

Article

Breast Cancer Diagnosis in Thermography Using Pre-Trained VGG16 with Deep Attention Mechanisms

Alia Alshehri *  and Duaa AlSaeed 

College of Computer and Information Sciences, King Saud University, Riyadh 11451, Saudi Arabia

* Correspondence: aliaalshehri96@gmail.com

Abstract: One of the most prevalent cancers in women is breast cancer. The mortality rate related to this disease can be decreased by early, accurate diagnosis to increase the chance of survival. Infrared thermal imaging is one of the breast imaging modalities in which the temperature of the breast tissue is measured using a screening tool. The previous studies did not use pre-trained deep learning (DL) with deep attention mechanisms (AMs) on thermographic images for breast cancer diagnosis. Using thermal images from the Database for Research Mastology with Infrared Image (DMR-IR), the study investigates the use of a pre-trained Visual Geometry Group with 16 layers (VGG16) with AMs that can produce good diagnosis performance utilizing the thermal images of breast cancer. The symmetry of the three models resulting from the combination of VGG16 with three types of AMs is evident in all its stages in methodology. The models were compared to state-of-art breast cancer diagnosis approaches and tested for accuracy, sensitivity, specificity, precision, F1-score, AUC score, and Cohen's kappa. The test accuracy rates for the AMs using the VGG16 model on the breast thermal dataset were encouraging, at 99.80%, 99.49%, and 99.32%. Test accuracy for VGG16 without AMs was 99.18%, whereas test accuracy for VGG16 with AMs improved by 0.62%. The proposed approaches also performed better than previous approaches examined in the related studies.

Keywords: breast cancer; thermography; deep learning; VGG16; attention mechanisms; machine learning; detection; diagnosis; artificial intelligence



Citation: Alshehri, A.; AlSaeed, D. Breast Cancer Diagnosis in Thermography Using Pre-Trained VGG16 with Deep Attention Mechanisms. *Symmetry* **2023**, *15*, 582. <https://doi.org/10.3390/sym15030582>

Academic Editors: João Ruivo Paulo, Cristina P. Santos and Gabriel Pires

Received: 1 February 2023
Revised: 17 February 2023
Accepted: 21 February 2023
Published: 23 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Human diseases have increasingly grown due to environmental or personal factors despite various diagnoses and preventative strategies. The most prevalent disease in the world is cancer, characterized by erratic cell growth that spreads to other body organs. Breast, prostate, lung, skin, and pancreatic cancers have all been reported. Cancer is one of the causes of losing many lives [1].

The likelihood of survival has increased with the introduction of early diagnosis and treatment techniques for breast cancer. The patient is screened utilizing several imaging modalities as one of the extensively practiced medical techniques to identify breast cancer. Imaging modalities screening includes mammography, ultrasound, magnetic resonance imaging (MRI), computed tomography imaging (CT), and thermal imaging. Results from these modalities rely on several circumstances; it is advised to use multiple methods to check the results because they each have unique techniques and instruments.

The most popular method of detecting breast cancer is mammography. Although it is the gold standard, mammography has several risks for patients. Due to the fact that it poses less danger than mammography, thermography has generated interest recently [2].

Skin lesions and temperature are the main inputs for digital infrared thermal imaging; thus, it has not yet replaced other modalities as the primary method of finding early breast cancer [3]. Nevertheless, encouraging findings suggest that radiologists can use it in conjunction with mammography to determine the condition of the breast for an accurate evaluation and diagnosis of breast cancer [3]. Compared to mammography, it has not been

utilized as frequently; although, it has shown good outcomes of early diagnosis. It also offers several benefits because it is painless, safe, non-invasive, and affordable.

The accuracy of breast cancer screening and diagnosis using machine learning (ML) and deep learning (DL) approaches has increased recently. Due to the encouraging outcomes obtained when DL algorithms for breast cancer screening were coupled with deep AMs, deep attention approaches have also gained increased attention in this sector [4]. As a result, there is a higher chance that the performance of the used approaches will be improved since notable regions of the image are given additional attention rather than treating all batches as equally important.

Despite the advancements in breast cancer diagnosis, additional study is required to investigate various modalities and improve detection accuracy when utilizing DL algorithms with deep AMs, and to further the area of breast cancer detection, this research intends to do so. This study's primary goal is to offer a trustworthy and efficient approach based on AMs and DL to identify breast cancer utilizing thermal imaging, assisting medical professionals in establishing a more thorough and credible screening of breast cancer tumors. Moreover, poor and developing countries will benefit from improved diagnostic techniques using thermal imaging due to it being inexpensive, which enhances the quality of life in those areas.

The contributions were mentioned in this research study as follows:

- This research study presents accurate models of breast cancer detection based on the pre-trained model VGG16 and AMs using thermographic images.
- This research study evaluates the performance of VGG16 with and without AMs to determine the effect of AMs on the performance of VGG16.
- This research study compares the proposed models with related studies.

Following is the arrangement of the remaining sections of this article. The background overview of the DL approaches is given in Section 2. A brief overview of deep AMs is provided in Section 3. The related works of breast cancer detection are reviewed in the following section. The suggested approaches are covered in Section 5, along with the experiment design and assessment procedure. The experimental findings are discussed in Section 6. The conclusions of this paper, including the main findings, limitations, and future directions, are presented in Section 7.

2. Deep Learning and Convolutional Neural Network (CNN)

A subset of artificial intelligence (AI) called machine learning (ML) enables computers to automatically learn from experiments and get better without explicitly being programmed. We can divide AI inventions into two groups: conventional ML and DL. Conventional ML methods rely on system learning from training data to construct a trained model. A more sophisticated class of ML algorithms called DL exceeded the performance of conventional ML methods [5].

DL algorithms efficiently handle vast data as opposed to small data. In the literature, many DL method categories have been used. Convolutional neural networks are one of the major DL algorithms (CNNs).

Regarding image-based classification and recognition for supervised learning, a CNN is a powerful and prevalent DL technique. DL has generally transformed computer vision by deploying superhuman accuracy techniques for a variety of jobs and applications. There have been several prominent methods for classifying and processing images that have enhanced performance, particularly in the medical industry.

In recent years, CNN has experienced rapid growth and the development of several architectures, such as VGG16, which is considered one of the most famous of CNN. The VGG16 architecture comprises five blocks of convolutional layers followed by three fully connected layers and the finally 1000 fully connected softmax layer. The VGG16 model has the drawback of being expensive to assess and requiring a large amount of memory and parameters. There are over 138 million parameters in VGG16. The fully connected layers include the majority of these characteristics (123 million in total) [6].

3. Deep Attention Mechanisms

Typically, when performed on images, ML approaches assign each image patch identical weight without paying particular “attention” to significant parts. The most prominent portions of the image are highlighted when AMs and ML are combined, raising the likelihood that the productivity of the techniques used will be enhanced. The performance of the dynamically generated outcome can be improved by using the AMs to highlight the essential segments of the inputs, and thereby enhance the capacity to extract the most pertinent information for each outcome segment and repress, or even disregard utterly, unrelated information [7]. We shall briefly explore each of the three basic categories of AMs.

- **Soft Attention:** The categorical distribution is computed through a set of elements, then the weights are generated by the resulting probabilities. The probabilities that result represent the importance of each element and are utilized as weights to generate a context-aware encoding that is the weighted sum of all elements. Due to the interdependence between the mechanism and the purpose of the deep neural network, it determines how much focus should be given to each input element by giving each element a weight between zero and one [8].
- **Hard Attention:** A subset of elements are selected from the input sequence. The weight allotted to an input element in hard AMs can be zero or one, forcing the approach to concentrate only on the critical elements and ignore the rest. Due to the input elements being either observed or not, the goal is non-differentiable [9].
- **Self-Attention:** The interdependence of the mechanism’s input elements is estimated because it permits the input to interact with the other “self” and identify what it should focus on more. One of the essential benefits of the self-attention layer against hard and soft mechanisms is its parallel computing capability for a lengthy input. To verify that all the same input elements are being paid attention to, this mechanism layer performs straightforward and efficiently parallelizable matrix calculations [9].

4. Related Work

A review of recent research on utilizing thermal imaging for breast cancer screening is presented in this section.

Several recent investigations have been carried out, employing ML and DL to diagnose breast cancer from mammograms.

The U-net structure inspired the authors’ novel CNN design of the two datasets of CBIS-DDSM mass and macrocalcification mammography [10], which included 692 mass and 603 macrocalcification images. The model has a 94.31% accuracy.

The BC-DROID approach allowed for one-step automated detection and classification using a CNN [11]. The approach was trained using 10,480 whole mammograms derived from DDSM data, with a detection accuracy of 90%, classification accuracy of 93.5%, and an AUC of 92.315%.

In [12], a grayscale co-occurrence matrix and a grayscale run-length matrix were extracted and fed two inputs to a hybrid CNN and RNN model called CRNN for mammographic breast cancer detection. The results and AND operation outcomes of the classifier achieved a diagnostic accuracy of 90.59%, which exceeded conventional models.

Recent years have seen more studies into utilizing thermal imaging to detect and classify breast cancer due to the quick development of infrared cameras. Several studies have used ML and DL algorithms on thermal imaging data.

The authors of [13] categorized 1052 thermograms from the University Hospital at The Federal University of Pernambuco. They employed a variety of models, including the Bayes network, naive Bayes, J48 decision tree, SVM, random forest (RF), multi-layer perceptron (MLP), ELM, and random tree (RT). The findings indicated that MLP performed well when compared to the other classifiers, with an accuracy of 73.38%, a kappa value of 0.6007, a sensitivity of 78%, and a specificity of 88%. The accuracy was enhanced to 76.01% by employing a 10-fold cross-validation procedure with a kappa index of 0.6402. The overall efficiency of the system was 83%.

The performance of SVM, ANN, DT, and KNN classifiers in diagnosing thermal images from the DBT-TU-JU and DMR-IR datasets was improved [14] using the features SSigFS, FStat, and STex. The performance of the classifiers in the two databases was compared: in the DBT-TU-JU dataset, SVM-RBF and ANN acquired the greatest accuracy of 84.29%, while in the DMR-IR dataset, ANN and SVM-linear earned the maximum accuracy of 87.50%.

On the other side, DL algorithms, particularly a CNN, have been applied to thermal images and have demonstrated competitive performance in breast diagnosis. A CNN was used for the first time to classify thermal images and was optimized using the Bayes optimization technique [15]. The accuracy was 98.05% using 1116 images taken from the DMI dataset. The model outperformed previous studies that utilized the same dataset with different features and classifiers on a larger number of images.

Authors of a different study [16] used a DCNN model to detect breast cancer using thermal images. They transformed 680 thermal images from a Visual Lab-IR dataset to grayscale before pre-processing, segmenting, and classifying them. They attained a 95.8% prediction accuracy. They surpassed the study in [17], which had a 93.30% accuracy with 50 thermograms. Furthermore, compared to a study published in 2012 [18] that employed the DT and fuzzy classifiers, it reached the most fantastic average accuracy of 93.30%, indicating a substantial increase by DCNN.

The combining of GVF breast segmentation and CNN classification was studied by the authors in [19] as a potential method for detecting breast cancer. They utilized 63 images from the DMR-IR dataset as samples of normal and abnormal breasts. Using a two-fold cross-validation methodology, the model was assessed. It performed better than tree random forest (TRF), MLP, and Bayes Network, achieving 100% accuracy, sensitivity, and specificity.

For the purpose of early breast cancer detection, the authors of [20] used a multi-input classifier model based on CNNs that integrates thermal images from various viewpoints with the personality and clinical data of 287 patients. They analyzed the performance of seven models, and the findings showed that the M.4ncd model had the highest accuracy (97%) and the best specificity (100%) and sensitivity (83%) values. The M.4ncd model performed better in AUC ROC and specificity metrics compared to the findings of other literature research.

Using static and dynamic protocols, CNNs were trained to categorize thermal images [21]. There were 300 thermal images in the DMR-IR dataset that were classified using the static protocol. The dynamic methodology was used to classify 2740 images. In both protocols, the suggested technique produced competitive results. The dynamic protocol obtained 95% accuracy for color images and 92% for grayscale, while the static protocol obtained 98% accuracy for color images and 95% for grayscale. It performed better than other techniques used on the same dataset.

Deep transfer learning models are used for transmission to classify medical images. In [22], the authors suggested training a visual geometry group 16 (VGG16) model to classify normal or abnormal breast thermal images with the help of a static and dynamic DMR-IR dataset of 1345 images with multi-view and single view. For the first time, conventional frontal, left-, and right-view breast thermal images from the Mastology Research database are sequenced with an infrared image to produce multi-view thermal images. This approach improves the system's accuracy by providing a more comprehensive and informative thermal temperature. Using multi-view images, VGG16 achieves an encouraging test accuracy of 99% on the dynamic breast imaging test dataset. To compare the VGG16 model performance with other deep transfer learning models, VGG19, ResNet50V2, and inceptionV3 were trained and tested to achieve test accuracy of 95%, 94%, and 89%, respectively. This indicates that the VGG16 model outperformed the other models even though these models are more complex and provide better results in other medical imaging classification tasks.

The studies that employed thermal imaging to diagnose breast cancer showed that, when compared to other imaging techniques, thermal imaging had positive outcomes. They were also combined with ML and DL techniques.

As we can see, the prior research that was evaluated accorded all patches of an image the same weight or “attention” and did not pay especially important consideration to any of the more prominent regions. On the other hand, focusing extra attention on important regions of the breast image may improve the model’s performance and detection outcomes.

The authors of [23] enhanced CNNs by using a novel SE-Attention mechanism to categorize 18,157 gathered mammograms and create a new benchmarking dataset. The model outperformed prior research with an accuracy of 92.17%.

In conclusion, the presented concise overview of the relevant study indicated that DL with deep AMs had not been utilized on thermal images for diagnosing breast cancer, as shown in Table 1.

On the other hand, based on the studies mentioned in the preceding, breast cancer diagnosis in thermal imaging has demonstrated encouraging accuracy outcomes when compared to other methods. This has motivated us to investigate the impact of integrating DL approaches with AMs in diagnosing breast cancer using thermal imaging in our previous study [24].

In our work [24], we applied CNN with self, soft, and hard AMs on the DMR-IR dataset of thermographic images for breast cancer diagnosis. The model was trained on 4146 images and achieved test accuracy of 99.46%, 99.34%, 99.32%, and 84.92% for hard, soft, self, and CNN alone without AMs, respectively. The CNN model with AMs showed results that outperformed many of the studies discussed in related work.

The promising results achieved with CNN have encouraged us to explore different types of AMs using the same methodology with pre-trained DL methods, namely VGG16, in detecting breast cancer in thermal images and comparing the results with those achieved with CNN in our previous study.

Table 1. Breast cancer detection and classification studies.

Ref.	Approaches	Imaging Modalities	Datasets	Results
[10]	U-Net CNN	Mammography	CBIS-DDSM mass images CBIS-DDSM microcalcification images	Acc = 94.31%
[11]	CNN	Mammography	DDSM	Acc = 93.5% AUC= 0.92315
[12]	CNN-RNN	Mammography	Mammogram image dataset	Acc = 90.59% Sn = 92.42% Sp = 89.88%
[13]	Bayes Network Naïve Bayes SVM Knowledge Tree J48 MLP RF RT ELM	Thermal	University Hospital of the Federal University of Pernambuco	Acc = 76.01%
[14]	SVM KNN DT ANN	Thermal	DBT-TU-JU DMR-IR	Acc = 84.29% Acc = 87.50%

Table 1. Cont.

Ref.	Approaches	Imaging Modalities	Datasets	Results
[15]	CNNs-Bayes optimization algorithm	Thermal	DMI	Acc: 98.95%
[16]	DCNNs	Thermal	DMR-IR	Acc = 95.8% Sn = 99.40% Sp = 76.3%
[17]	DNN	Thermal	DMR-IR	Conf (Sick) = 78% Conf (Healthy) = 94%
[18]	DT Fuzzy Sugeno Naïve Bayes <i>k</i> -Nearest Neighbor Gaussian Mixture Model Probabilistic Neural Network	Thermal	Singapore General Hospital-NEC-Avio Thermo TVS2000 MkIIIST System	Acc = 93.30% Sn = 86.70% Sp = 100%
[19]	CNNs-GVF	Thermal	DMR-IR	Acc = 100% Sn = 100% Sp = 100%
[20]	Multi-input CNN	Thermal	DMR-IR	Acc = 97% Sn = 83% Sp = 100% AUC = 0.99
[21]	CNNs	Thermal	DMR-IR	Acc(color) = 98% Acc(grayscale) = 95% Acc(color) = 95% Acc(grayscale) = 92%
[22]	VGG16	Thermal	DMR-IR	Acc = 99% Sn = 98.04% Sp = 100% Prec = 100% F1-score = 99.01%
[23]	CNNs + SE-Attention	Mammography	New Benchmarking dataset	Acc = 92.17%
[24]	CNNs + AMs	Thermal	DMR-IR	Acc = 99.49% Sn = 99.71% Sp = 99.71% Prec = 99.28% F1-score = 99.49% AUC = 0.999

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

5. Materials and Methods

This section shows how we achieved our aims and objectives in practice. An experimental study method was used to address the following research question: “To what degree may DL approaches using deep AMs improve performance for the task of breast cancer diagnosis in thermography images?”

Our suggested technique uses AMs and the VGG16 methodology to enhance breast cancer diagnosis and classification utilizing thermal images via five components: pre-processing, feature extraction, bidirectional long short-term memory (BLSTM), AMs, and image classification. The stages in the suggested technique are illustrated in Figure 1. Following that, we will go over these steps in further detail.

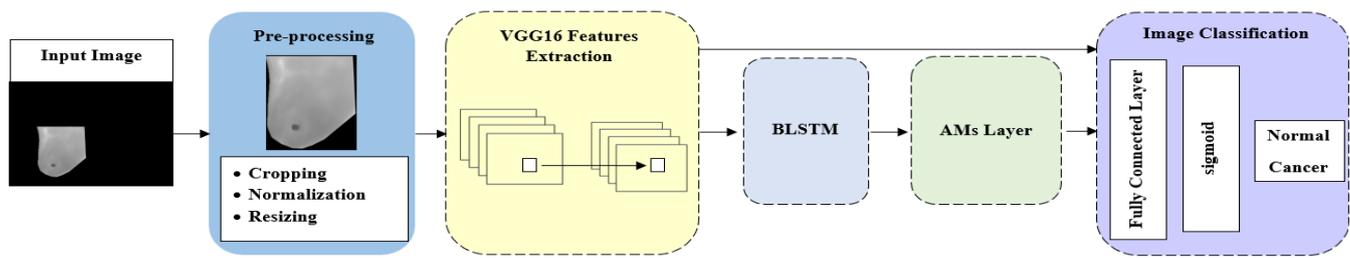


Figure 1. Model architecture.

5.1. Dataset

According to our review of the studies on the use of thermal images to diagnose breast cancer, DMR-IR, which includes static and dynamic protocols, was the most used thermal image dataset [25]. Images are captured at the University Hospital of UFPE (Brazil) and recorded into a database with other details, such as age, family history, exam date, and patient preparation. The Brazilian Ministry of Health’s ethical committee approved the acquisition procedure and storage of images. Some influential aspects are considered when taking thermographic images to detect breast cancer, such as examination room conditions, instructions that the patient must adhere to, and capture positions. To prevent physiological changes in the body that could occur in an uncontrolled setting, the image must be captured in a controlled environment. The patient must be informed to avoid things that contribute to changes in an uncontrolled setting, such as exercises, deodorants, cosmetics, lotions, caffeine, alcohol, and smoking, as well as removing any jewelry and preventing exposure to the sun. The room’s temperature must be between 18 and 25 °C to obtain a thermal image and 40% to 75% humidity. Any source of heat or wind, as well as sunshine, must be excluded from the area. The FLIR SC-620 thermal camera at a resolution of 640×480 with a thermal sensitivity of 45 mk was used to capture thermal images of 287 volunteer women, ranging in age from 29 to 85 years old. The dynamic protocol is often taken in patches of 20 each, while the static protocol is obtained from five different angles. The diagnostic has been prior confirmed via mammography, ultrasound, and biopsies. In addition, the radiologist has thoroughly authenticated these thermal images and their associated annotations.

5.2. Data Pre-Processing

We utilized DMR-IR datasets that included segmented grayscale images. At this stage, we pre-processed the images via cropping, normalizing, and resizing. These thermal images were after pre-processing from their original 640×480 dimensions to the VGG16 model’s default input size of 224×224 . To reduce overfitting, all images were augmented in the following stage to offer a larger dataset for training.

After pre-processing the dataset, we used a stratified 10-fold cross-validation technique to divide it into training and testing sets. The data were split into ten roughly equal segments, with one segment serving as a roughly equal test set.

5.3. Feature Extraction

Pattern recognition that discriminates between cancer and healthy breasts requires feature extraction. Pre-trained models are superior to recently created from-scratch CNN architectures in accuracy and efficiency, particularly for classification purposes. It is considered that the pre-trained VGG16 model represents a developed version of the Alex Net neural network, so we investigated VGG16 in our study for feature extraction. The transfer learning techniques using VGG16 were pre-trained on the ImageNet dataset to extract features. When loading a VGG16 model, we set the “include_top” parameter to False, in which case the fully connected output layers of the model used to make predictions are not loaded, which are replaced by the following layers below, significantly reducing the number of necessary parameters. In addition, we freeze the convolutional base before build-

ing and training the model to prevent losing any of the data they include in subsequent training rounds.

5.4. Bidirectional Long Short-Term Memory Layer

We employed a BLSTM network to extract long-term information from spatial features, which are widely acknowledged to have spatial features in images. A bidirectional LSTM was used to extract temporal features from forward and backward order in order to fully use the past and future context information of a sequence in classification.

5.5. Attention Mechanisms Layer

Using the output from the BLSTM, we generated a variety of attention values that indicated the importance of the feature vector. We applied three AMs separately in this layer, using similar procedures as in the layers preceding and following. The three AMs were self-attention (SL), soft attention (SF), and hard attention (HD).

- SL layer: Using an attention approach that considers the context of each timestamp when processing sequential data. It is implemented in this study by the package Keras-self-attention with multiplicative type, using the following formulae [26]:

$$e_t, t' = \sigma(x_t^T W_a x_{t'} + b_a)$$

- SF layer: It uses a low weight to multiply the associated feature map to discard unimportant regions. As a result, an area with high attention maintains its original value, whereas an area with low attention drifts closer to 0 (and becomes dark in the visualization). To calculate a weight α_i for covering each sub-section of an image, we utilize the hidden state $C = h_t - 1$ from the previous time step. To determine how much attention is being paid, we create a score s_i using the formula [27]:

$$s_i = \tanh(W_c C + W_x X_i) = \tanh(W_c h_{c-1} + W_x x_i)$$

We feed s_i through a softmax for normalization to obtain the weight α_i .

$$\alpha_i = \text{softmax}(s_1, s_2, \dots, s_i)$$

Using softmax, we can calculate a weighted average for x_i by using α_i , which adds up to 1.

$$Z = \sum_i \alpha_i x_i$$

- HD layer: The weight applied to an input portion is either 0 or 1; this causes the model to concentrate only on the critical elements while disregarding others. The outcome is that the input parts are either observed or not, making the goal non-differentiable. Instead of utilizing a weighted average as in SF [27], HD is computed using α_i as a sample rate to select one x_i as the input to the next layer.

$$Z \sim x_i, \alpha_i$$

5.6. Fully Connected Layer

The output from the layers above was flattened and given to the fully connected layer. We merged all the information gained from the network’s previous layers to classify the input image.

5.7. Sigmoid Layer

A sigmoid function was utilized to transform the output of the fully connected layer into binary (0 or 1), which can be interpreted as classification probabilities.

5.8. Classification Layer

In the network, this was the last layer. By utilizing the classification probabilities produced by each input's sigmoid, the layer classified each input into one of the two classes (normal or cancer). The trained network performed the test dataset classification after training.

Three AM approaches, SF, SL, and HD with VGG16, were developed, which we called VGG16-SL, VGG16-HD, and VGG16-SF. For comparison and to pinpoint the effect of the AMs on the VGG16 performance, we evaluated a VGG16 without them. All layers except the BLSTM and AMs were applied in the VGG16 model without AMs, whereas all layers were applied in the VGG16 models with AMs, as shown in Figure 1.

5.9. Model Evaluation

The main evaluation metrics were utilized to assess and compare the effectiveness of the various generated models for detecting breast cancer using AMs and deep neural networks in thermal breast images: accuracy, specificity, sensitivity, precision, F1-score, and others. The following formulae are provided for these metrics [28]:

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

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (2)$$

$$\text{Sensitivity (also known as Recall)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (4)$$

$$\text{F1 - score} = 2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

where TP, TN, FP, and FN indicate true positive, true negative, false positive, and false negative, respectively.

Accuracy is expressed as a percentage of all instances that were properly classified. Even while it is simple to grasp, it can be deceptive, particularly when there is an imbalance in class. The percentage of negative events is called specificity. Precision measures how many accurately positively labeled (abnormal) instances there are compared to all positively classified cases. The percentage of positive (abnormal) instances that were correctly classified as positive is known as sensitivity (recall). However, the F1-score is a more accurate measure of the classifier's performance. Integrating the two competing measures of precision and recall summarizes the predictive performance of a model [28].

The AUC ROC, which has a value range between 0.5 and 1, measures the model's ability to distinguish across classes. It evaluates the model's diagnostic performance in the context of medicine.

Values closer to 1 denote good test findings, while values closer to 0.5 denote poor test findings [20]. Furthermore, the constructed models were evaluated against the state-of-the-art studies covered in Section 4, Related Work.

6. Results

On the DMR-IR dataset of thermal images, we compared the performances of the developed approaches employing SF, SL, and HD AMs with a VGG16, resulting in four different approaches (VGG16, VGG16-SF, VGG16-SL, and VGG16-HD). We contrasted the four approaches' produced findings with the state-of-the-art approaches discussed in the Related Work section to assess our proposed approaches against competing approaches.

The succeeding parts in this section are structured as follows: The experiments are briefly described in Section 6.1. The results of the classification are described in Section 6.2. The comparison of recent technologies will be made in Section 6.3.

6.1. Experimental Settings

Based on the dynamic protocol using the DMR-IR dataset, the proposed approaches were trained for diagnosing breast cancer based on thermal images. The dataset includes 1542 thermal images of 56 patients' breasts, with 762 of them representing breasts that are cancerous and 780 of them healthy cases. A training set of 1302 thermal images and a testing set of 240 thermal images were each split from the dataset. We used image augmentation techniques, such as rotation and brightness change, to increase the training dataset size to overcome the issue of the small dataset size. We rotated the original images 90 degrees. We adjusted the images' brightness, a color augmentation technique, so that the resulting image was lighter than the original images. As a result, we had two augmented copies of each original image. Information regarding the dataset is provided in Table 2.

Table 2. Dataset details.

Dataset	Size	Abnormal Breasts	Normal Breasts
DMR-IR-Original	1542	762	780
DMR-IR-Augmented	2604	1284	1320
Total	4146	2046	2100

As inputs (images) and outputs (labels) of training and test sets were combined, the approaches were trained and tested using a 10-fold cross-validation method. The source data were divided into 10 relatively equal, disjointed segments after stratified sampling. The performance of approaches was tested using a subset of the test sets, which amounted to 415 thermal images and labels, and training sets, which amounted to 3731 thermal images and labels.

The initialization parameters of the approaches were obtained by pretraining of these networks using a binary class of the DMR-IR dataset. We employed the ReLU and sigmoid activation functions, and we trained all AMs using the Adam optimizer. Periodic learning was employed to further increase the accuracy of the image classification findings. The four approaches had a learning rate of 0.001. For the fitting models, the batch size was set to 64 and the number of epochs to 50.

6.2. Classification Results

Through experiments, proposed models employing the VGG16 model with/without the soft, self, and hard AMs were used to classify thermal images of breast cancer (VGG16, VGG16-SF, VGG16-SL, and VGG16-HD). We resized all images to 224×224 pixels, which is the default input size of the VGG16 model. In the majority of situations, downsampling an image to its default size on standard architectures is effective to achieve higher accuracy [29].

To further guarantee that each image was examined in both the test and training sets, we employed stratified 10-fold cross-validation. It also reduced the overfitting-related generalization problems. On the DMR-IR dataset, the proposed models' performances were assessed, and the outcomes were compared to identify the most accurate model.

Each model's accuracy, sensitivity, specificity, Cohen's kappa, and AUC ROC scores are estimated. Table 3 shows the average results of the VGG16 with/without AMs to observe the effect on the performance of the VGG16 model.

Table 3. Model classification results.

Models	Accuracy (%)	Specificity (%)	Sensitivity (Recall) (%)	Precision (%)	F1-Score (%)	AUC	Cohen's Kappa
VGG16	99.18%	98.48%	99.85%	98.55%	99.20%	0.999	0.983
VGG16-SL	99.49%	99.56%	99.42%	99.57%	99.49%	0.994	0.989
VGG16-HD	99.80%	99.75%	99.85%	99.76%	99.80%	0.999	0.996
VGG16-SF	99.32%	99.46%	99.19%	99.48%	99.33%	0.999	0.986

Three AM models with the VGG16 also exhibited convergence in the results. The VGG16-HD indicated that using hard attention with the VGG16 had a higher effect in all metrics.

We found that the use of AMs with the VGG16 model had an improvement effect on VGG16 performance in several metrics, such as accuracy, specificity, precision, and F1-score, while VGG16 achieves a higher result in sensitivity than VGG16-SF and VGG16-SL. In contrast, it has a similar result to VGG16-HD. Nevertheless, the VGG16-SF and VGG16-SL achieved impressive and encouraging results. Sensitivity is a crucial factor in medical applications, as we know; thus, the findings of the suggested model were promising.

In addition, we provided the average AUC ROC of 10-fold cross-validation results, where the AUC score shows the model's ability to differentiate between classes—in our situation, normal and cancer. The VGG16 models with/without the three AMs obtained values that were close to 0.999, while the VGG16-SL was 0.994. In Figure 2, the ROC curves of four approaches for each fold are shown, as is the AUC findings and average. A mighty statistic that can be used to assess intra- and inter-rater reliability is Cohen's kappa. It can have values between -1 and $+1$, similar to correlation coefficients, where 0 represents the degree of agreement expected by chance and 1 represents perfect agreement between the raters. A standardized value can be interpreted similarly across various research in all correlation statistics, including the kappa. To understand the kappa result, Cohen advised using the following scale: values ≤ 0 denoting no agreement, and 0.01–0.20 denoting none to slight, 0.21–0.40 denoting fair, 0.41–0.60 denoting moderate, 0.61–0.80 denoting substantial, and 0.81–1.00 denoting almost perfect agreement [30]. Using 10-fold cross-validation, we provided the aggregate Cohen's kappa scores. The VGG16 models with/without the three AMs showed excellent results in this metric, with scores ranging from 0.983 to 0.996 for all models, which indicated almost perfect agreement.

Each fold included 204 normal and 210 cancer images in the aggregate confusion matrix after 10-fold cross-validation of actual and predicted labels, as illustrated in Figure 3. For the VGG16, the 2015 normal breast class images were correctly classified, with 31 misclassifications, and in the cancer class, the 2097 images were correctly classified, with 3 images misclassified. In contrast, for the VGG16-SL, 2037 normal breast class images were correctly classified, and 9 were misclassified, while in the cancer class, 2088 images were classified correctly, with 12 misclassified. For the VGG16-HD, 2041 normal breast images were correctly classified with 5 misclassified, and 2087 cancer images were correctly classified, and 3 were misclassified. In the VGG16-SF confusion matrix, 2035 images in the normal breast class were correctly classified, while 11 were incorrectly classified. Of the 2083 images in the cancer class, 2083 were correctly classified, while 17 were incorrectly classified.

Figure 4 displays the four tables of the classification report along with the classification results for the VGG16, VGG16-SF, VGG16-SL, and VGG16-HD in terms of accuracy, recall, precision, F1-score, and support, where support is the number of samples.

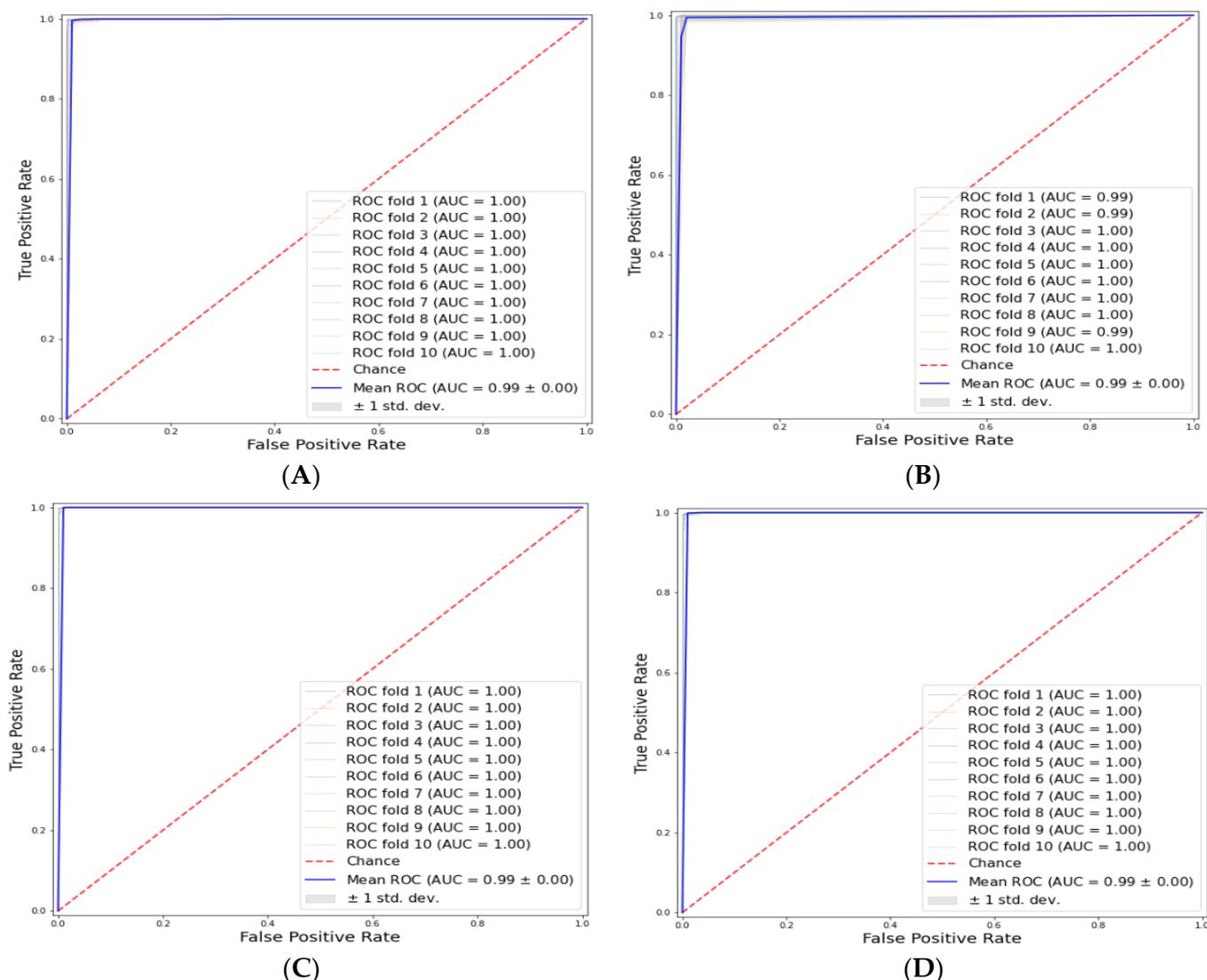


Figure 2. ROC curves of models for breast cancer. The ROC curve of the VGG16 model (A); the ROC curve of VGG16-SL model (B); the ROC curve of VGG16-HD model (C); the ROC curve of VGG16-SF model (D).

As a conclusion and summary of the findings, we can see from Table 3 and Figures 2–4, that early detection of breast cancer by thermal imaging utilizing VGG16 and the suggested AMs has been proven successful. Due to the lack of AMs, the VGG16 had the highest false alarm rate among all models, with a classification rate of 99.18% and a low false alarm rate of 0.81%. In contrast, the VGG16-HD had the lowest false alarm rate among all the models, with a classification rate of 99.80% and a low false alarm rate of 0.19%, while the VGG16-SL had a classification rate of 99.49% and the low false alarm rate of 0.50%. The VGG16-SF classification rate was 99.32%, while its low false alarm rate was 0.67%. The accuracy achieved using the VGG16 models, with/without the three AMs, made it apparent that superb results were obtained in all suggested models. Based on our findings, the proposed models achieved the highest accuracy, where hard attention comes first, then self-attention, soft attention, and finally, the VGG16 model.

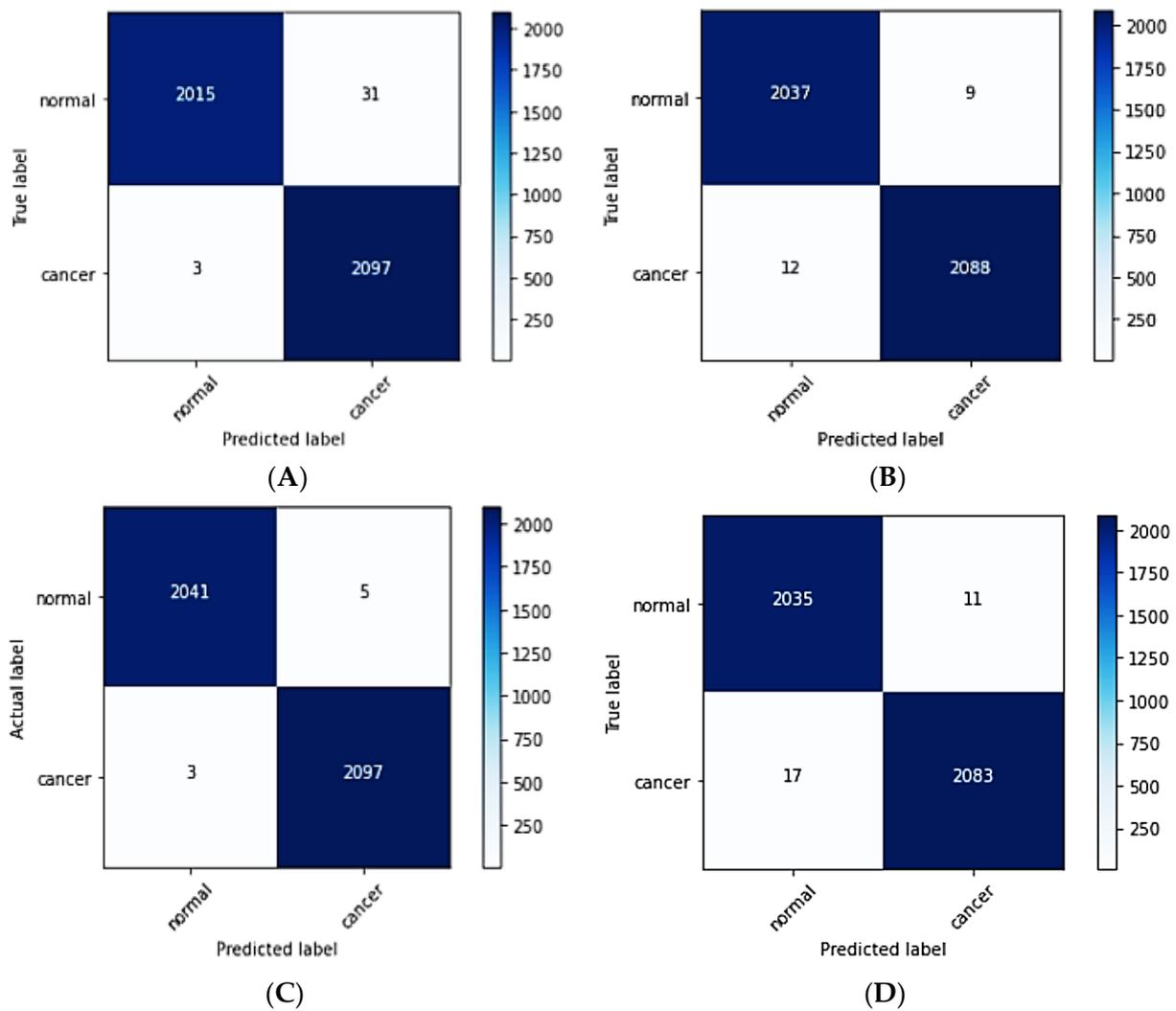


Figure 3. VGG16 model with/without AMs experiment results. VGG16 model (A); VGG16-SL model (B); VGG16-HD model (C); VGG16-SF model (D).

	precision	recall	f1-score	support
0.0	0.98	1.00	0.99	2018
1.0	1.00	0.99	0.99	2128
accuracy			0.99	4146
macro avg	0.99	0.99	0.99	4146
weighted avg	0.99	0.99	0.99	4146

(A)

	precision	recall	f1-score	support
0.0	1.00	0.99	0.99	2049
1.0	0.99	1.00	0.99	2097
accuracy			0.99	4146
macro avg	0.99	0.99	0.99	4146
weighted avg	0.99	0.99	0.99	4146

(B)

	precision	recall	f1-score	support
0.0	1.00	1.00	1.00	2044
1.0	1.00	1.00	1.00	2102
accuracy			1.00	4146
macro avg	1.00	1.00	1.00	4146
weighted avg	1.00	1.00	1.00	4146

(C)

	precision	recall	f1-score	support
0.0	0.99	0.99	0.99	2052
1.0	0.99	0.99	0.99	2094
accuracy			0.99	4146
macro avg	0.99	0.99	0.99	4146
weighted avg	0.99	0.99	0.99	4146

(D)

Figure 4. Aggregate confusion matrix of all folds for each model. VGG16 model (A); VGG16-SL model (B); VGG16-HD model (C); VGG16-SF model (D).

6.3. Comparison with Recent Methods

In this section, we will compare our results in our proposed method of VGG16 with AMs with the most recent study on breast cancer detection, which was based on the same pre-trained model of VGG16 [22]. In addition, in our previous study [24], we investigated the use of CNN with AMs to detect breast cancer using thermographic images and it outperformed the reviewed models in the literature. Thus, we will compare the results of our previous work with the performance of our proposed method of VGG16 with AMs.

6.3.1. Comparison with Other Pre-Trained VGG16 Model

According to our review of the literature, a recent study [22] was conducted using the same pre-trained model of VGG16 on thermal images for breast cancer detection. Thus, we found it necessary to compare our results with those achieved in that study.

Deep multi-view VGG16 achieved the highest accuracy on the dynamic DMR-IR dataset. Their results showed the highest accuracy obtained through multi-view (99%), specificity (100%), sensitivity (98.04%), precision (100%), and F1-score (99.01%). Moreover, when the model was tested on a single-view (frontal), it achieved lower results: accuracy (94%), specificity (92.31), sensitivity (95.83%), precision (92%), and F1-score (93.88%), see Table 4.

On the other hand, in our proposed approach, all models developed (tested on a single view “frontal”) achieved higher accuracy than the multi- and single-view in their study. Where VGG16-HD was the highest with 99.80% accuracy, 99.85% sensitivity, 99.80% F1-score, and competitive results in specificity and precision.

Based on the results in Table 4, AMs showed their ability to improve the performance of the pre-trained VGG16 model, which contributes to enhancing the detection ability of breast cancer using thermal images with more accuracy.

6.3.2. Comparison with Convolutional Neural Network (CNN) Models

In this subsection, we will discuss and compare the results achieved in our previous work with CNN [24] on the same dataset. Table 4 shows the performance results of CNN [24] and VGG16 with and without AMs.

Table 4. Comparison with recent models.

Approaches	Accuracy	Specificity	Sensitivity (Recall)	Precision	F1-Score	AUC	Cohen’s Kappa
Recent study [22]: Deep multi-view VGG16 on thermal images							
VGG16-Multi-View	99%	100%	98.04%	100%	99.01%	-	-
VGG16-Frontal View	94%	92.31%	95.83%	92%	93.88%	-	-
Our previous work [24]: CNN with and without AM on thermal images							
CNN	84.92%	89.61%	89.61%	90.23%	83.91%	0.851	0.69
CNN-SL	99.32%	99.52%	99.52%	99.14%	99.32%	0.999	0.98
CNN-HD	99.49%	99.71%	99.71%	99.28%	99.49%	0.999	0.98
CNN-SF	99.34%	99.21%	99.21%	99.52%	99.36%	0.999	0.98
Proposed approach: VGG16 pre-trained DL method with and without AM on thermal images							
VGG16	99.18%	98.48%	99.85%	98.55%	99.20%	0.999	0.983
VGG16-SL	99.49%	99.56%	99.42%	99.57%	99.49%	0.994	0.989
VGG16-HD	99.80%	99.75%	99.85%	99.76%	99.80%	0.999	0.996
VGG16-SF	99.32%	99.46%	99.19%	99.48%	99.33%	0.999	0.986

As we can see, VGG16 showed a significant improvement over CNN in all metrics. CNN without AMs achieved an accuracy 84.92%, specificity 89.61%, sensitivity 89.61%, precision 90.23%, F1-score 83.91%, AUC 0.851, Cohen's kappa 0.69; while VGG16 without AMs has achieved higher numbers, see Figure 5A.

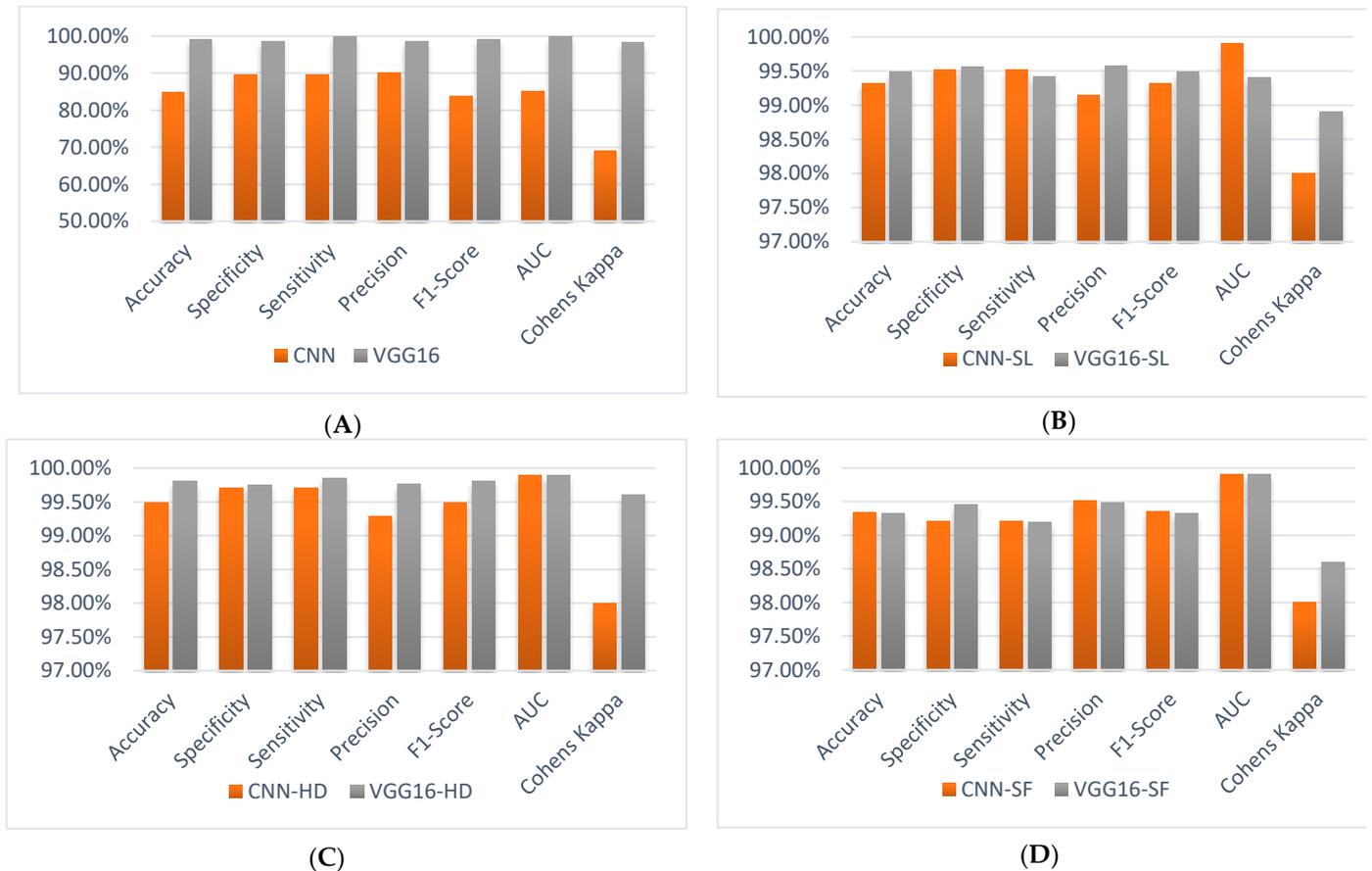


Figure 5. Charts for comparison of CNN and VGG16 results with/without AMs. VGG16 and CNN model (A); VGG16-SL and CNN-SL model (B); VGG16-HD and CNN-HD model (C); VGG16-SF and CNN-SF model (D).

On the other hand, SL with CNN provides an accuracy of 99.32%, a specificity and a sensitivity of 99.52%, a precision of 99.14%, an F1-score of 99.32%, an AUC of 0.999, and a Cohen's kappa of 0.98. VGG16 achieved convergence performance with CNN when using SL. Furthermore, VGG16-SL shows improvement in some metrics, such as accuracy and precision, see Figure 5B.

The use of AMs demonstrated a significant improvement in CNN performance. HD shows the highest performance with CNN with an accuracy of 99.34%, a specificity of 99.21%, a precision of 99.52%, an F1-score of 99.36%, an AUC of 0.999, and a Cohen's kappa of 0.98. The performance of the HD with VGG16 outperformed CNN, as shown in Figure 5C.

The results present a convergence in performance between CNN and VGG16 with SF, see Figure 5D.

We infer from the comparison presented in this subsection that VGG16 achieves superior performance over CNN without AMs and when used with HD, while there is a convergence in the performance of VGG16 and CNN with SF and SL. Consequently, the AMs provides an impressive improvement in the performance of CNN and VGG16.

The results showed an improvement in the performance of the AMs models, with the pre-trained model VGG16 proving better than the CNN model, as Table 4 shows, and this

is due to the fact that the pre-trained models have the ability to perform better, as they are trained on a large dataset (Imagenet) and with more computing resources.

We conclude from the experiments conducted in this paper, that the three AMs (self, hard, and soft) have the ability to improve the performance of the CNN and VGG16 models, as they help focus on the most prominent regions, which contributed to raising the accuracy of the diagnostic thermal images of breast cancer. Moreover, the pre-trained models (VGG16) performed better than simple CNN, encouraging the experience of other pre-trained models in breast cancer detection. From our point of view, the small size of the dataset is one of the limitations of this study, which we tried to address using augmentation techniques. In addition, we believe that experimenting with the proposed models and generalizing them on a larger dataset and for other types of cancers will support our findings.

7. Conclusions

Three AM models with a VGG16 were proposed to apply to thermal imaging images for the diagnosis of breast cancer. Our argument demonstrated how the VGG16 model performed better with AMs in several metrics. Nevertheless, the VGG16 provided great results in sensitivity and precision. Moreover, several related studies were presented and compared with our proposed models, showing outperformed results. The test accuracy results of the suggested models for diagnosing breast cancer were 99.18% for VGG16, 99.49% for self-attention, 99.80% for hard attention, and 99.32% for soft attention. Compared to recent studies that have used the same pre-trained VGG16 model on the same DMR-IR dataset [22], our proposed model outperformed their results, even though we used a single-view image (frontal). In contrast, they used a multi-view (frontal, left, right), providing more information. In addition, when compared to our previous work on CNNs [24] on the same dataset, which, in turn, has outperformed the reviewed models in the literature, our VGG16 model with and without AMs has achieved better results than the counterpart models with CNN. The limited availability of thermal breast imaging datasets is a limitation of our investigation. Additionally, the dataset size is small, and these problems were resolved in our work utilizing augmentation methods. As a result, a future extension of this study would involve experimenting with datasets of huge sizes by combining different datasets and investigating various augmentation methods. The suggested models will be assessed in further work using various biomedical imaging datasets. Another future work direction would be exploring the use of other pre-trained DL models with and without AMs and on multiple datasets.

Author Contributions: Conceptualization, D.A.; methodology, D.A.; software, A.A.; investigation, A.A.; writing—original draft preparation, A.A.; writing—review and editing, D.A.; supervision, D.A.; project administration, D.A.; funding acquisition, D.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research project was funded by the Research Supporting Program Project Number (RSP2023R281), King Saud University, Riyadh, Saudi Arabia.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This research project was supported by a grant from the Research Supporting Program Project Number (RSP2023R281), King Saud University, Riyadh, Saudi Arabia.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Rajinikanth, V.; Kadry, S.; Taniar, D.; Damaševičius, R.; Rauf, H.T. Breast-Cancer Detection Using Thermal Images with Marine-Predators-Algorithm Selected Features. In Proceedings of the 2021 Seventh International conference on Bio Signals, Images, and Instrumentation (ICBSII), Chennai, India, 25–27 March 2021; pp. 1–6.
2. Singh, D.; Singh, A.K. Role of Image Thermography in Early Breast Cancer Detection-Past, Present and Future. *Comput. Methods Programs Biomed.* **2020**, *183*, 105074. [[CrossRef](#)] [[PubMed](#)]
3. Parisky, Y.R.; Skinner, K.A.; Cothren, R.; DeWitney, R.L.; Birbeck, J.S.; Conti, P.S.; Rich, J.K.; Dougherty, W.R. Computerized Thermal Breast Imaging Revisited: An Adjunctive Tool to Mammography. In Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Biomedical Engineering towards the Year 2000 and Beyond, Hong Kong, China, 1 November 1998; Volume 2, pp. 919–921.
4. Wu, P.; Qu, H.; Yi, J.; Huang, Q.; Chen, C.; Metaxas, D. Deep Attentive Feature Learning for Histopathology Image Classification. In Proceedings of the 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy, 8–11 April 2019; pp. 1865–1868.
5. Palminteri, S.; Pessiglione, M. Chapter Five—Reinforcement Learning and Tourette Syndrome. In *International Review of Neurobiology*; Martino, D., Cavanna, A.E., Eds.; Advances in the Neurochemistry and Neuropharmacology of Tourette Syndrome; Academic Press: Cambridge, MA, USA, 2013; Volume 112, pp. 131–153.
6. Rezende, E.; Ruppert, G.; Carvalho, T.; Theophilo, A.; Ramos, F.; Geus, P.D. Malicious Software Classification Using VGG16 Deep Neural Network’s Bottleneck Features. In *Information Technology—New Generations*; Latifi, S., Ed.; Advances in Intelligent Systems and Computing; Springer International Publishing: Cham, Germany, 2018; Volume 738, pp. 51–59. ISBN 978-3-319-77027-7.
7. Tian, H.; Wang, P.; Tansey, K.; Han, D.; Zhang, J.; Zhang, S.; Li, H. A Deep Learning Framework under Attention Mechanism for Wheat Yield Estimation Using Remotely Sensed Indices in the Guanzhong Plain, PR China. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *102*, 102375. [[CrossRef](#)]
8. Shen, T.; Zhou, T.; Long, G.; Jiang, J.; Wang, S.; Zhang, C. Reinforced Self-Attention Network: A Hybrid of Hard and Soft Attention for Sequence Modeling. In Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, International Joint Conferences on Artificial Intelligence Organization, Stockholm, Sweden, 5 July 2018; pp. 4345–4352.
9. de Santana Correia, A.; Colombini, E.L. Attention, Please! A Survey of Neural Attention Models in Deep Learning. *Artif. Intell. Rev.* **2022**, *55*, 6037–6124. [[CrossRef](#)]
10. Rashed, E.; El Seoud, M.S.A. Deep Learning Approach for Breast Cancer Diagnosis. In Proceedings of the 2019 8th International Conference on Software and Information Engineering, ACM, Cairo, Egypt, 9 April 2019; pp. 243–247.
11. Platania, R.; Shams, S.; Yang, S.; Zhang, J.; Lee, K.; Park, S.-J. Automated Breast Cancer Diagnosis Using Deep Learning and Region of Interest Detection (BC-DROID). In Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, ACM, Boston, MA, USA, 20 August 2017; pp. 536–543.
12. Patil, R.S.; Biradar, N. Automated Mammogram Breast Cancer Detection Using the Optimized Combination of Convolutional and Recurrent Neural Network. *Evol. Intel.* **2021**, *14*, 1459–1474. [[CrossRef](#)]
13. Santana, M.A.; Pereira, J.M.S.; Silva, F.L.; Lima, N.M.; Sousa, F.N.; Arruda, G.M.S.; de Lima, R.C.F.; Silva, W.W.A.; Santos, W.P. Breast Cancer Diagnosis Based on Mammary Thermography and Extreme Learning Machines. *Res. Biomed. Eng.* **2018**, *34*, 45–53. [[CrossRef](#)]
14. Gogoi, U.R.; Bhowmik, M.K.; Ghosh, A.K.; Bhattacharjee, D.; Majumdar, G. Discriminative Feature Selection for Breast Abnormality Detection and Accurate Classification of Thermograms. In Proceedings of the 2017 International Conference on Innovations in Electronics, Signal Processing and Communication (IESC), Shillong, India, 6–7 April 2017; pp. 39–44.
15. Ekici, S.; Jawzal, H. Breast Cancer Diagnosis Using Thermography and Convolutional Neural Networks. *Med. Hypotheses* **2020**, *137*, 109542. [[CrossRef](#)] [[PubMed](#)]
16. Mishra, S.; Prakash, A.; Roy, S.K.; Sharan, P.; Mathur, N. Breast Cancer Detection Using Thermal Images and Deep Learning. In Proceedings of the 2020 7th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 12–14 March 2020; pp. 211–216.
17. Mambou, S.J.; Maresova, P.; Krejcar, O.; Selamat, A.; Kuca, K. Breast Cancer Detection Using Infrared Thermal Imaging and a Deep Learning Model. *Sensors* **2018**, *18*, 2799. [[CrossRef](#)] [[PubMed](#)]
18. Mookiah, M.R.K.; Acharya, U.R.; Ng, E.Y.K. Data Mining Technique for Breast Cancer Detection in Thermograms Using Hybrid Feature Extraction Strategy. *Quant. InfraRed. Thermogr. J.* **2012**, *9*, 151–165. [[CrossRef](#)]
19. Tello-Mijares, S.; Woo, F.; Flores, F. Breast Cancer Identification via Thermography Image Segmentation with a Gradient Vector Flow and a Convolutional Neural Network. *J. Healthc. Eng.* **2019**, *2019*, e9807619. [[CrossRef](#)] [[PubMed](#)]
20. Sánchez-Cauce, R.; Pérez-Martín, J.; Luque, M. Multi-Input Convolutional Neural Network for Breast Cancer Detection Using Thermal Images and Clinical Data. *Comput. Methods Programs Biomed.* **2021**, *204*, 106045. [[CrossRef](#)] [[PubMed](#)]
21. de Freitas Oliveira Baffa, M.; Grassano Lattari, L. Convolutional Neural Networks for Static and Dynamic Breast Infrared Imaging Classification. In Proceedings of the 2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), Parana, Brazil, 29 October–1 November 2018; pp. 174–181.
22. Tiwari, D.; Dixit, M.; Gupta, K. Deep Multi-View Breast Cancer Detection: A Multi-View Concatenated Infrared Thermal Images Based Breast Cancer Detection System Using Deep Transfer Learning. *TS* **2021**, *38*, 1699–1711. [[CrossRef](#)]

23. Deng, J.; Ma, Y.; Li, D.; Zhao, J.; Liu, Y.; Zhang, H. Classification of Breast Density Categories Based on SE-Attention Neural Networks. *Comput. Methods Programs Biomed.* **2020**, *193*, 105489. [[CrossRef](#)] [[PubMed](#)]
24. Alshehri, A.; ALSaeed, D. Breast Cancer Detection in Thermography Using Convolutional Neural Networks (CNNs) with Deep Attention Mechanisms. *Appl. Sci.* **2022**, *12*, 12922. [[CrossRef](#)]
25. Silva, L.F.; Saade, D.C.M.; Sequeiros, G.O.; Silva, A.C.; Paiva, A.C.; Bravo, R.S.; Conci, A. A New Database for Breast Research with Infrared Image. *J. Med. Imaging Health Inform.* **2014**, *4*, 92–100. [[CrossRef](#)]
26. CyberZHG Keras-Self-Attention: Attention Mechanism for Processing Sequential Data That Considers the Context for Each Timestamp. Available online: <https://pypi.org/project/keras-self-attention/> (accessed on 12 December 2022).
27. Soft & Hard Attention. Available online: <https://jhui.github.io/2017/03/15/Soft-and-hard-attention/> (accessed on 12 December 2022).
28. Murugan, R.; Goel, T. E-DiCoNet: Extreme Learning Machine Based Classifier for Diagnosis of COVID-19 Using Deep Convolutional Network. *J Ambient Intell Hum. Comput.* **2021**, *12*, 8887–8898. [[CrossRef](#)] [[PubMed](#)]
29. Luke, J.; Joseph, R.; Balaji, M. Impact of Image Size on Accuracy and Generalization of Convolutional Neural Networks. *Int. J. Res. Anal. Rev. (IJRAR)* **2019**. Available online: https://www.researchgate.net/profile/Mahesh-Balaji/publication/332241609_IMPACT_OF_IMAGE_SIZE_ON_ACCURACY_AND_GENERALIZATION_OF_CONVOLUTIONAL_NEURAL_NETWORKS/links/5fa7a715299bf10f732fdc1c/IMPACT-OF-IMAGE-SIZE-ON-ACCURACY-AND-GENERALIZATION-OF-CONVOLUTIONAL-NEURAL-NETWORKS.pdf (accessed on 20 December 2021).
30. McHugh, M.L. Interrater Reliability: The Kappa Statistic. *Biochem. Med.* **2012**, *22*, 276–282. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.