Iranian Journal of Medical Physics

ijmp.mums.ac.ir

A Framework for Promoting Passive Breast Cancer Monitoring: Deep Learning as an Interpretation Tool for Breast Thermograms

Mohamad Firouzmand[1], Keivan Majidzadeh[2] , Maryam Jafari[2], Shahpar Haghighat[2], Rezvan Esmaeili[2], Leila Moradi[1], Nima Misaghi[3] , Mahsa Ensafi[4] , Fatemeh Batmanghelich[3] , Mohammad Reza Keyvanpour[4], Seyed Vahab Shojaedini[1*]

1. Biomedical engineering Department, Iranian Research Organization for Science & Technology, Tehran, Iran
2. Breast Cancer Research Center, Motamed Cancer Institute, ACECR, Tehran, Iran
3. Department of Computer Engineering, Faculty of Engineering, Islamic Azad University E-Campus, Tehran, Iran
4. Department of Computer Engineering, Faculty of Engineering, Alzahra University, Tehran, Iran

ARTICLE INFO

Article type:
Original Paper

Article history:
Received: Apr 10, 2023
Accepted: July 02, 2023

Keywords:
Breast Cancer
Medical Imaging
Artificial Intelligence
Deep Learning
Temperature Mapping

ABSTRACT

Introduction: Several types of cancer can be detected early through thermography, which uses thermal profiles to image tissues in recent years, thermography has gained increasing attention due to its non-invasive and radiation-free nature. There is a growing need for thermographic images of breast cancer lesions in different nationalities and ages to develop this technique, however. This study aims to introduce a dataset of breast thermograms.

Material and Methods: In this study, thermographic images of breast cancer from Iranian samples were prepared and confirmed due to the limited number of breast thermogram databases. The prepared database was tested using artificial intelligence and another well-known DMR database (Database for Mastology Research) in this study to determine its reliability.

Results: A variety of deep learning architectures and transfer learning are used to evaluate these databases for accuracy, sensitivity, speed, training compliance, and validation compliance. According to best-fitted structures for both types of databases, the database obtained from this study has a quality comparable to the DMR reference database, with minimum accuracy, sensitivity, specificity, precision, and F-score of 80%, 86%, 86%, 88%, and 87%, respectively.

Conclusion: Using thermography as a method of early breast screening is demonstrated to be effective**.** In comparison to DMR, the lower statistics of the proposed database (between 2 and 7 percent) indicates that more diverse breast thermograms should be captured in conjunction with improvements to imaging equipment as well as adherence to thermography recording protocols in order to improve the reliability and efficiency of the database.

►*Please cite this article as:*
Firouzmand M, Majidzadeh K, Jafari M, Haghighat Sh, Esmaeili R, Moradi L, Misaghi N, Ensafi M, Batmanghelich F, Keyvanpour MR, Shojaedini SV. A Framework for Promoting Passive Breast Cancer Monitoring: Deep Learning as an Interpretation Tool for Breast Thermograms. Iran J Med Phys 2024; 21: 237-248. 10.22038/ijmp.2023.71683.2268.

Introduction

Among women worldwide, breast cancer is the second most common and most deadly cancer, after lung cancer [1,2]. Early detection of breast cancer greatly improves patient survival [3,4]. Breast screening methods like Mammography have a high false-positive rate and can increase radiation-induced cancer risks [5-7]. Ultrasound is another method for breast screening, but its performance depends on the device and radiologist [8].

Living tissue has heat necessarily. Changes in surface temperature and blood circulation affect body temperature. It is possible to map quickly the distribution of heat throughout the body with infrared cameras. Thermographic cameras differ in their spectral response, response time, and sensitivity. Each pixel in a thermal image is affected by both emitted and reflected radiation [9]. Micro bolometers provide basic thermal imaging technology [9]. There are two main types of thermal imaging cameras: thermal detectors and quantum detectors [10]. Most thermal detectors use uncooled micro-bolometers, which have lower sensitivity. The FLIR Lepton is a widely used long-wave infrared (LWIR) camera module designed for easy integration with standard mobile interfaces [9]. Infrared radiation from body tissues varies continuously, allowing for the measurement of heat distribution and changes. Thermography measures skin heat radiation, aiding in the diagnosis of vascular diseases and tumor types [8].

Early-stage cancer cells produce nitric oxide, leading to arterial dilation and increased blood flow, raising the temperature of affected tissue. Thus, deep breast lesions can also cause temperature changes on the skin surface [11]. Temperature changes are

*Corresponding Author: Tel: +98-21-56276311; Email: shojadini@irost.ir

gaining attention for breast cancer screening. Thus, thermographic imaging is an economical, non-invasive, and painless method for early breast cancer detection, lacking ionizing radiation unlike other imaging techniques. Thermography offers less discomfort, no breast pressure, can identify precancerous areas, and is safe for young or dense breasts, as well as during lactation and pregnancy, increasing its popularity in breast cancer detection over the past decade [8].

Additionally, factors such as menstruation, stress, contraceptive use, pregnancy, and breastfeeding should be considered in final decisions regarding breast screening [12]. The quality of thermal images relies on room and patient conditions, necessitating a standard imaging protocol [6]. Articles [10,13] discuss patient preparation, imaging environments, and post-processing of thermal images.

Despite these advantages, thermography is not yet recognized as the standard breast screening method [5,14]. The main reason is that diagnosing lesions in thermographic images is often more challenging than with conventional imaging methods. Two main techniques have been proposed to diagnose cancerous lesions in breast thermography images.

The first category extracts features from thermographic images for analysis. Some methods use morphological operators to detect subtle temperature changes in cancerous tissue versus background tissue. This technique relies on the normal heat distribution of healthy breast tissue, which shows slight variations from the background. Cancerous tissue temperature is higher than the background and correlates with malignancy grade and extent [15].

Other methods in this category utilize edge detection and gradient operators [16]. Techniques like Wiener filtering and histogram equalization enhance contrast, improving thermographic image interpretation. Furthermore, the estimation of tissue features by using a co-occurrence matrix [17], and attributes such as mean, standard deviation, entropy, skewness, and kurtosis have been examined in this family of techniques [18]. Other extracted features in this category include entropy, dual entropy, histogram-based features, and central computing-based features like center of gravity and geometric center [16].

The second category focuses on classification methods, with fuzzy approaches extensively examined for analyzing breast thermographic data. In [17], thermography images were used to detect breast cancer using statistical features of both breasts and a fuzzy classifier. Additionally, classifiers like simple Bayesian, support vector machines, hill-climbing algorithms, and decision trees are commonly used in this category [18]. Neural networks have also proven effective for classification in this category. Research [19] proposes a system that classifies breast thermographic images as normal or abnormal using a multilayer neural network with eight statistical features.

Classification methods primarily require effective feature extraction from thermograms. Moreover, manual feature extraction is difficult, time-consuming, and requires expertise. Thus, the results of computer-aided detection using these classifiers heavily depend on the feature extraction stage. In recent years, deep learning has emerged as an effective technique for classifying complex data. Deep learning offers several advantages, including direct feature extraction from training data, reduced complexity in feature selection, and the ability to perform feature extraction, selection, and classification within the same architecture [20]. Convolutional Neural Networks (CNNs) are the most common deep learning algorithms for medical image analysis [21].

Article [22] suggests combining thermography and deep learning to improve early breast cancer detection. The authors of [23] propose a CAD method to classify patients as cancer, no cancer, or non-cancerous. In article [24], the authors present an automatic method for extracting the breast area and classifying thermograms as normal or abnormal using deep learning. The authors of [25] propose a novel method for early breast cancer detection by combining thermal images with personal and clinical data. The authors of [26] propose a mobile self-screening framework for early breast cancer detection based on temperature characteristics.

Training CNNs requires a large amount of labeled data, which is complex and costly to collect in medical applications. This challenge can be addressed using transfer learning from pre-trained CNN networks [21], applying knowledge from large datasets to similar problems.

According to [27], transfer learning is followed by fine-tuning to save computational resources and data augmentation to tackle data scarcity. In article [28], the pre-trained DenseNet121 model serves as a feature extractor for the classifier. The authors use Prewitt and Roberts edge detectors on thermal breast images prior to feature extraction. The authors of [29] use three CNNs with transfer learning to classify thermography images as sick or healthy. The authors of [30] fine-tuned CNN models for breast cancer detection using ResNet101, MobileNetV2, and ShuffleNetV2.

Deep neural networks exhibit high predictive accuracy in breast cancer radiology due to their robustness and scalability [31]. Deep learning models like CNNs can classify thermograms from thermography as normal or abnormal, akin to frameworks in studies [5,7]. A key limitation of deep learning for breast thermograms is the need for a large dataset. This issue is rarely addressed in studies, emphasizing the need for a new national database for thermography breast cancer analysis. To address this issue, this study introduces a database of breast

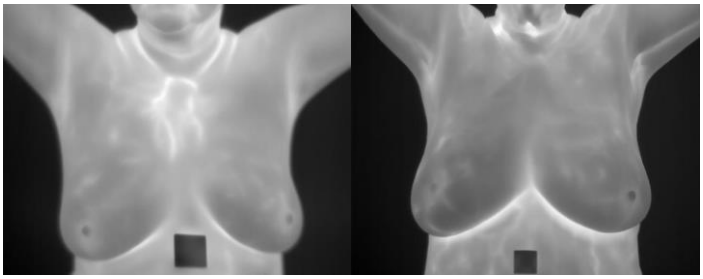
thermograms from Iranian women. According to this study, accuracy, sensitivity, specificity, precision, and F-score had to be at least 80%, 86%, 86%, 88%, and 87% respectively. Pre-trained deep learning architectures were validated using the introduced and a reference database, assessing accuracy, sensitivity, speed, and training compliance.

Materials and Methods

Data

Two datasets of thermographic images were used, with the first sourced from the DMR-IR database [32]. The dataset contains infrared (IR) images and clinical data, including healthy and sick individuals, using 640 x 480 pixel IR images [33]. IR images in DMR are captured using FLIR (Forward Looking Infrared) SC620 thermal cameras, with a sensitivity of less than 0.04 °C and a temperature range of 40 to 500 °C. This database offers IR imaging from five views: frontal, left 45°, left 90°, right 45°, and right 90°, using static and dynamic protocols [33]. Figure 1 displays some breast thermograms from the DMR-IR database.

The second dataset, the Jihad database, was obtained from a collaboration between Research Organization for Sciences and Technology (IROST) and the Motamed Cancer Institute, (MCI). Figure 2 shows some breast thermograms in Jihad database. Over two years, 407 volunteers provided thermograms for this dataset through a static protocol, created under ethics approval IR.ACECR.IBCRC.REC.1398.009. The imaging device was a VisIR 640 thermal camera from Thermoteknix Systems Ltd. The dimensions of the thermograms in Jihad database are 1008 by 528. The Jihad database includes ultrasound imaging alongside thermography, as well as ultrasound, mammography, and pathology reports. To label the data, breast processes in the pathology reports must be identified and their status determined. A breast process is classified into four categories: normal, benign, high-risk benign, and malignant. In Table 1, the two databases are compared in detail. The Thermoteknix VisIR 640 camera has a resolution of 480x640, a spectral range of 7.5 to 13 µm, and can detect temperatures from -20 to 350 °C.



a) b)

Figure 1. Examples of DMR-IR database. a. Healthy thermogram b. Sick thermogram

a) b) c)

Figure 2. Examples of Jihad database a. The left breast is benign and the right breast is normal b. The left breast is normal and the right breast is malignant c. The both of breast are benign

Table 1. Details of available images for each database

	DMR-IR	Jihad
The number of patients	296	407
Age range	25 - 123	14 - 75
The period for taking images	2007-2020	2017-2019
The number of smoker	13	30
The number of patient with history of surgery	27	24
The number of patient with history of biopsy	111	62
The number of patient with history of radiography	44	13
The number of patient with family history	13	160
The number of patient with products at breasts or armpits region	65	110
The number of patients with a history of alcohol consumption	0	10

The proposed method

This study explores various deep learning methods for analyzing thermographic images of breast cancer. To this end, convolutional neural network-based methods are employed. This section first examines the function of convolutional neural networks, followed by an exploration of the capabilities of transfer learning modified versions of this architecture.

Convolutional neural network

Deep learning, a subset of machine learning and AI, uses a hierarchical structure for feature extraction from raw input data, unlike classical neural networks. Convolutional neural networks are essential for image analysis, including face, text, human body, and biological image identification. Figure 3 shows that this structure consists of convolutional, pooling, and fully connected layers [34]. In this architecture, the convolution layer is the first applied to the input image, where various filters slide over it as mathematical kernels, extracting unique features from the raw image.

As shown in figure 3, the result of the internal multiplication of the input image and the filter, along with the application of the activation function, forms the pooling layer [35]. In the next step, the pooling layer performs nonlinear sampling of features, reducing computational complexity by minimizing dimensions and dividing the input into smaller parts. Depending on the reduction type, the average, median, or maximum of each part is replaced by an index value [36]. Finally, the fully connected layer, functioning as a conventional perceptron, classifies features extracted from previous layers, with each input node connected and weighted. The output is the sum of the products of inputs and their weights, with a threshold set by the activation function, which may be a simple or multilayer perceptron or other classifiers [19].

Transfer learning

The structure of deep neural networks, including the convolutional network described earlier, consists of various parameters, and effective learning requires proper parameter setting. Optimal performance is achieved by deep learning models trained on large volumes of annotated data [37]. Large datasets are essential for optimal network performance, but often hard to obtain in applications like medical image processing, necessitating the use of trained networks. This limitation can be addressed through transfer learning, an effective and useful method. Transfer learning enables a model for a similar application to initialize its weights from a pre-trained model, potentially improving performance [38]. Thus, the learning process does not start from scratch, as the convolution layer weights are frozen while the fully connected layer is updated based on the database [29]. Key benefits of transfer learning include improved classification accuracy and faster training processes [39].

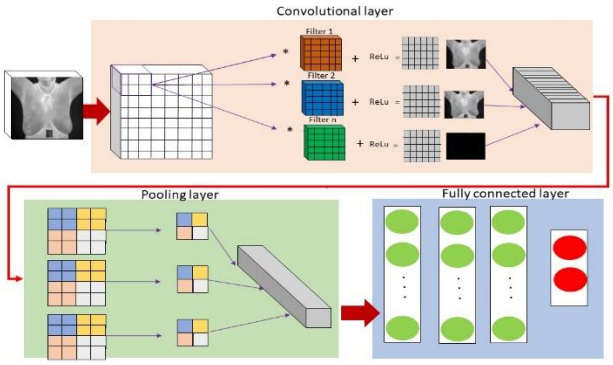


Figure 3. Description of the basic structure of the convolutional neural network

Transfer learning can be implemented in two main ways: as a feature extractor, which freezes the convolutional layers, or through fine-tuning, which updates the model parameters during training [40]. The present study employs a feature extractor approach, utilizing well-known pre-trained networks such as VGG16, VGG19, ResNet50, and InceptionV3. These models were trained on the ImageNet database, which includes over 14 million images across 1,000 classes [40]. Tables 2, 3, and 4 show the details of the architectures used in this study. Models VGG16 and VGG19 belong to the VGG (Visual Geometry Group) family. A major disadvantage of VGG is its large number of parameters that need to be trained [39]. The Inception V3 with was developed by the Google research team. By factoring large convolution layers into smaller ones, this network reduces the number of parameters [40]. The ResNet50 model prevents overfitting by using identity mapping, allowing the model to bypass unnecessary CNN weight layers [41].

Table 2. The detail of the trainable layers

Network	Internal structure	Number of parameters	Input size
VGG16(Visual Geometry Group)	5 convolution layers, and 3 max pooling layers	138 million	244 × 244
VGG19(Visual Geometry Group)	5 convolution layers, and 3 max pooling layers	138 million	244 × 244
Xception	8 MBconv block, and 1 convolution block	23 million	299 × 299
ResNet50	5 convolution layers	23 million	244 × 244
InceptinV3	4 convolution layers	25 million	299 × 299
MobileNet	6 convolution layers	13 million	244 × 244
DenseNet121	3 convolution layers	7,628,484	244 × 244
EfficientNetB0	4 convolution layers	11 million	244 × 244

Table 3. The detail of the used parameters

Optimizer	Batch size	Activation function	Loss function
Adam	32	ReLu	Binary_Crossentropy

Table 4. The detail of the non-pre-trained CNN (Convolution Neural Network)

Number of layers	6
Convolution 1	[2-D conv(3x3), 16, 2-D conv (3x3), 16]
Convolution 2	2-D conv (3x3), 32 2-D conv (3x3), 32]
Convolution 3	2-D conv (3x3), 64 2-D conv (3x3), 64]
Activation function	ReLu

To comprehensively test the capability of deep neural networks in thermographic image modeling, this study uses nine models, starting with a CNN without training. For the other models, transfer learning neural networks from the Keras library are applied, including VGG16, VGG19, Xception, ResNet50, InceptionV3, MobileNet, DenseNet121, and EfficientNetB3.

The proposed algorithm was applied to the mentioned databases through random selection of patients, forming three sets: training, validation, and test. In both scenarios, 60% of the data is allocated for training, 20% for validation, and 20% for testing. In the Jihad dataset, normal patients are classified as healthy, while benign high-risk patients and malignant patients are considered sick. In the DMR-IR dataset, thermograms are labeled as healthy or sick, so benign thermograms are excluded in the second scenario. Additionally, due to the higher number of healthy individuals in both databases, the images of this group were reduced to match the number of the sick group, and unlabeled thermograms were subsequently removed for better neural network performance.

As mentioned earlier, in the Jihad database, right and left breast thermographic images were separated, and network training or testing was conducted using images from the same breast. This separation is important because one breast may be healthy while the other is sick. In the DMR-IR database, however, it is not specified which breast is healthy or sick. Thus, each thermogram of a sick individual was labeled as sick, while each healthy individual was labeled as healthy.

Figure 4 displays the different thermograms in the DMR-IR and Jihad databases. In the Jihad database, as shown, the right and left breasts in each image are cropped and labeled as healthy or diseased based on available pathology reports before entering the neural network.

Given these issues and the use of two distinct thermographic databases, the study was conducted in two separate scenarios. This approach not only evaluates the efficiency of deep learning methods across these databases but also justifies the differences in results based on the variations between the databases.

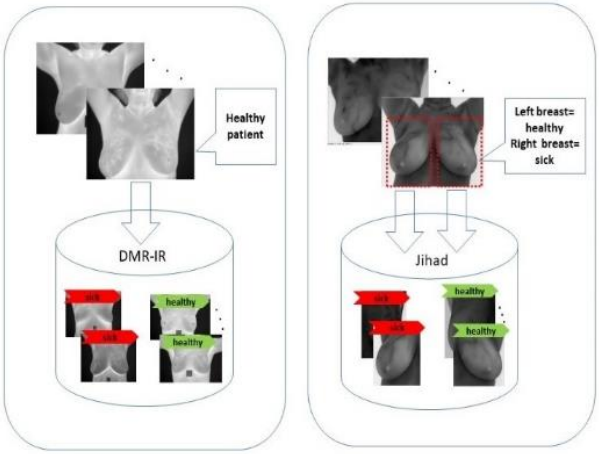


Figure 4. Example of thermograms: Left: Collection of thermograms of jihad database separately for each beast and according to available pathology reports. Right: Collection of DMR-IR database thermograms by each patient according to available pathology report

Statistical analysis

Following the completion of testing, various potential outcomes may arise when validating the results. One scenario is when the deep learning model correctly identifies an individual as sick, known as a true positive. Misdiagnoses are termed false positives. Conversely, correct identifications of healthy individuals are called true negatives, while incorrect identifications of healthy individuals are referred to as false negatives. Thus, the sensitivity parameter represents the percentage of correctly diagnosed sick individuals from the total actual patients in equation 1 [39].

$$sensitivity = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (1)$$

In equation 2, the specificity parameter indicates the percentage of true healthy individuals among total healthy samples, reflecting the classifier's ability to identify genuine healthy cases [39].

$$Specificity = \frac{True\ Negaive}{True\ Negative + False\ Positive} \quad (2)$$

In equation 3, the accuracy parameter represents the ratio of correct diagnoses (both healthy and sick) by the classifier to the total samples, reflecting the classifier's overall diagnostic capability [39].

$$Accuracy = \frac{True\ Positive+True\ Negative}{True\ Positive+True\ Negative+False\ Positive+False\ Negative} \quad (3)$$

Positive cases correctly identified are measured by the precision parameter, which indicates the percentage of individuals known to be sick who are accurately diagnosed as such. This value is derived from equation 4 [39].

$$precision = \frac{True\ Positive}{True\ Positive+\ False\ Positive} \quad (4)$$

The F-score, obtained from equation 5, represents the sensitivity and precision by balancing them through harmonic means to mitigate the influence of extreme values [33].

$$F-score = \frac{2\times precision\times recall}{precision+recall} \quad (5)$$

Results

Table 5 and Table 6 show the number of thermograms used in each database. The difference between the number of available and used images was due to the lack of biopsy, low quality, and wrong angle of capture in different groups of images.

The first scenario involved an examination of the DMR. Table 7 shows the results of the above comparisons.

In the second scenario, the Iranian Jihad database was examined. Table 8 shows the results of the above comparisons.

Table 5. The number of used thermograms in DMR-IR (Database for Mastology Research) database

	Training	validation	Test
Healthy	33	11	11
Sick	26	9	9

Table 6. The number of used thermograms in Jihad database

	training	validation	Test
Healthy	90	30	30
Sick	50	16	18

Table 7. Results of the DMR-IR (Database for Mastology Research) database

Model	Accuracy (%)	sensitivity (%)	Specificity (%)	precision (%)	F-score (%)
CNN	63.15	37.5	81.81	60.00	46.15
VGG16(Visual Geometry Group)	73.68	50.00	90.90	80.00	61.53
VGG19(Visual Geometry Group)	73.68	62.50	81.81	71.42	66.66
Xception	89.47	87.50	90.90	87.50	87.50
ResNet50	42.10	37.50	45.45	33.34	35.29
Inception	57.89	37.50	72.73	50.00	42.85
MobileNet	73.68	62.50	81.81	71.42	66.66
DenseNet121	84.21	75.00	90.90	85.71	80.00
EfficientNet	67.30	61.53	73.07	62.03	61.77

Table 8. Results of the Jihad database

Model	Accuracy (%)	sensitivity (%)	Specificity (%)	precision (%)	F-score (%)
CNN	71.73	78.26	65.21	79.04	78.65
VGG16(Visual Geometry Group)	78.26	86.95	69.56	88.12	87.53
VGG19(Visual Geometry Group)	80.43	78.26	82.60	79.05	78.65
Xception	76.08	78.26	73.91	78.53	78.40
ResNet50	67.39	73.91	60.86	74.66	74.28
Inception	63.04	65.21	60.86	66.27	65.73
MobileNet	78.26	69.56	86.95	69.59	70.05
DenseNet121	67.39	73.91	60.86	75.06	74.48
EfficientNet	58.69	52.17	65.21	54.78	62.92

Discussion

Breast thermography data were analyzed across two separate databases using nine deep architectures, with biopsy results serving as the standard reference for diagnostic validation in both scenarios. To evaluate and compare the results of both scenarios, the parameters of accuracy, sensitivity, specificity, precision, and F-score were utilized. In calculating the mentioned parameters, the biopsy result was used as a reference due to its greater reliability compared to results from other modalities, such as mammography or sonography.

Data processing based on deep learning paradigm were performed on the TensorFlow platform (version 2.10.0) by using the Tensor Processor Unit hardware developed by Google Colab. Computers with a Core I7-7700 processor and 16 GB of RAM, and a 2 GB graphics card (NVIDIA GeForce GT 710) were used in order to performing simultaneous calculations.

In the first scenario, the DMR-IR database was analyzed, with training and testing of deep models conducted exclusively using data obtained from the frontal view. This strategy was implemented because the second database (Jihad) contains only images taken from the same view, allowing for a more meaningful comparison between the results of the two databases. High sensitivity and minimal false-negative errors are crucial in medical decision-making.

Results reveal the CNN-based model without transfer learning is significantly weak, as the available thermographic data is insufficient for effective training. Since access to a large amount of thermographic data is often unlikely, employing transfer learning methods is essential to avoid overfitting or under-fitting the model. Another result supports the use of deep learning and thermographic data for breast cancer diagnosis, as confirmed by several evaluated models. Table 7 shows that the Xception method achieved higher quality for the test database, with an accuracy approximately 5% better than its nearest competitor, DenseNet121. Its sensitivity is about 12% higher than DenseNet121, the following best sensitivity. This model matches or exceeds all reported methods in specificity; however, the current results only indicate its potential for cancer diagnosis. For instance, Table 7 shows that the recent model's 87.5% sensitivity indicates that over 12% of breast cancer cases may go undetected, highlighting the need for further research to enhance performance. Similarly, achieving 90.90% for the specificity parameter indicates the diagnosis of the disease for about 10% of healthy people, which is not as challenging as the previous case but indicates the need to use more images or more complex networks. In addition, Table 9 shows the comparison of the results obtained in this scenario with several studies that have used this database.

In the second scenario, the Iranian Jihad database was analyzed, using only frontal data for training and testing, as in the first scenario. Table 8 shows the results of the above comparisons. The interpretation of results using these parameters indicates that the CNN-based model without transfer learning performs poorly. It is evident that no method is definitively superior for the database in question. Therefore, the accuracy of the VGG19 architecture was about 2% better than that of its closest competitors. However, regarding specificity and sensitivity, this method did not perform the best. MobileNet and VGG16 outperformed it, with VGG16 showing an 8% higher sensitivity than its closest alternative, while MobileNet had a 4% advantage in specificity over its nearest competitor.

The results from both scenarios indicate that the deep learning framework has the potential to effectively distinguish between healthy and cancerous thermographic images. Achieving 90%, 87%, 89%, 87%, and 87% for specificity, sensitivity, accuracy, precision, and F-score, respectively, indicates that the DMR-IR data effectively aids in diagnosing healthy and sick individuals. However, the sensitivity parameter in this database reveals a 15% error in identifying sick patients as healthy, indicating the need for a larger database and potentially more complex deep network. The Jihad database in the best structure obtained 86%, 86%, 80%, 88% and 87% values for the five parameters of specificity, sensitivity, accuracy, precision and F-score respectively. The values of the criteria are significantly weaker—by 2 to 7%—compared to those in the DMR-IR data. Nonetheless, a consistent trend in the parameter values from DMR-IR testing is evident in the Jihad database, indicating that the deep learning framework for modeling and analyzing this data also possesses the necessary potential. The lower results related to the Jihad database compared to DMR-IR can be attributed to two main factors: first, the Jihad database has fewer patients, and second, it contains more images of lower quality and higher noise. The limited number of images likely contributes significantly to the reduced performance observed in this dataset. Another noteworthy point is that the models yielding the best results in the two tested scenarios were not necessarily the same or similar; for instance, the Xception deep model stood out in this regard. While it achieved the best results among competitors for the DMR-IR data, the same model performed weaker than its alternatives in the Jihad database.

Table 9. Comparison of the results obtained with several studies

Model	Accuracy (%)	sensitivity (%)	Specificity (%)	precision (%)	F-score (%)
ours	89.47	87.50	90.90	87.50	87.50
[42]	80.00	83.33	77.77	71.43	76.89
[43]	73.10	92.00	53.00	-	-
[44]	88.89	-	-	-	-

Conversely, the VGG19 model yielded the best results for the Jihad database, while its performance on the DMR-IR database was relatively poor. This difference can be attributed to the insufficient number of training images, which may lead each examined network structure to acquire optimal weights for specific information only after transfer learning from pre-trained networks. These pre-trained networks may deliver better results for specific tasks based on their training data, data structure, and network architecture. For example, the VGG16 model outperforms the VGG19 model by a few hundredths of a percent on ImageNet data, despite VGG19 having a more complex structure with more layers than VGG16. These results indicate that achieving an optimal deep model for distinguishing healthy and cancerous thermographic images depends on the type and characteristics of the image data, rather than adhering to a single architecture. A similar limitation has been observed in other applications of deep learning [18]. In applying this research, two crucial factors emerge: first, the non-uniqueness of deep architectures, and second, the type of images used for network pre-training in transfer learning, specifically their compatibility with breast thermography images.

Learning curves for these models in both scenarios are shown in figure 5 and 6. The diagram for each model expresses the training process performed in that model for DMR-IR and the Jihad data simultaneously. As can be seen, the superiority of the first database, argued in the previous lines, has been significantly illustrated in these figures.

Given the random nature of deep neural network parameters, particularly at the start of training, measuring the variance in the output of these networks is a crucial factor in assessing model performance. Calculating this parameter from 30 runs for the networks are shown in figure 7 and 8. Thus, Figure 7 reports the values of these scatterings for the models trained in the first scenario, and figure 8 reports the values in the second scenario.

Figure 7 shows that in the case of the DMR-IR database, the VGG16 and DenseNet121 models had the least scattering, indicating that in these cases, the relevant neural network performed a more concentrated result in most tests. Based on this, as it turns out, the DenseNet121 architecture, which was the best case in figure 8, also performed well in most applications.

Figure 8 shows that the ResNet50, EfficientNetB3, and DenseNet121 models in the Jihad database had the lowest scattering rates after CNN. It should be noted that in the non-transfer learning CNN method, all the results have been poor and the different executions have not resulted in very different results. This underscores the fact that multiple applications of this method on small data do not lead to practical training.

Alongside accuracy parameters for learning and testing, execution speed is a vital criterion for evaluating various processing methods, primarily due to the bulk and inherent slowness of deep learning techniques. This factor is doubly critical in studies such as this research. Based on this, the time taken to train the nine mentioned models on the same hardware system reflects the duration required for each model's training. Based on the values in figure 9, it can be concluded that the MobileNet neural network is approximately 1 second faster than its nearest competitor, the CNN without transfer learning. However, if we exclude this model from the comparison, the CNN without transfer learning has provided unacceptable results, as noted earlier in the section. In that case, the MobileNet deep model completes the training process about 8 seconds faster than the next alternative, VGG16. Considering that this model is more straightforward than other neural networks, its higher speed may be explained. This aspect of its performance can be attributed to the simpler architecture of this network than the other mentioned neural networks. Similarly, for the EfficientNetB3 and ResNet50, due to their different architectures the fact that they need more time to train may be explained. An important point is the required time to train the networks with the best performance in both databases in this study. It can be seen that the Densnet121 models (optimal method for DMR-IR data) and VGG19, VGG16, and MobileNet models (superior method for jihad data) had good performance due to the average speed.

Another important measure in evaluating the performance of neural networks used in this study is the study of learning curve behavior for both used databases. Figure 5 and 6 show the training and validation curves for both DMR-IR and Jihad databases, respectively. The curves for the first database indicate that in the DMR-IR, the validation curve in all tested networks had both a decreasing trend. This phenomenon indicates the absence of problems such as overfitting. Although the above trend may be observed in all used architectures, in particular, the DenseNet121 neural network has been learning from the network parameters until the last execution of the program, and it is completely aligned with the best results obtained by this network. In the learning diagrams of other models, including ResNet50, Inception V3, Xception, VGG16, and EfficientNetB5, although the network learning curves are relatively similar to DenseNet121, in the validation process, these neural networks, especially in the last epoch, have not been able to achieve the alignment in the DenseNet121 diagrams. This is more evident in the case of the comparing the DenseNet121 learning and validation curves with other models from epoch 15 onwards. Based on learning and validation curves, most of the tested models showed a decreasing trend in Jihad database. However, by examining the curves reported in figure 6, it is clear that the VGG19 neural network has the best performance in validating and learning. This is more evident in comparing the VGG19 training and validation curves with other structures from epoch 15 onwards. This issue is aligned with the best results obtained by this network in the test scenario with the jihad database described in the previous sections. As can be seen in the learning

diagrams of other architectures in figure 6, such as VGG16, ResNet50, Xception, InceptionV3, and EfficinetNetB5, these networks did not perform better in the validation process than VGG19. It is especially true in cases like CNN with poor results from the earlier discussed architectures.

Conclusion

In this research, a new database of breast cancer thermograms in Iran has been prepared and is being validated. In addition, nine well-known deep architectures such as Xception have been used to evaluate the proposed database.

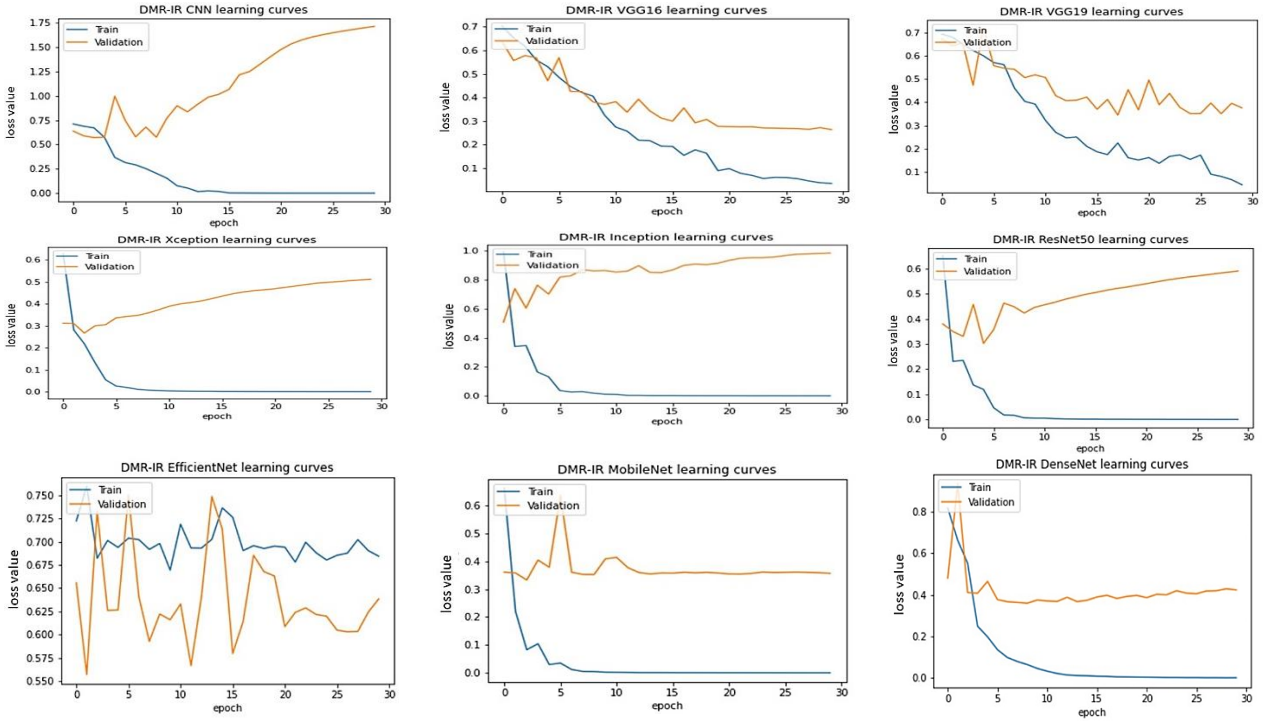
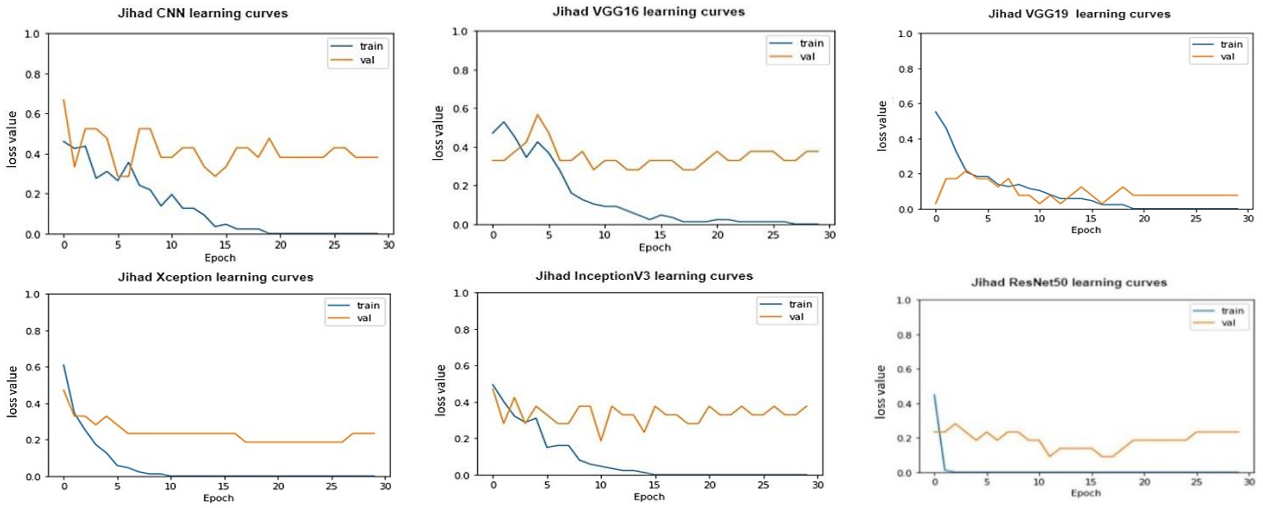


Figure 5. DMR-IR database training diagrams



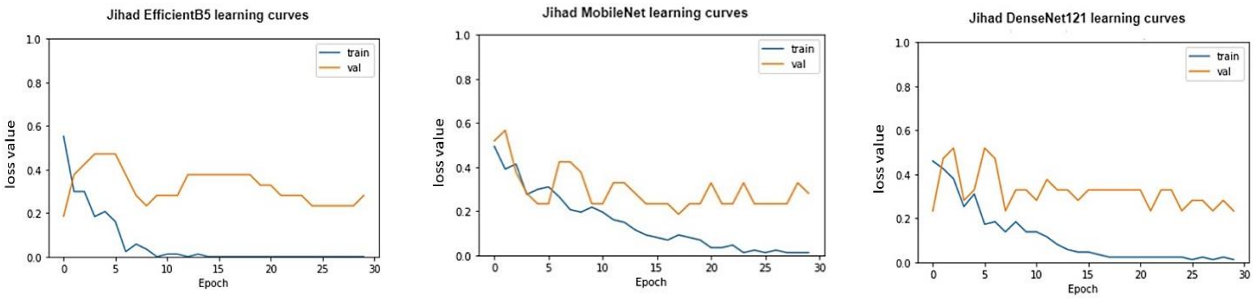


Figure 6. Jihad database training diagrams

Besides the proposed database, the DMR-IR database has also been used to achieve a plenary result. Obtaining 90%, 87%, 89%, 87% and 87% for specificity, sensitivity, accuracy, precision and F-score parameters for DMR-IR data, respectively, showed the considerable potential of deep learning methods in interpreting this data. Meanwhile, the results obtained from the Iranian database also obtained values of 86%, 86%, 80%, 88% and 87% for the above five parameters, respectively. Based on the results of the study, artificial intelligence and deep learning may be applied as a confirmation method to introduce thermography as a confirmatory breast cancer diagnosis method. The study will be further investigated by taking more images, using a more suitable thermal imaging device, and providing better temperature conditions during the imaging process.

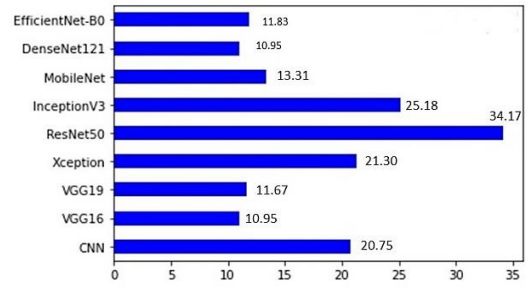


Figure 7. The variance of DMR-IR database

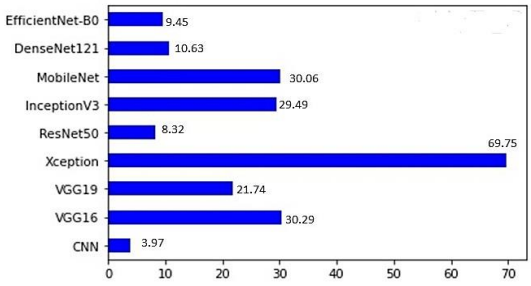


Figure 8. The variance of Jihad database

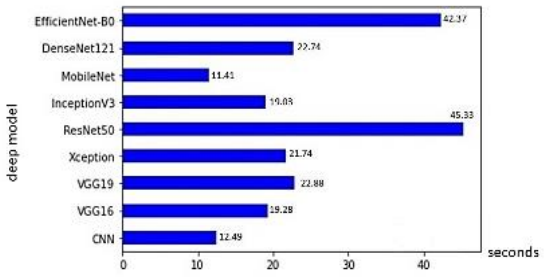


Figure 9. Duration of training (vertical axis is the name of the method - the horizontal axis is seconds)

Acknowledgment

All authors have been personally and actively engaged in the significant work that led to the development of this manuscript, and will collectively and individually be responsible for its content.

References

1. Fouladi N, Pourfarzi F, Amani F, Ali-Mohammadi H, Lotfi I, Mazaheri E. Breast Cancer in Ardabil Province in the North-West of Iran: an Epidemiological Study. Asian Pacific Journal of Cancer Prevention. 2012; 13:1543-5
2. Díaz-Cortés MA, Ortega-Sánchez N, Hinojosa S, Oliva D, Cuevas E, Rojas R, et al. A multi-level thresholding method for breast thermograms analysis using Dragonfly algorithm. Infrared Physics & Technology. 2018; 93:346-611.
3. Roslidar R, Rahman A, Muharar R, Syahputra M.R, Arnia F, Syukri M, et.al . A Review on Recent Progress in Thermal Imaging and Deep Learning Approaches for Breast Cancer Detection. IEEE Access. 2020; 8:116176 –94.
4. Baffa MD, Lattari LG. Convolutional Neural Networks for Static and Dynamic Breast Infrared Imaging Classification. 2018;31st Conference on Graphics, Patterns and Images (SIBGRAPI):174-81.
5. Ekici S, Jawzal H. Breast cancer diagnosis using thermography and convolutional neural networks. Medical Hypotheses. 2020;37:109542.
6. Zuluaga-Gomez J, Zerhouni N, Al Masry Z, Devall C, Varnier C. A survey of breast cancer screening techniques: thermography and electrical impedance tomography. Journal of Medical Engineering & Technology. 2019; 43(5):305-22.
7. Mambou S. J, Maresova P, Krejcar O, Selamat A, Kuca K. Breast Cancer Detection Using. Infrared Thermal Imaging and a Deep Learning Model. Sensors (Basel Switzerland). 2018; 18(9):2799.

8. Ensafi, M., Keyvanpour, M.R. & Shojaedini, S.V: ABT: a comparative analytical survey on Analysis of Breast Thermograms. Multimed Tools Appl, 2023. 83, 53293−53346.
9. Stančić I, Kuzmanić Skelin A, Musić J, Cecić M. The Development of a Cost-Effective Imaging Device Based on Thermographic Technology. Sensors. 2023; 23(10):4582.
10. Kandlikar SG, Perez-Raya I, Raghupathi PA, Gonzalez-Hernandez J-L, Dabydeen D, Medeiros L, et al. Infrared imaging technology for breast cancer detection − current status, protocols and new directions. Int J Heat Mass Transf. 2017;108:2303−20.
11. Wang SH, Muhammad K, Phillips P, Dong Z, Zhang YD. Ductal carcinoma in situ detection in breast thermography by extreme learning machine and combination of statistical measure and fractal dimension. Journal of Ambient Intelligence and Humanized Computing. 2017:1-11
12. Mashekova A, Zhao Y, Ng EYK, Zarikas V, Fok SC, Mukhmetov O. Early detection of the breast cancer using infrared technology—A comprehensive review. In: Thermal science and engineering progress 2022; 27:101142.
13. Resmini R, Silva LF, Medeiros PR, Araujo AS, Muchaluat-Saade DC, Conci A. A hybrid methodology for breast screening and cancer diagnosis using thermography. Computers in Biology and Medicine. 2021. 135:104553.
14. Amri A, Pulko SH, Wilkinson AJ. Potentialities of steady-state and transient thermography in breast tumor depth detection: A numerical study. Computer Methods and Programs in Biomedicine. 2016; 123:68-80.
15. Tang X, Ding H, Yuan Y, Wang V. Morphological measurement of localized temperature increase amplitudes in breast infrared thermograms and its clinical application. Biomedical Signal Processing and Control. 2008;3(4):312-8.
16. Kapoor, P., Prasad, S. V. A. V., & Patni, S. Automatic analysis of breast tomograms for tumor detection based on biostatistical feature extraction and ANN. IJETED. 2012; 7(2): 2249-6149.
17. Schaefer G, Závišek M, Nakashima T. Thermography based breast cancer analysis using statistical features and fuzzy classification. Pattern Recognition. 2009; 42(6):1133-7.
18. Nicandro CR, Efrén MM, Maria Yaneli AA, Enrique MD, Hector Gabriel AM, Nancy PC, et al. Evaluation of the Diagnostic Power of Thermography in Breast Cancer Using Bayesian Classifier, Computational and Mathematical Methods in Medicine. 2013; 2013(1):264246.
19. Lessa V, Marengoni M. Applying Artificial Neural Network for the Classification of Breast Cancer Using Infrared Thermographic Images. International Conference on Computer Vision and Graphics (CCVG). 2016; 9972:429-38
20. Jiménez-Gaona Y, Rodríguez-Álvarez MJ, Lakshminarayanan V. Deep-Learning Based Computer-Aided Systems for Breast Cancer Imaging: A Critical Review. Applied Sciences. 2020;10(22), 8298.
21. Wang J, Zhu H, Wang SH, and Zhang YD. A Review of Deep Learning on Medical Image Analysis. Mobile Networks and Applications. 2021;26(1):351-80.
22. Mishra S, Prakash A, Roy SK, Sharan P, Mathur N. Breast Cancer Detection using Thermal Images and Deep learning. 7th International Conference on Computing for Sustainable Global Development (INDIACom). 2020; 211-6.
23. Allugunti V.R. Breast cancer detection based on thermographic images using machine learning and deep learning algorithms. International Journal of Engineering in Computer Science. 2022; 4(1): 49-56.
24. MohamedI EA, Rashed EA, GaberI T, Karam O. Deep learning model for fully automated breast cancer detection system from thermograms. Plos one. 2022; 17(1):e0262349.
25. Sánchez-Cauce R, Pérez-Martín J, Luque M. Multi-input convolutional neural network for breast cancer detection using thermal images and clinical data. Computer Methods and Programs in Biomedicine. 2021; 204:106045.
26. Muchamad MK, Arnia F, Syukri M, Munadi K. A Conceptual Framework of Deploying a Trained CNN Model for Mobile Breast Self-Screening. 18th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON). 2021; 533-7.
27. Zuluaga-Gomez J, Al Masry Z, Benaggoune K, Meraghni S, Zerhouni N. A CNN-based methodology for breast cancer diagnosis using thermal images. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization. 2021; 9(2):131-45.
28. Dey S, Roychoudhury R, Malakar S, Sarkar R. Screening of breast cancer from thermogram images by edge detection aided deep transfer learning model. Multimedia Tools and Applications. 2022;81(7):9331-49.
29. Gonçalves CB, Souza JR, Fernandes H. Fernandes. Classification of static infrared images using pre-trained CNN for breast cancer detection. 34th International Symposium on Computer-Based Medical Systems (CBMS). 2021; 101-6
30. Roslidar R, Saddami K, Arnia F, Syukri M, Munadi KA Study of Fine-Tuning CNN Models Based on Thermal Imaging for Breast Cancer Classification. IEEE International Conference on Cybernetics and Computational Intelligence (IEEE CYBERNETICSCOM). 2019; 77-81.
31. Tsietso D, Yahya A, and Samikannu R. A Review on Thermal Imaging-Based Breast Cancer Detection Using Deep Learning. Mobile Information System. 2022; 2022(1):8952849.
32. [online]: Available from: http://visual.ic.uf.br/.
33. Silva LF, Saade DC, Sequeiros GO, Silva AC, Paiva AC, Bravo RS, et al. A New Database for Breast Research with Infrared Image. Journal of Medical Imaging Heal. Informatics. 2014; 4(1):92−100.
34. Houssein E, et al. Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review. Expert Systems with Applications. 2021; 167:114161.
35. Tan M, Le Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks.

International Conference on Machine Learning. 2019; 97: 6105-14.
36. Sarvamangala DR, Kulkarni RV. Convolutional neural networks in medical image understanding: a survey. Evolutionary Intelligence. 2022;15(1):1-22.
37. Lotter W, Diab AR, Haslam B, Kim JG, Grisot G, Wu E, et al. Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach. Nat Med. 2021;27(2):244- 9.
38. Ensafi M, Keyvanpouy MR, Shojaedini SV. A New method for promote the performance of deep learning paradigm in diagnosing breast cancer: improving role of fusing multiple views of thermography images. Health and Technology. 2022; 12(6):1097-107.
39. Saber A, Sakr M, Abo-Seida OM, Keshk A, Chen H. A novel deep-learning model for automatic detection and classification of breast cancer using the transferlearning technique. IEEE Access. 2021; 9:71194–209
40. Kim HE, Cosa-Linan A, Santhanam N, Jannesari M, Maros ME, Ganslandt T. Transfer learning for medical image classification: a literature review. BMC Med Imaging. 2022; 22(1): 69.
41. Eldin SN, Hamdy JK, Adnan GT, Hossam M, Elmasry N, Mohammed A. Deep Learning Approach for Breast Cancer Diagnosis from Microscopy Biopsy Images. In Proceedings of the 2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC). 2021; 216-22.
42. Farooq M.A, and Corcoran P. Infrared Imaging for Human Thermography and Breast Tumor Classifcation using Thermal Images. 31st Irish Signals and Systems Conference (ISSC). 2020; 1-6.
43. Torres-Galván, JC, Guevara E, Kolosovas-Machuca ES, Oceguera-Villanueva A, Flores JL, González FJ. Deep convolutional neural networks for classifying breast cancer using infrared thermography. Quantitative InfraRed Thermography Journal. 2022:19(4):283-94.
44. Kiymet S, Aslankaya MY, Taskiran M, and Bolat B. Breast cancer detection from thermography based on deep neural networks. In 2019 Innovations in Intelligent Systems and Applications Conference (ASYU). IEEE, 2019:1-5.