM results improved with the addition of the segmented database in the DMR-IR database, enabling the model to classify thermograms based on their regions of interest.

4.5. Cross-Validation

Cross-validation is a method used to evaluate our model. The training data were splitted into k-subsets. In each subset, the data is divided into 2 parts training and testing, fit a model on the training set and evaluate it on the testing set. The evaluation score is saved for each subset and combined to summarise the evaluation of the model. For image classification, we used a stratified K-Fold in which each split will have a certain percentage of each class. Figure 14 represents an example on how the training data containing 3 classes are splitted into 5 subsets.

The Table 6 shows the results obtained using cross validation with different K-Fold numbers. With the help of cross validation, we observed that the results for all the models

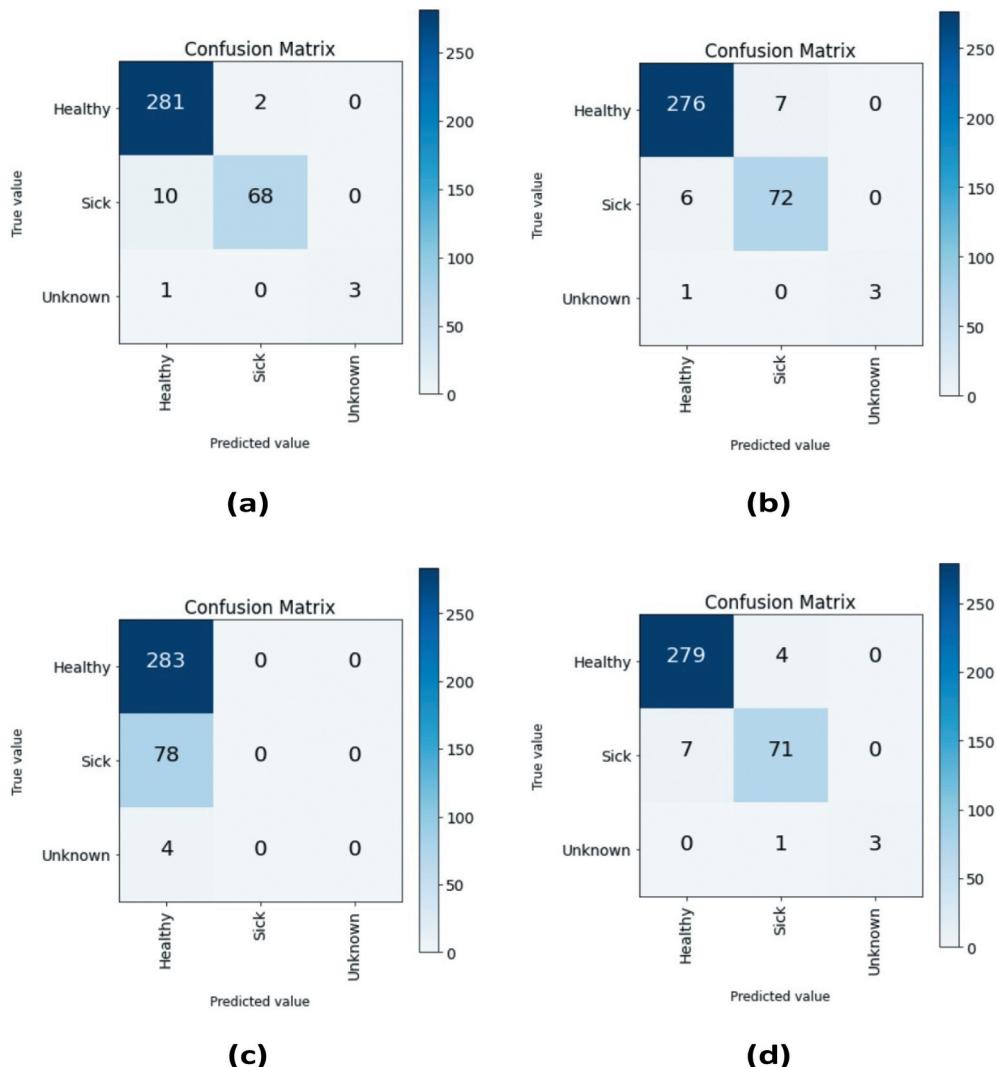


Figure 10. DMR-IR database - confusion matrix: (a) EfficientNet-B7; (b) ResNet-50; (c) VGG-16; (d) DenseNet-201.

have been improved, especially for VGG-16 model. The previous accuracy of VGG-16 was 77.53% and with cross validation, it has been improved up to 98.25%.

5. Discussion

Despite only training with 446 thermograms, the pretrained TransUNet model was able to predict the mask for the entire database and gave the best curve than the manually created mask. The results from each model differ for classification. In the DMR-IR database, the best performance was obtained with the EfficientNet-B7 model, which had an accuracy of 98.36%. However, using the Grad-CAM, we observed that, in addition to the breast area, several other objects were taken into account in the

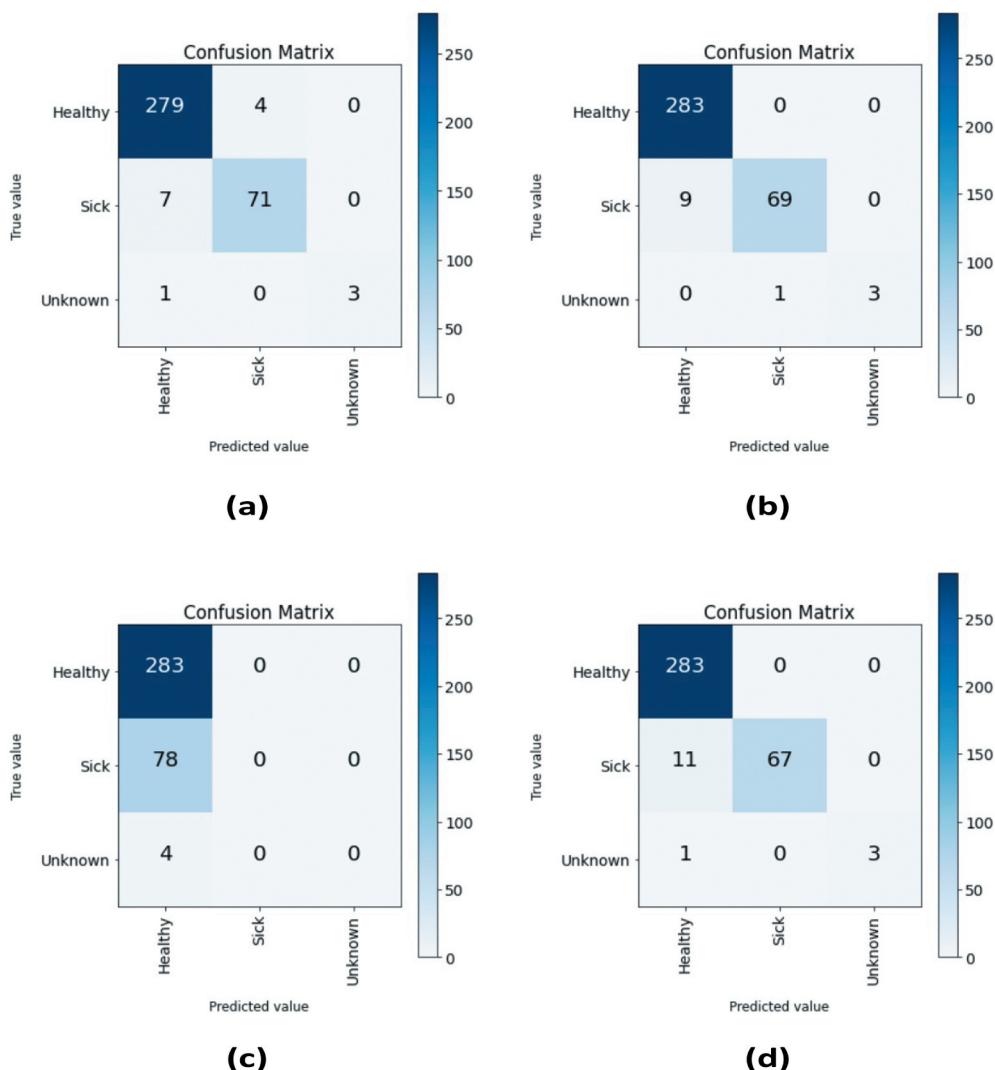


Figure 11. DMR-IR + segmented database - Confusion matrix: (a) EfficientNet-B7; (b) ResNet-50; (c) VGG-16; (d) DenseNet-201.

model. To improve the Grad-CAM model, we combined the DMR-IR database with the segmented thermograms so that the model could focus more on the region of interest. With the DMR-IR + segmented database, the results of the Grad-CAM have improved, allowing the model to classify the thermograms based on their region of interest. Despite using a pretrained model, there is no improvement when using a VGG-16 model. Even though having a limited number of layers, the model was unable to classify the DMR-IR or DMR-IR + segmented database, nor localise the breast region in Grad-CAM. Therefore, only EfficientNet-B7, ResNet-50 and DenseNet-201 models were able to classify and highlight the region of interest.

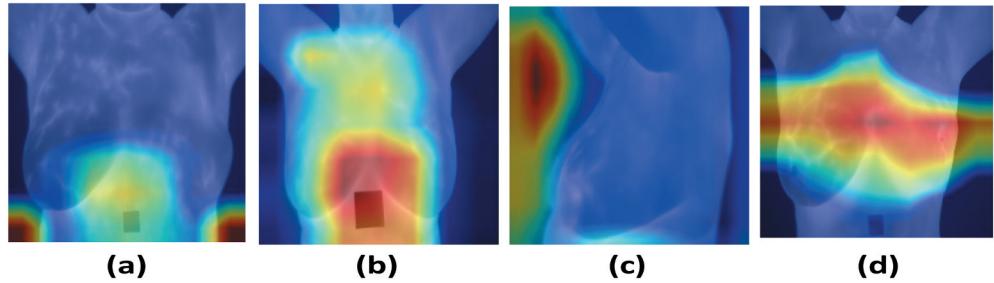


Figure 12. DMR-IR database - Grad-CAM: (a) EfficientNet-B7; (b) ResNet-50; (c) VGG-16; (d) DenseNet-201.

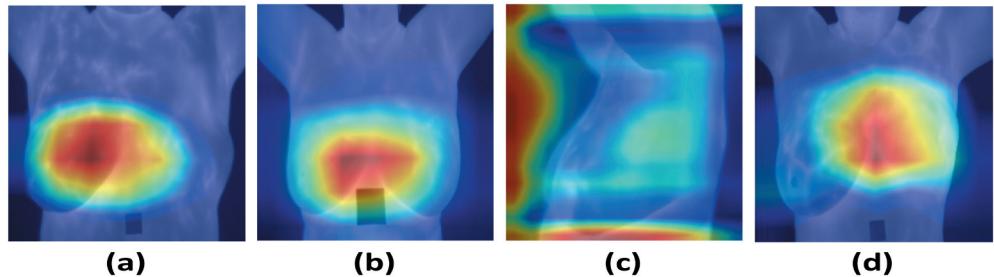


Figure 13. DMR-IR + segmented database - Grad-CAM: (a) EfficientNet-B7; (b) ResNet-50; (c) VGG-16; (d) DenseNet-201.

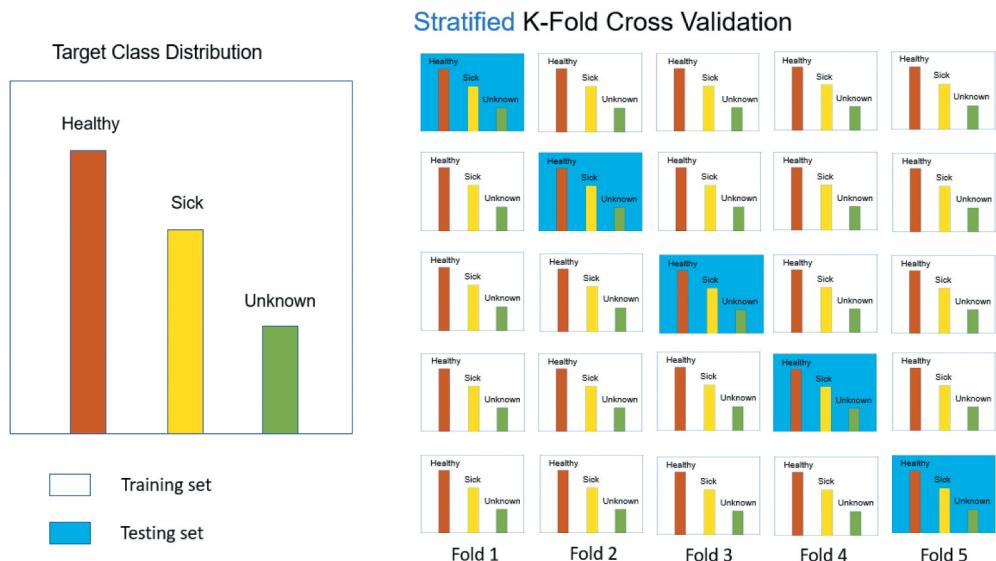


Figure 14. Stratified K-Fold cross-validation representation of training data splitted into 5 subsets.

Table 6. Cross validation with stratified K-Fold using DMR-IR + segmented database.

Model	K-Fold=3		K-Fold=5		K-Fold=10	
	Accuracy (%)	F1-score (%)	Accuracy (%)	F1-score (%)	Accuracy (%)	F1-score (%)
EfficientNet-B7	97.94	97.92	93.50	93.67	93.50	93.18
ResNet-50	97.93	97.94	98.25	98.25	98.09	98.09
VGG-16	98.25	98.24	97.93	97.93	97.46	97.45
DenseNet-201	97.30	97.23	97.62	97.64	98.09	98.10

6. Conclusion

The aim of this project is to segment and classify thermograms using deep convolutional neural networks and Transformers. After segmenting the breast regions, we used ImageNet's pretrained weights and an adaptive learning rate to train four different models on the DMR-IR database and the DMR-IR + segmented database. The results obtained from both databases were comparable. For DMR-IR database, the accuracies obtained by EfficientNet-B7, ResNet-50, VGG-16 and DenseNet-201 were 98.39%, 97.26%, 77.53% and 97.81% respectively and for DMR-IR + segmented database, 96.44%, 97.26%, 77.53% and 96.44% respectively. However, we performed a Grad-CAM visualisation for DMR-IR database to highlight the region of interest of these models and observed that different objects were taken into consideration. The Grad-CAM visualisation of these models was improved by including segmented thermograms in the DMR-IR database. Using DMR-IR + segmented database, the EfficientNet-B7, ResNet-50, and DenseNet-201 models were able to highlight the breast region.

Our next step will be to use a vision transformer to automate the classification process, taking into account each patch of the image.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was enabled in part by support provided by the New Brunswick Health Research Foundation (NBHRF).

ORCID

Ella Mahoro  <http://orcid.org/0000-0001-8451-8496>

Moulay A. Akhloufi  <http://orcid.org/0000-0002-4378-2669>

References

- [1] Mayo Clinic. Breast cancer, 2022. [cited 2022 June 30]. Available from: <https://www.mayoclinic.org/diseases-conditions/breast-cancer/symptoms-causes/syc-20352470>

- [2] Breastcancer.org. Genetic. 2022 [cited 2022 June 30]. Available from: <https://www.breastcancer.org/risk/risk-factors/genetics>
- [3] Andrew Futreal P, Liu Q, Shattuck-Eidens D, et al. Brca1 mutations in primary breast and ovarian carcinomas. *Science*. 1994;266:120–122.
- [4] MedicineNet. Breast cancer prevention. 2022 [cited 2022 June 30]. Available from: <https://www.medicinenet.com/breastcancerprevention/article.htm>
- [5] Knechtges PM, Carlos RC. The evolving role of radiologists within the health care system. *J Am College Radiol*. 2007;4(9):626–635.
- [6] Dongola N. Mammography in breast cancer. 2018 [cited 2022 June 30] . Available from: <https://emedicine.medscape.com/article/346529-overview>
- [7] Mayo Clinic. Mammogram. 2022 [cited 2022 June 30]. Available from: <https://www.mayoclinic.org/tests-procedures/mammogram/about/pac-20384806>
- [8] Holistic Breast Health. Comparison of breast screenings. 2022 [cited 2022 June 30]. Available from: <https://holisticbreasthealth.com/good-breast-health/compare-breast-screenings/>
- [9] Burke E. Mammography vs. Thermography: which is better at detecting breast cancer? 2022 [cited 2022 June 30]. Available from: <https://www.wakerad.com/expert-feature/Mammography-vs-thermography-which-is-better-at-detecting-breast-cancer/>
- [10] Cowley G, Veazey K. Comparing mammography and thermography. 2022 [cited 2022 June 30]. Available from: <https://www.medicalnewstoday.com/articles/316632>
- [11] Keyserlingk JR, Ahlgren PD, Yu E, et al. Functional infrared imaging of the breast. *IEEE Eng Med Biol Mag*. 2000;19:30–41.
- [12] Caplan L. Delay in breast cancer: implications for stage at diagnosis and survival. *Front Public Health*. 2014;2:87.
- [13] Richards MA, Westcombe AM, Love SB, et al. Influence of delay on survival in patients with breast cancer: a systematic review. *Lancet*. 1999;353(9159):1119–1126.
- [14] Miotto R, Wang F, Wang S, et al. Deep learning for healthcare: review, opportunities and challenges. *Brief Bioinform*. 2018;19(6):1236–1246.
- [15] Khan M-H-M, Boodoo-Jahangeer N, Dullull W, et al. Multi-Class classification of breast cancer abnormalities using deep convolutional neural network (CNN). *PLoS One*. 2021;16(8):e0256500.
- [16] Johnson JM, Khoshgoftaar TM. Survey on deep learning with class imbalance. *J Big Data*. 2019;6(1):1–54.
- [17] Kösters JP, Gøtzsche PC. Regular self-examination or clinical examination for early detection of breast cancer. *Cochrane Database Syst Rev*. 2003;2.
- [18] Khan M-H-M, Boodoo-Jahangeer N, Dullull W, et al. Multi-class classification of breast cancer abnormalities using deep convolutional neural network (CNN). *PLoS One*. 2021;16:1–15.
- [19] Kalaf JM. Mammography: a history of success and scientific enthusiasm. *Radiologia Brasileira*. 2014;47:7–8.
- [20] DeFrank JT, Rimer BK, Bowling JM, et al. Influence of false-positive mammography results on subsequent screening: do physician recommendations buffer negative effects? *J Med Screen*. 2012;19(1):35–41.
- [21] de Santana M, Pereira JM, da Silva F, et al. Breast cancer diagnosis based on mammary thermography and extreme learning machines. *Res Biomed Eng*. 2018;34:45–53.
- [22] Torres-Galván J, Guevara E, González FJ. Comparison of deep learning architectures for pre-screening of breast cancer thermograms. 2019 Photonics North (PN). 2019;CFP1909V-ART:1–2.
- [23] Ekici S, Jawzal H. Breast cancer diagnosis using thermography and convolutional neural networks. *Med Hypotheses*. 2020;137:109542.
- [24] Krawczyk B, Schaefer G. Breast thermogram analysis using classifier ensembles and image symmetry features. *IEEE Syst J*. 2014;8:921–928.
- [25] Al Husaini MAS, Habaebi MH, Gunawan TS, et al. Self-Detection of early breast cancer application with infrared camera and deep learning. *Electronics*. 2021;10:2538.

- [26] Baffa MDFO, Lattari LG. Convolutional neural networks for static and dynamic breast infrared imaging classification. In: *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*; 174–181. IEEE; 2018.
- [27] Chebbah NK, Ouslim M, Benabid S. New computer aided diagnostic system using deep neural network and svm to detect breast cancer in thermography. *Quant Infrared Thermogr J.* 2022;1–16.
- [28] Torres-Galvan JC, Guevara E, Kolosovas-Machuca ES, et al. Deep convolutional neural networks for classifying breast cancer using infrared thermography. *Quant Infrared Thermogr J.* 2022;19(4):283–294.
- [29] Pramanik S, Bhattacharjee D, Nasipuri M. Wavelet based thermogram analysis for breast cancer detection. In: *2015 International Symposium on Advanced Computing and Communication (ISACC)*, 2015. p. 205–212.
- [30] Ma J, Shang P, Lu C, et al. A portable breast cancer detection system based on smartphone with infrared camera. *Vibroengineering Procedia.* 2019;26:57–63.
- [31] Milosevic M, Jankovic D, Peulic A. Thermography based breast cancer detection using texture features and minimum variance quantization. *Excli J.* 2014;13:1204.
- [32] Resmini R, Silva L, Araujo AS, et al. Combining genetic algorithms and svm for breast cancer diagnosis using infrared thermography. *Sensors.* 2021;21(14):4802.
- [33] Mohamed EA, Rashed EA, Gaber T, et al. Deep learning model for fully automated breast cancer detection system from thermograms. *PLoS One.* 2022;17:1–20.
- [34] Chen J, Lu Y, Yu Q, et al. Transunet: transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306.* 2021.
- [35] Silva LF, Saade DCM, Sequeiros GO, et al. A new database for breast research with infrared image. *J Med Imaging Health Inform.* 2014;4(1):92–100.
- [36] Tan M, Le QV. Efficientnet: rethinking model scaling for convolutional neural networks. CoRR. 2019. Available from: <https://arxiv.org/abs/1905.11946>
- [37] Atila Ü, Uçar M, Akyol K, et al. Plant leaf disease classification using efficientnet deep learning model. *Ecol Inf.* 2021;61:101182.
- [38] Chougrad H, Zouaki H, Alheyane O. Deep convolutional neural networks for breast cancer screening. *Comput Methods Programs Biomed.* 2018;157:19–30.
- [39] Ali L, Alnajjar F, Al Jassmi H, et al. Performance evaluation of deep CNN-based crack detection and localiza-tion techniques for concrete structures. *Sensors.* 2021;21:1688.
- [40] Simonyan K, Zisserman A, Zisserman A. Learning local feature descriptors using convex optimisation. 2014. *arXiv preprint arXiv:1409.1556.*
- [41] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017. 4700–4708.
- [42] Yu X, Zeng N, Liu S, et al. Utilization of densenet201 for diagnosis of breast abnormality. *Mach Vision Appl.* 2019;30:1135–1144.
- [43] Selvaraju RR, Cogswell M, Das A, et al. Grad-Cam: visual explanations from deep networks via gradient-based localization. In: *Proceedings of the IEEE international conference on computer vision*, 2017. p. 618–626.