



## Enhanced Lesion Classification Based on YOLO Architectures Using Thermal Breast Images on a Patient by Patient Basis

Kerim Kürşat Çevik<sup>1</sup>, Soner Çivilibal<sup>2</sup>, Ahmet Bozkurt<sup>3</sup>, Emre Dandil<sup>4\*</sup>

<sup>1</sup> Department of Management Information Systems, Faculty of Applied Sciences, Akdeniz University, Antalya 07070, Türkiye

<sup>2</sup> Department of Biomedical Engineering, Graduate School of Natural and Applied Sciences, Akdeniz University, Antalya 07070, Türkiye

<sup>3</sup> Department of Computational Science and Engineering, Informatics Institute, Istanbul Technical University, Istanbul 34467, Türkiye

<sup>4</sup> Department of Computer Engineering, Faculty of Engineering, Bilecik Seyh Edebali University, Bilecik 11230, Türkiye

Corresponding Author Email: [emre.dandil@bilecik.edu.tr](mailto:emre.dandil@bilecik.edu.tr)

Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.410617>

### ABSTRACT

Breast cancer classification using deep learning architectures plays a crucial role in assisting clinicians with early-stage diagnosis. In this study, we present a comprehensive evaluation of YOLO architectures-YOLOv2, YOLOv3, YOLOv4, and YOLOv5-for the classification of breast lesions in thermal breast images. By employing these architectures, we enhanced the identification of relevant regions of interest (ROIs) for lesion contouring. The dataset for this research was sourced from a publicly available repository, and divided on a patient-by-patient basis. This patient-based split enhances the robustness and clinical relevance of the model's performance compared to prior studies that relied on random data partitioning. Experimental results demonstrate that YOLOv5, trained with the Stochastic Gradient Descent with Momentum (SGDM) optimizer, achieved superior performance, with 0.83, 0.66, 0.97 and 0.79 for the key metrics of accuracy, precision, recall and F1-score, respectively. These results underscore the model's potential for reliable breast lesion classification and emphasize the importance of robust dataset partitioning to enhance clinical applicability.

**Received:** 11 May 2024

**Revised:** 7 October 2024

**Accepted:** 6 November 2024

**Available online:** 31 December 2024

### Keywords:

*breast lesions, thermal imaging, classification, deep learning, YOLO architecture*

## 1. INTRODUCTION

The most commonly diagnosed cancer in women worldwide and the second leading cause of cancer-related deaths after lung cancer, breast cancer is a major health concern [1, 2]. Breast cancer has an incidence rate of 11.7% and a mortality rate of 13.6% among cancer patients worldwide, according to the World Health Organization's (WHO) Global Cancer Observatory [3]. Early detection of breast cancer significantly enhances the efficacy of treatment, with the diagnostic process typically starting with routine self-examinations and advancing to imaging techniques followed by biopsies if required [4]. While biopsy is the gold standard in diagnosis, it presents risks such as bleeding, infection, and bruising, making imaging modalities a critical step before invasive procedures.

Imaging techniques are used before biopsy, with the technique chosen depending on the size of the suspected cancer and the size of the breast. Due to its ability to detect cancers in smaller breast volumes and its non-invasive nature, mammography is considered the gold standard for breast cancer screening [5]. However, mammography has limitations, including patient discomfort due to compression, difficulty in imaging large breasts, and the use of ionizing radiation (X-rays). In addition, mammography is often supplemented by

ultrasound and magnetic resonance imaging (MRI). Ultrasound is useful for imaging larger breasts, but has limited penetration for deep lesions. MRI, although more sensitive, is less preferred due to its high cost and longer examination times [4].

Thermal imaging, or thermography, presents a promising non-invasive alternative for breast cancer detection, offering the advantage of no exposure to ionizing radiation [6]. Thermography captures the thermal emissions of the body, producing temperature-based imaging of breast tissue. It can be divided into active and passive as two types [7]. In passive thermography, the natural heat emitted by the breast is captured without any external stimulation. Active thermography, on the other hand, introduces external stimuli (such as cold or hot) to elicit temperature changes in the region of interest, and the resulting thermal variations are analyzed over time to distinguish between healthy and abnormal tissue [8]. In this way, images obtained at different times can be analyzed and computerized systems can be used to make the most appropriate distinction between lesioned and healthy areas.

The rapid evolution of artificial intelligence (AI), particularly with the advancement of hardware and computational capabilities, has led to widespread application in medical image recognition and classification [9]. AI

approaches to image analysis can be categorized into machine and deep learning. Traditional machine learning involves feature extraction followed by the application of algorithms to classify the images. However, this manual process can result in critical image information being overlooked, which can limit the model's accuracy. Deep learning, on the other hand, automates the feature extraction process, allowing for more robust analysis. Recent studies have consistently shown that deep learning models, particularly convolutional neural networks (CNN), exceed traditional machine learning methods in image-based tasks [10-12].

Deep learning architectures stand out as a highly effective technology for detecting breast lesions from thermogram images. In thermograms, cancer tissue typically shows higher temperatures than healthy tissue [13]. By analyzing the temperature distributions contained in thermal images, these architectures can accurately and quickly detect anomalies such as cancerous lesions. Unlike other breast imaging techniques, thermography is non-invasive and does not use radiation [14]. This is a significant advantage, especially for frequent scans. Thermography integrated with deep learning architectures has the potential to further increase clinical use by providing more reliable and faster lesion detection results.

Breast cancer diagnosis is an area where early detection significantly improves clinical outcomes. Among various imaging techniques, thermal imaging (thermography) has gained attention due to its non-invasive nature and ability to detect abnormal heat patterns associated with malignancies. This study investigates the use of advanced deep learning techniques, specifically YOLO architectures, to detect and classify breast regions on thermograms. A publicly available repository, the DMR-IR database [15], used in this study. In order to investigate the impact of different YOLO architectures on breast region classification, we experimented with different versions such as YOLOv2, YOLOv3, YOLOv4 and YOLOv5. Each version was evaluated in terms of its ability to classify both breasts in each image, identifying which region was relevant for contouring and lesion detection. The choice of YOLO architectures also allowed the integration of different backbones, facilitating transfer learning and optimization of the models for better performance on the thermographic dataset. A summary of the study's contributions can be found in the following:

- This study evaluates the impact of different YOLO architectures on the classification of breast regions.
- YOLO's single-shot detection system allows it to localize lesions with high precision in thermal images.
- This study enhances the state-of-the-art in breast cancer diagnosis using thermography by addressing key limitations of existing methods.
- Specifically, it ensures that both breasts are correctly classified before any further lesion analysis, significantly improving diagnostic accuracy.

## 2. RELATED WORK

The detection of breast cancer and other lesions by means of thermal imaging has gained increasing attention due to its non-invasive nature and the absence of ionizing radiation. Numerous studies have focused on improving breast cancer diagnosis using machine learning [16, 17] and deep learning

[18] techniques. Over the past decade, deep learning techniques have advanced significantly and offer important advantages over machine learning algorithms in accurately diagnosing abnormal breast tissue [19, 20].

Machine learning techniques on thermal images have many applications in breast cancer diagnosis and classification. Schaefer et al. [21] proposed a fuzzy classification method using statistical features such as mean, standard deviation, median for temperature to diagnose breast cancer. Acharya et al. [22] used support vector machines (SVM) to identify breast cancer with high accuracy. In another study, Mookiah et al. [23] classified tissue features obtained from thermograms for the diagnosis of breast cancer using discrete wavelet transform (DWT). Their results showed that fuzzy sugeno and decision tree algorithms gave the best results. Milosevic et al. [24] evaluated breast cancer classification using machine learning algorithms such as SVM, k-nearest neighbors (k-NN). Bayes and k-NN had the greatest classification performance. In another study, Pramanik et al. [25] used a feed-forward multilayer perceptron network (MLP) algorithm to classify breast cancer. Karim et al. [7] applied to classify breast images using the SVM method with gray level co-occurrence matrix (GLCM) features and obtained high classification results.

On the other hand, many studies have been proposed for the segmentation of breast lesions. Golestani et al. [26] investigated the segmentation of breast lesions using k-means, fuzzy c-means, and a level determination method and qualitatively characterized the achieved segmentation performances. Ghayoumi Zadeh et al. [27] studied the automatic segmentation processes on thermograms using semi-automatic and deep learning networks.

Some previous studies have used DL architectures in addition to studies dealing with the detection and classification of thermal breast images. Baffa and Lattari [8] used a CNN model for breast cancer detection and achieved high classification accuracy in the static and the dynamic datasets. In another study, Fernández-Ovies et al. [28] studied breast cancer diagnosis using thermograms with CNN and they reported that ResNet34 and ResNet50 architectures achieved the best results using transfer learning. Tello-Mijares et al. [29] used machine learning algorithms such as decision trees, artificial neural networks (ANN), naïve bayes and CNN architectures for breast cancer diagnosis. Farooq and Corcoran [30] reported that they achieved high performance for breast cancer classification using transfer learning with the InceptionV3 architecture after preprocessing on thermograms. In their study, Zuluaga-Gomez et al. [11] investigated breast cancer detection using transfer learning architectures such as ResNet, InceptionV3, SeResNet, Xception VGG16, and InceptionResNetV2 in addition to the CNN architecture based on segmented thermograms. Civilibal et al. [31] employed Mask R-CNN technique for automatic detection, segmentation, and classification of breast lesions on thermograms and achieved high scores using Res-Net50 architecture.

Although the previous studies provide higher accuracy values for breast lesion detection and classification, their convergence speed is relatively slower. Recent studies have focused on developing faster architectures for medical image recognition problems. Al-Antari et al. [32] investigated computer-aided identification of breast lesions using mammography images. As a result of their study, they obtained the higher results using InceptionResNet-V2 based YOLO architecture on the in a breast dataset [33] and DDSM dataset [34]. In their study, Baccouche et al. [35] proposed the

simultaneous diagnosis of breast lesions from mammograms using YOLO-based fusion models. Aly et al. [36] studied the identification and classification of breast cancer from mammography images using YOLO architectures. They achieved the highest performance of the three designs using InceptionV3. Hamed et al. [37] proposed YOLOv4-based breast lesion identification and classification on mammography images and reported that their proposed model achieved higher classification results on test datasets. Kolchev et al. [38] studied breast lesion detection using YOLOv4 architecture on mammography images and performed classification.

Despite these advancements in mammography and other imaging modalities, there has been limited exploration of YOLO architectures for breast lesion detection in thermal images. Existing studies on thermographic imaging have focused on segmentation and classification, but often lack specificity in identifying which breast contains the lesion. For example, even if breast cancer is diagnosed using whole-image categorization, the results may not specify which breast is malignant. Similarly, segmentation-based classification of the hottest region can be misinterpreted as a lesion due to temperature differences between different slices. Moreover, traditional segmentation approaches can misclassify hot regions as lesions due to temperature variations. The YOLO architecture, known for its speed and accuracy, addresses these

limitations by detecting and classifying breast lesions in a single pass, reducing the risk of misinterpretation and improving clinical applicability. This study fills that gap by investigating YOLO architectures for lesion detection and classification in thermal images. By training the models on a patient-by-patient basis, this work aims to increase the reliability of the results and to avoid biases related to random image splitting as observed in previous studies.

### 3. MATERIAL AND METHODS

In this study, YOLO architectures are used to detect and classify breast lesions on thermogram images. The lesion detection process is followed that includes the steps of dataset creation, pre-processing, image annotation, detection and classification of the images with YOLO architectures. The training and test sets were processed separately after the resulting images were divided in different sets. In order to clearly illustrate this process, a block diagram in Figure 1 is presented that explains the lesion detection and classification processes on thermograms step by step. This diagram covers the training, testing and performance evaluation of the model. In methodology, the YOLO architectures were trained using the thermograms in training set and finally, the classification were applied to the test set.

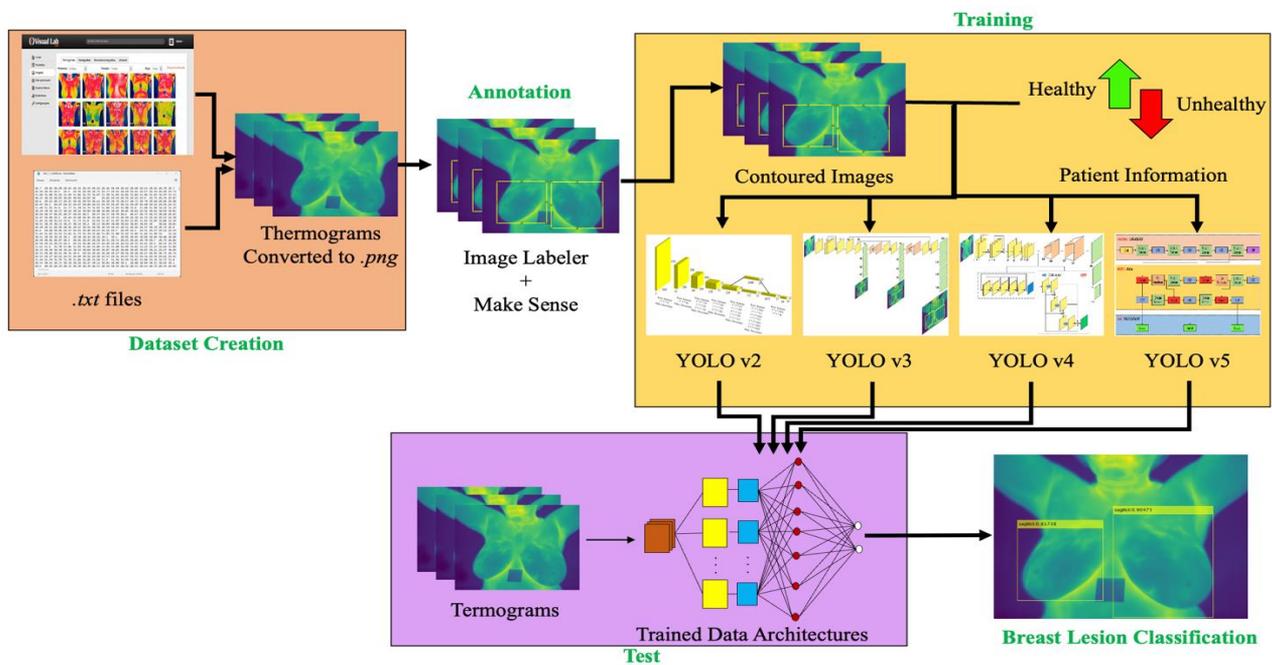


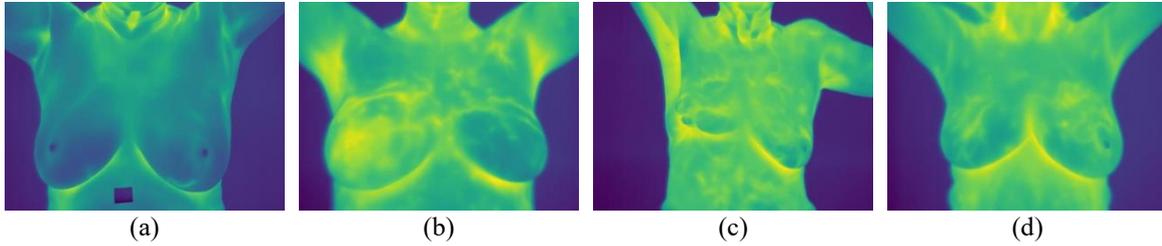
Figure 1. Flowchart diagram of the proposed detection and classification model for breast lesion on thermograms

#### 3.1 Dataset

In this study, the publicly available Database for Mammal Research with Infrared Image (DMR-IR) database [15] was used to classify breast lesions from thermal images. The infrared images in this dataset were captured by a FLIR thermal imaging camera and are 640×480 pixels in size. In addition, the images in the dataset were collected from a total of 149 patients, with an average of 27 images per patient. In this study, a total of 1120 images from 19 healthy and 37 patients were obtained from the DMR-IR dataset to classify breast lesions from thermal images using the proposed method. Thermograms of healthy individual (a), patient with lesion (b),

patient with mastectomy (c), and patient with asymmetrical breast (d) sample from DMR-IR dataset are shown in Figure 2.

In preparing the dataset, the images from the '.txt' format dataset were first converted to '.png' format using a Python code. These images were then parsed and distributed to the training and test sets. The contouring and annotation processes for the images were saved with '.txt' and '.mat' extensions using Image Labeler [39] and Make Sense [40] software. Here, two different actual class labels, healthy and unhealthy, are created for each image. Finally, YOLO architectures were trained using thermograms and localization information after preprocessing for training.



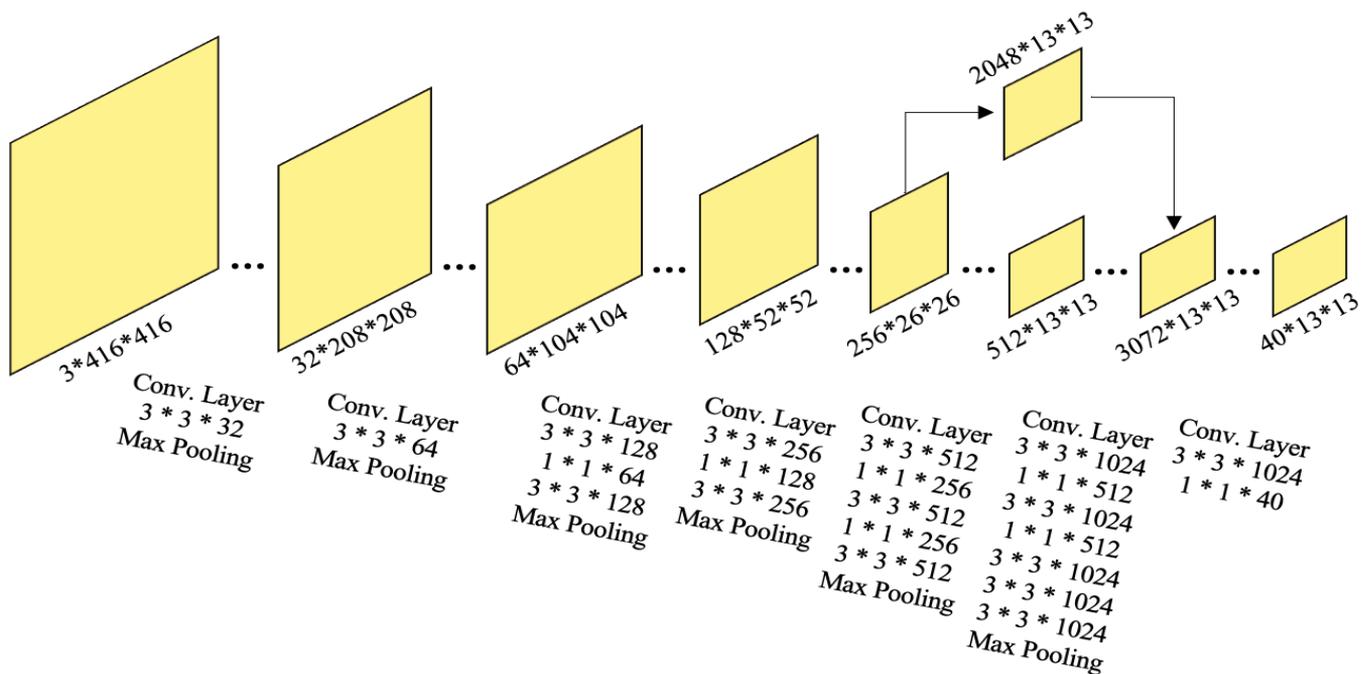
**Figure 2.** Sample thermograms (a) healthy individual, (b) patient with lesion, (c) patient with mastectomy, and (d) patient with asymmetrical breast from DMR-IR database [15]

### 3.2 YOLO architectures

In this study, we compared YOLOv5 with YOLOv2, YOLOv3 and YOLOv4 to evaluate the performance improvements of different versions of the YOLO architecture in the context of breast lesion detection and classification using thermal images. By comparing these different versions, we aimed to determine which YOLO architecture provides the best balance between detection accuracy, precision, sensitivity, and computational efficiency for thermal breast images. This comparative analysis provides insights into the evolution of YOLO architectures and their suitability for medical imaging tasks.

The You Only Look Once (YOLO) architecture is based on a CNN and serves as a deep learning technique for tracking and detection objects in an image [41, 42]. Because it treats an object as a regression problem at once and only runs the image through the neural network once, it is faster than the other state-of-the-arts techniques. It uses CNNs as the basis for feature extraction and categorization. The architecture starts operations after resizing the input photographs. Double linear interpolation is used to adapt the resizing procedures by scaling the input image first to the intermediate image and then to the final image. Instead of working on the whole image, the YOLO architecture divides the image into equal parts. It creates box drawings, known as bounding boxes, that show the elements in each zone it divides. Confidence scores are then

calculated, which indicate the percentage similarity of objects in boxes that can be drawn in different sizes [43]. In determining whether an object is present in the drawn boxes, it also determines whether the box is within the object's central point if an object is present there. It also predicts a vector containing the object's length, height and class details according to the controls. The YOLO architecture is inadequate when there is more than one object center in a drawn box. To address this issue, researchers created the YOLOv2 architecture, which incorporates the anchor box approach and eliminates the interconnected layers [44]. As a result, the boxes are constructed independently for each object, and a confidence score, Intersection over Union (IoU), class likelihood, and size information are computed accordingly. Although it can be possible to detect more than one object in the box in this situation, there would be numerous superfluous box drawings. The architecture uses the non-max suppression algorithm to overcome this problem. This technique reduces the number of boxes containing the objects to one by eliminating the boxes with a confidence value less than a certain threshold, the boxes with the highest confidence value and the boxes with a value greater than  $\text{IoU}=0.5$ . The architectural output is then created by drawing the bounding box and class information on the source image, which is finally condensed into a single box. The structure of the YOLOv2 architecture is shown in Figure 3.



**Figure 3.** The structure of the YOLOv2 architecture adopted from previous study [45]

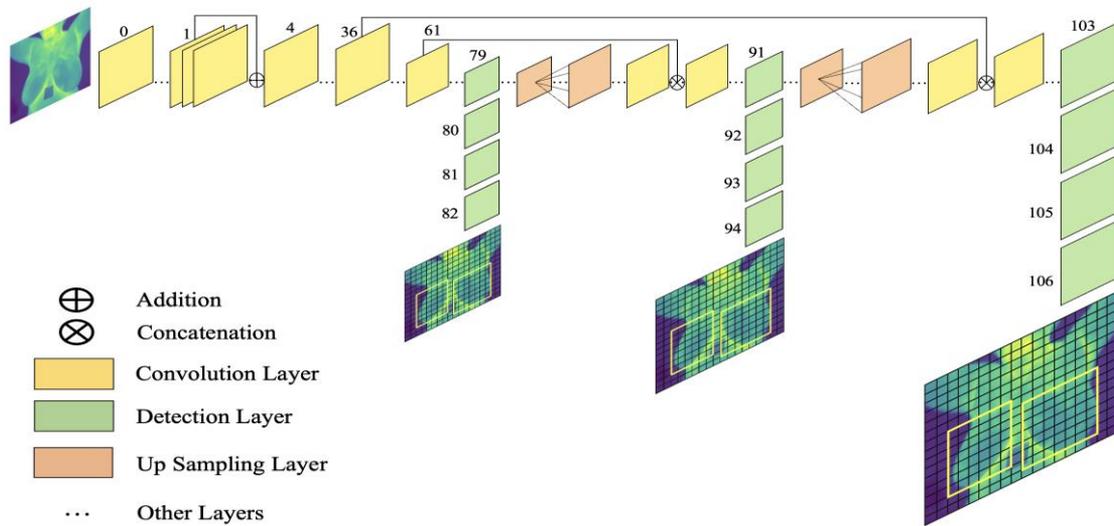


Figure 4. YOLOv3 architecture adopted from previous studies [46, 47]

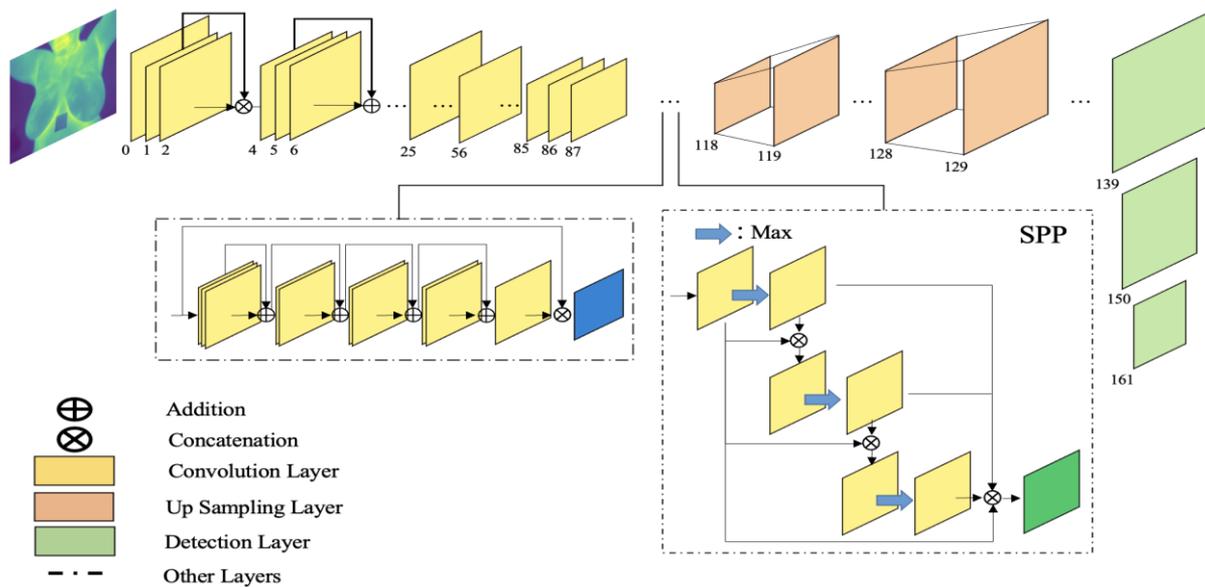


Figure 5. The network structure of the YOLOv4 architecture adapted from previous study [47]

The DarkNet-19 network architecture with 19 convolutional layers and 5 pooling layers is used in the YOLOv2 architecture [48]. Instead of using fully connected layers like its predecessor, YOLOv2 incorporates bulk normalization and uses connection boxes [49]. Although these changes increase the speed and accuracy of the model, they are insufficient for recognizing small objects. The YOLOv3 architecture avoids these shortcomings. Through the use of novel techniques and the Featured Pyramid Networks (FPN) it creates, YOLOv3 makes it possible to recognize objects of different dimensions. The YOLOv3 architecture uses the DarkNet53 feature extractor and a 53-layer convolutional neural network [49]. It also uses three estimation scales at each location to extract feature maps on the input images, as shown in Figure 4. On the other hand, the YOLOv4 model was created by tweaking and increasing the number of mesh structures in the YOLOv3 model, selecting the best hyperparameter and performing normalization operations on small groups. As a result, useful statistics could be collected over many training iterations using collective normalization techniques. Furthermore, in YOLOv4, CSPDarkNet53 is used instead of DarkNet53 in the backbone structure. The CSPDarkNet53 backbone consists of 29

convolutional layers, 725.725 receiver regions, and 27.600 parameters [50]. The network structure of the YOLOv4 architecture is denoted in Figure 5. In addition, in YOLOv4, Spatial Pyramid Pooling (SPP) helps improve object detection performance, especially for objects of varying scales.

When it was discovered that the YOLOv4 model had too many layers and was too slow, the YOLOv5 model was created, increasing the speed [51]. Figure 6 shows the network structure of the YOLOv5 architecture. The backbone is for feature extraction, the neck is for feature map generation, and the head is for detection and classification. The YOLOv5 architecture uses CSPDarkNet53 as the backbone, which provides better detection accuracy than DarkNet53 in YOLOv3 [52]. The Path Aggregation Network (PANet) is used for the parametric polymerization mechanism in the neck section. In addition, In YOLOv5, Cross Stage Partial (CSP) is a block used to improve the efficiency and performance of the backbone and neck sections of the model and uses as BottleNeckCSP in YOLOv5. The final section, the head section, uses the same categorization and detection mechanisms as YOLOv3 and YOLOv4.

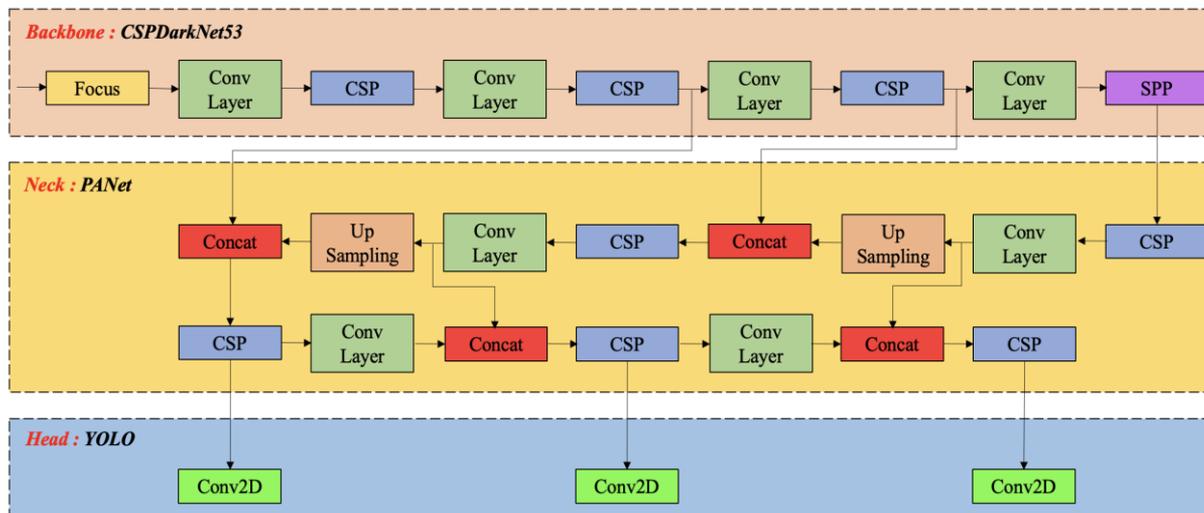


Figure 6. The network structure of the YOLOv5 architecture adapted from previous study [53]

#### 4. RESULTS AND DISCUSSION

In this study, we performed comprehensive experiments using various versions of YOLO architectures to classify breast lesions from thermal images. A total of 1120 thermal images from the DMR-IR dataset were divided into 80% training and 20% test sets on a patient-by-patient basis. A total of 44 patients (15 healthy and 29 unhealthy) were selected for the training set and 12 patients (4 healthy and 8 unhealthy) were selected for the test set. As 20 images for the right and left breast were obtained from the DMR-IR database for each patient, there are a total of 880 thermal images in the training set and 240 thermal images in the test set. Furthermore, considering two breast images per patient, the training set contains 1180 healthy and 580 unhealthy breast images, while the test set contains 320 healthy and 160 unhealthy breast images. Thus, in the patient-by-patient approach, 20 images from each patient are either in the training set or in the test set. In other words, each patient does not have images in both the training and test sets.

In the proposed study for classification of breast lesions from thermal images, images obtained from DMR-IR dataset are labelled with ground truth (actual class) labels for two different classes as healthy and unhealthy using labelling software. In the proposed approach, after the training of YOLO architectures, these labels are determined for thermal images. Firstly, breast detection with bounding box is performed using YOLO architectures. In the next step, the classification process is completed by predicting the labels for the corresponding breast.

The experimental results in this study were achieved using the UHeM Altay [54] server with applications written in Python programming language. In the experimental studies, a total of 16 tests, denoted in Table 1, were performed according to different values of some hyperparameters to evaluate the performance of YOLO architectures. The tests were performed on different architectures, specifically YOLOv2, YOLOv3, YOLOv4 and YOLOv5, with different backbones such as SqueezeNet, Tiny YOLO, DarkNet and YOLOv5x. On the other hand, Adam and Stochastic Gradient Descent with Momentum (SGDM) optimizers have been used in YOLO architectures. While SGDM is reliable and generalizes well but may require more tuning of the learning rate, converges

faster and is easier to fine-tune, it may not generalize. On the other hand, the experimental analyses evaluated the effect of three different values of batch size, 4, 16 and 64, on the results. The batch size is a crucial hyperparameter that affects both the training speed and the ability of the model to generalize. Finally, the effect of the number of epochs for 10, 20, 50, 100, 200 and 500 is also observed. In all these performance tests, the learning rate for each of the YOLO architectures was set to 0.001.

Table 1. Hyperparameters optimization in training YOLOv2, YOLOv3, YOLOv4 and YOLOv5 architectures for breast lesion classification on thermal images

Test Number	YOLO Architecture	Backbone	Optimizer	Batch Size	Epoch
1	YOLOv2	SqueezeNet	SGDM	64	100
2	YOLOv3	SqueezeNet	SGDM	64	100
3	YOLOv4	Tiny YOLO	Adam	16	50
4	YOLOv4	Tiny YOLO	Adam	16	100
5	YOLOv4	Tiny YOLO	Adam	16	100
6	YOLOv4	Tiny YOLO	SGDM	16	100
7	YOLOv4	Tiny YOLO	Adam	16	200
8	YOLOv4	Tiny YOLO	SGDM	16	200
9	YOLOv4	Tiny YOLO	Adam	16	500
10	YOLOv4	DarkNet	Adam	4	10
11	YOLOv4	DarkNet	Adam	4	20
12	YOLOv4	DarkNet	Adam	4	30
13	YOLOv4	DarkNet	Adam	4	50
14	YOLOv5	YOLOv5x	SGDM	4	100
15	YOLOv5	YOLOv5x	SGDM	16	100
16	YOLOv5	YOLOv5x	SGDM	16	200

Table 2. The confusion matrix design for breast lesion classification on thermal images

		Predicted Class	
		Unhealthy	Healthy
Actual Class	Unhealthy	<b>TP</b> True Positive	<b>FN</b> False Negative
	Healthy	<b>FP</b> False Positive	<b>TN</b> True Negative

In this study, which is proposed to classify breast lesions as healthy and unhealthy from thermal images, the classification

performances of YOLO architectures are evaluated using the key metrics of Accuracy, Precision, Recall and F1-score obtained from the confusion matrix. These metrics are given in Eqs. (1)-(4), respectively. The confusion matrix from which these equations are generated is shown in Table 2. Here, true positive (TP) predicts lesioned breasts as unhealthy and true negative (TN) predicts non-lesioned breasts as healthy. On the other hand, false negative (FN) indicates breast images with lesions but classified as healthy, while false positive (FP) is used for breast images without lesions but classified as unhealthy.

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

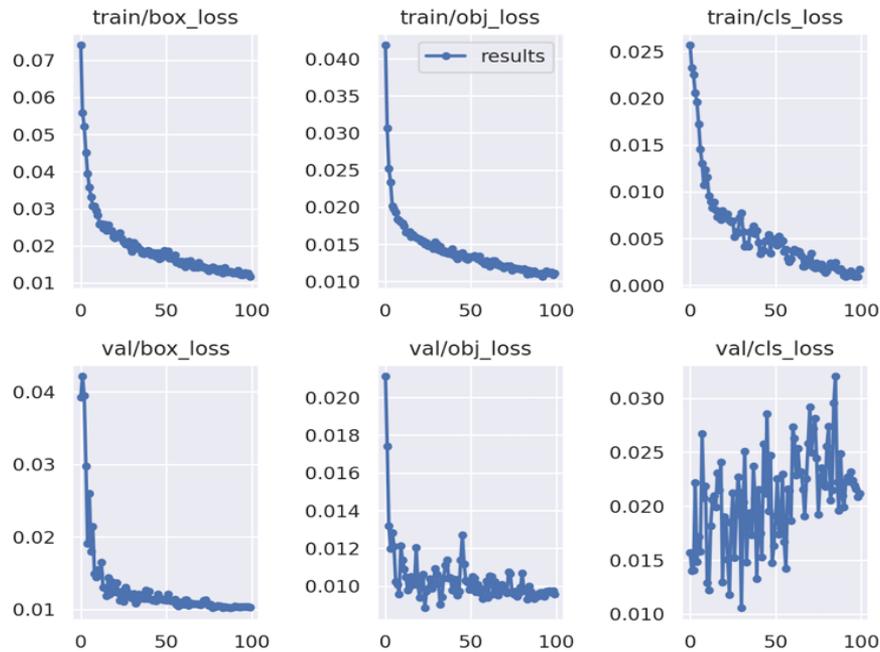
$$Precision = \frac{TP}{TP + FP} \quad (2)$$

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

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

**Table 3.** The confusion matrix results and breast lesion classification scores by key metrics for performance tests

Test Number	Architecture	TP	FP	TN	FN	Empty	Total	Precision	Recall	F1-Score	Accuracy
1	YOLOv2	125	75	245	35	0	480	0.63	0.78	0.70	0.77
2	YOLOv3	124	80	218	29	29	480	0.61	0.81	0.70	0.76
3	YOLOv4	76	78	220	43	63	480	0.49	0.64	0.58	0.71
4	YOLOv4	76	86	221	43	54	480	0.47	0.64	0.54	0.70
5	YOLOv4	73	65	210	35	97	480	0.53	0.68	0.59	0.74
6	YOLOv4	63	75	190	52	100	480	0.46	0.55	0.50	0.67
7	YOLOv4	79	87	222	61	31	480	0.48	0.56	0.52	0.67
8	YOLOv4	58	74	222	59	67	480	0.44	0.50	0.47	0.68
9	YOLOv4	72	71	231	47	59	480	0.50	0.61	0.55	0.72
10	YOLOv4	66	120	191	69	34	480	0.36	0.49	0.41	0.58
11	YOLOv4	45	55	248	58	74	480	0.45	0.44	0.44	0.72
12	YOLOv4	76	49	220	52	83	480	0.61	0.59	0.60	0.75
13	YOLOv4	43	51	180	33	173	480	0.46	0.57	0.51	0.73
14	YOLOv5	116	101	219	29	15	480	0.54	0.80	0.64	0.72
15	YOLOv5	141	72	230	5	32	480	0.66	0.97	0.79	0.83
16	YOLOv5	132	82	238	28	0	480	0.62	0.83	0.71	0.77

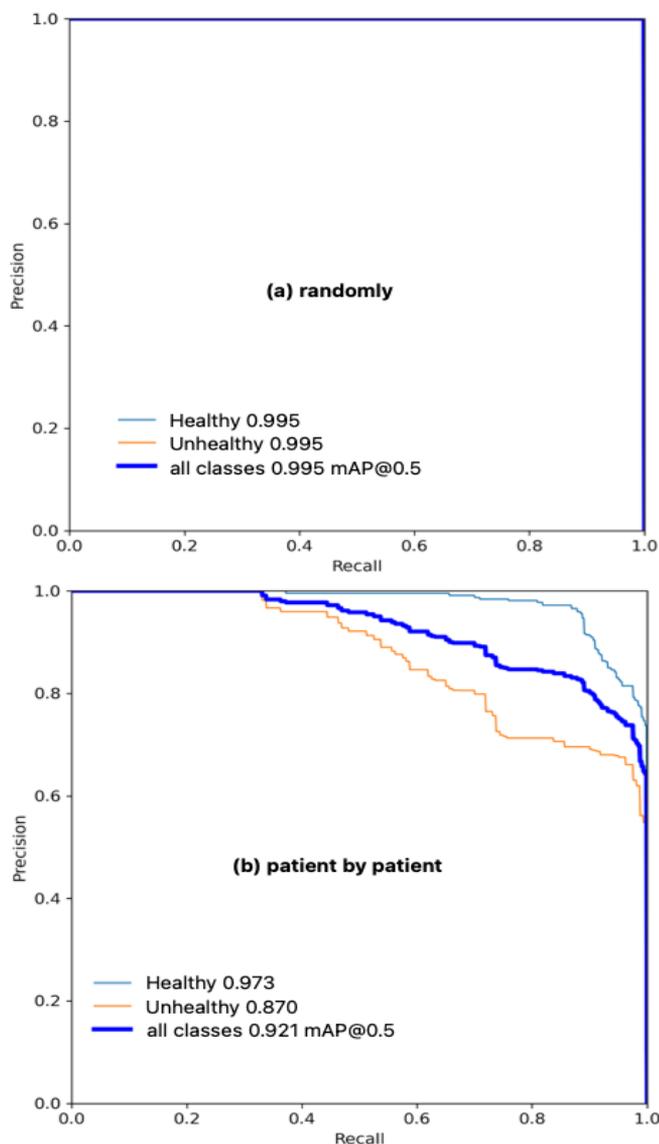


**Figure 7.** Training and validation loss curves for the training set of the YOLOv5 architecture

Table 3 shows the results obtained for 16 different performance tests conducted on the test set using YOLO architectures (YOLOv2, YOLOv3, YOLOv4 and YOLOv5) at different hyperparameter values for patient-by-patient detection and classification of breast lesions from thermal images. To evaluate the classification performance, 240 breast thermal images from 12 patients in the test set, 20 images from each patient, were used in the patient-by-patient approach. Since each breast in each patient has a separate class label as

healthy or unhealthy, the results were obtained over a total of 480 images for the test set, taking into account two breasts in each patient. Here, both the confusion matrix values obtained for each of the YOLO architectures and the Accuracy, Precision, Recall and F1-score key metric results are presented. On the other hand, in the event that breast detection could not be performed in thermal images with YOLO architectures, the relevant images were evaluated as Empty. In the patient-by-patient approach proposed in this study, 20 thermal images

(480 breast images for 12 patients) of the same patient are included in the test set. Evaluating the results in Table 3, it can be seen that the most successful classification results are obtained with YOLOv5 in Test 15. For this test, results of 0.83, 0.66, 0.97 and 0.79 were achieved for Accuracy, Precision, Recall and F1-score key metrics, respectively. The highest value was achieved for TP and the lowest for FN. On the other hand, 32 breast images were categorized as Empty, as no breast detection could be performed with YOLOv5. On the other hand, the highest accuracy score of 0.75 was achieved in the classification of breast lesions from thermal images using YOLOv4, while 0.77 and 0.76 accuracy scores were obtained for the YOLOv2 and YOLOv3 architectures, respectively. When the results of all architectures are compared, it is seen that either Empty responses or high FP values are common issues. On the other hand, this common issue shows that training with a low-dimensional dataset may lead to an inability to distribute the results proportionally or to choose the correct backbone.



**Figure 8.** Areas under Precision-Recall curves for randomly (a), and patient by patient (b) approaches in the training set for YOLOv5 architecture

Since the highest accuracy in breast lesion classification from thermal images was achieved with YOLOv5, Figure 7 shows the training and validation loss curves for the training and validation sets. Here, box\_loss, obj\_loss and cls\_loss graphs are shown for the training and validation sets. box\_loss shows the accuracy of the coordinates of the ground truth bounding box and the predicted bounding box. obj\_loss measures how confident the model is that an object exists within a predicted bounding box. On the other hand, cls\_loss evaluates the accuracy of the predicted class label for each object within a bounding box. From these graphs it can be seen that the training of YOLOv5 is successful and generalizes the results. It can also be seen that the cls\_loss graph for the validation set does not follow a regular decreasing trend and has a variable structure.

In order to evaluate the accuracy of the patient-by-patient approach proposed in this study for the detection and classification of breast lesions from thermal images, 1120 patient images in the prepared dataset were randomly divided into 80% (896 images) training set and 20% (224 images) test set. Thus, a total of 1792 breast images were randomly generated for evaluation in the training set for both breasts. Furthermore, for randomly approach, considering two breast images per patient, the training set contains 1182 healthy and 610 unhealthy breast images, while the test set contains 298 healthy and 150 unhealthy breast images. As a result of the training performed using the YOLOv5 architecture, Figure 8 (a) shows the area under the Precision-Recall curves obtained with the randomly approach for the healthy and unhealthy classes, and Figure 8 (b) shows the area under the Precision-Recall curves obtained with the patient-by-patient approach for the healthy and unhealthy classes. Although the area under the curves is larger for the random approach and YOLOv5 is considered to be more successful in classification, it is assumed that the network in this approach memorizes the images of the same patient from both training and test sets. Another important point supporting this is that the thermal images in the dataset were acquired at very close time intervals, so the images are very similar to each other.

Breast cancer diagnosis using thermograms has been investigated in many machine learning and deep learning studies in the previous [7, 25, 28]. In addition, these studies typically use techniques such as direct classification or automatic segmentation of breast regions, segmentation of the hottest areas of the thermogram, classification using segmented breast images, and methods that combine segmentation and classification. In Table 4, previous studies proposed using thermal breast images from the DMR-IR database are categorized as machine learning and deep learning approaches, and performance comparisons are provided. In addition, these studies are categorized into detection, segmentation and classification. It is observed that deep learning-based methods are more successful than machine learning based methods in the previously proposed studies for breast lesion detection from thermal images. On the other hand, it is observed that the most of the previously proposed studies are based on a randomized approach and the studies based on a patient-by-patient approach, as in this study, are limited. In this study, the training and test sets were separated on patient basis, and the YOLOv5 model achieved 83% accuracy for classification performance.

**Table 4.** The performance comparisons for previous studies proposed using thermal breast images from the DMR-IR dataset (NA: not applicable)

Method	Study	Approach	Model	Data Split Technique	Accuracy (%)
Machine Learning	Pramanik et al. [25]	Detection & Classification	MLP	Randomly	90.48
	Karim et al. [7]	Classification	SVM	NA	91.25
	Ghayoumi Zadeh et al. [27]	Classification	Autoencoder	NA	94.87
	Baffa and Lattari [8]	Classification	CNN	Randomly	98.0 (static) 95.0 (dynamic)
Deep Learning	Fernández-Ovies et al. [28]	Classification	CNN	Randomly	100
	Tello-Mijares et al. [29]	Detection & Classification	CNN	NA	100
	Civilibal et al. [55]	Segmentation & Classification	Mask R-CNN and transfer learning	Randomly	100
	Zuluaga-Gomez et al. [11]	Classification	CNN	Randomly	92.0
	Civilibal et al. [31]	Detection & Classification	Mask R-CNN	Randomly	97.1
	Our study	Detection & Classification	YOLOv5	On a patient-by-patient basis	83.0

## 5. CONCLUSIONS

In this study, YOLOv5 architecture based on patient-by-patient basis was proposed for the detection and classification of breast lesions on thermal breast images. In addition, the performance results of YOLO architectures such as YOLOv2, YOLOv3, YOLOv4, and YOLOv5 were compared for breast lesion classification. In the study, breast lesion diagnosis was achieved by categorizing full thermograms, segmenting the hottest parts, manually or automatically segmenting the breast regions and then training on the segmented breasts. As a result, using YOLO-based deep learning architectures, this study adopted a patient-based strategy, investigating the detection of breasts one by one and then applying a specific categorization procedure for each breast. For the best performance result, we achieved for YOLOv5 trained with the SGDM optimization algorithm, 100 epochs and 16 batch size hyperparameters, in yielded scores of 0.83, 0.66, 0.97 and 0.79 for the key metrics of accuracy, precision, recall and F1-score, respectively. These findings underline the model's ability to detect and classify lesions with high sensitivity, which is critical in clinical practice. The decision to divide the dataset on a patient-by-patient basis is indeed a strong approach, as it enhances the study's reliability by preventing data leakage-where similar or nearly identical images from the same patient could end up in both the training and test sets. This method ensures that the model does not unfairly benefit from memorizing features specific to individual patients, but rather learns to generalize across different individuals.

Our study also contributes to the growing body of research by providing a detailed comparison of different YOLO models and showing that YOLO architectures, due to their speed and one-pass design, are highly suitable for real-time medical applications. Nonetheless, we acknowledge the need for additional work to further refine the models, improve generalizability, and reduce any biases introduced by the dataset or methodology. Future work may include larger datasets, a more detailed comparison with other state-of-the-art deep learning models, and exploring different methods for improving precision.

## ACKNOWLEDGMENT

The authors would like to thank the researchers of the publicly available dataset for providing the thermogram data. In addition, computing resources used in this work were

provided by the National Center for High Performance Computing of Türkiye (UHeM) (Grant No.: 016482023).

## REFERENCES

- [1] Harbeck, N., Penault-Llorca, F., Cortes, J., Gnant, M., Houssami, N., Poortmans, P., Ruddy, K., Tsang, J., Cardoso, F. (2019). Breast cancer. *Nature Reviews Disease Primers*, 5(1): 66. <https://doi.org/10.1038/s41572-019-0111-2>
- [2] Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 74(3): 229-263. <https://doi.org/10.3322/caac.21834>
- [3] Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 71(3): 209-249. <https://doi.org/10.3322/caac.21660>
- [4] Iranmakani, S., Mortezaazadeh, T., Sajadian, F., Ghaziani, M.F., Ghafari, A., Khezerloo, D., Musa, A.E. (2020). A review of various modalities in breast imaging: technical aspects and clinical outcomes. *Egyptian Journal of Radiology and Nuclear Medicine*, 51: 1-22. <https://doi.org/10.1186/s43055-020-00175-5>
- [5] Greenwood, H.I., Dodelzon, K., Katzen, J.T. (2018). Impact of advancing technology on diagnosis and treatment of breast cancer. *Surgical Clinics*, 98(4): 703-724. <https://doi.org/10.1016/j.suc.2018.03.006>
- [6] Magalhaes, C., Vardasca, R., Rebelo, M., Valenca-Filipe, R., Ribeiro, M., Mendes, J. (2019). Distinguishing melanocytic nevi from melanomas using static and dynamic infrared thermal imaging. *Journal of the European Academy of Dermatology and Venereology*, 33(9): 1700-1705. <https://doi.org/10.1111/jdv.15611>
- [7] Karim, C.N., Mohamed, O., Ryad, T. (2018). A new approach for breast abnormality detection based on thermography. *Medical Technologies Journal*, 2(3): 245-254. <http://doi.org/10.26415/2572-004X-vol2iss1p257-265>
- [8] Baffa, M.D.F.O., Lattari, L.G. (2018). Convolutional neural networks for static and dynamic breast infrared imaging classification. In 2018 31st SIBGRABI

- Conference on Graphics, Patterns and Images (SIBGRAP), Parana, Brazil, pp. 174-181. <http://doi.org/10.1109/SIBGRAP.2018.00029>
- [9] Pinto-Coelho, L. (2023). How artificial intelligence is shaping medical imaging technology: A survey of innovations and applications. *Bioengineering (Basel)*, 10(12): 1435. <https://doi.org/10.3390/bioengineering10121435>
- [10] Roslidar, R., Rahman, A., Muharar, R., Syahputra, M.R., Arnia, F., Syukri, M., Pradhan, B., Munadi, K. (2020). A review on recent progress in thermal imaging and deep learning approaches for breast cancer detection. *IEEE Access*, 8: 116176-116194. <https://doi.org/10.1109/ACCESS.2020.3004056>
- [11] Zuluaga-Gomez, J., Al Masry, Z., Benaggoune, K., Meraghni, S., Zerhouni, N. (2021). A CNN-based methodology for breast cancer diagnosis using thermal images. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 9(2): 131-145. <https://doi.org/10.1080/21681163.2020.1824685>
- [12] Shoieb, D.A., Youssef, S.M., Aly, W.M. (2016). Computer-aided model for skin diagnosis using deep learning. *Journal of Image and Graphics*, 4(2): 122-129. <http://doi.org/10.18178/joig.4.2.122-129>
- [13] Al Husaini, M.A.S., Habaebi, M.H., Suliman, F.M., Islam, M.R., Elsheikh, E.A., Muhaisen, N.A. (2023). Influence of tissue thermophysical characteristics and situ-cooling on the detection of breast cancer. *Applied Sciences*, 13(15): 8752. <https://doi.org/10.3390/app13158752>
- [14] Ucuzal, H., Baykara, M., Küçükakçalı, Z. (2021). Breast cancer diagnosis based on thermography images using pre-trained networks. *The Journal of Cognitive Systems*, 6(2): 64-68. <https://doi.org/10.52876/jcs.990948>
- [15] Silva, L.F., Saade, D.C.M., Sequeiros, G.O., Silva, A.C., Paiva, A.C., Bravo, R.S., Conci, A. (2014). A new database for breast research with infrared image. *Journal of Medical Imaging and Health Informatics*, 4(1): 92-100. <https://doi.org/10.1166/jmih.2014.1226>
- [16] Yassin, N.I., Omran, S., El Houbay, E.M., Allam, H. (2018). Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review. *Computer Methods and Programs in Biomedicine*, 156: 25-45. <https://doi.org/10.1016/j.cmpb.2017.12.012>
- [17] Dhahri, H., Al Maghayreh, E., Mahmood, A., Elkilani, W., Faisal Nagi, M. (2019). Automated breast cancer diagnosis based on machine learning algorithms. *Journal of Healthcare Engineering*, 2019(1): 4253641. <https://doi.org/10.1155/2019/4253641>
- [18] Ahmed, L., Iqbal, M.M., Aldabbas, H., Khalid, S., Saleem, Y., Saeed, S. (2023). Images data practices for semantic segmentation of breast cancer using deep neural network. *Journal of Ambient Intelligence and Humanized Computing*, 14(11): 15227-15243. <https://doi.org/10.1007/s12652-020-01680-1>
- [19] Jiménez-Gaona, Y., Rodríguez-Álvarez, M.J., Lakshminarayanan, V. (2020). Deep-learning-based computer-aided systems for breast cancer imaging: A critical review. *Applied Sciences*, 10(22): 8298. <https://doi.org/10.3390/app10228298>
- [20] Abdelrahman, L., Al Ghamdi, M., Collado-Mesa, F., Abdel-Mottaleb, M. (2021). Convolutional neural networks for breast cancer detection in mammography: A survey. *Computers in Biology and Medicine*, 131: 104248. <http://doi.org/10.1016/j.compbimed.2021.104248>
- [21] Schaefer, G., Závisek, M., Nakashima, T. (2009). Thermography based breast cancer analysis using statistical features and fuzzy classification. *Pattern Recognition*, 42(6): 1133-1137. <https://doi.org/10.1016/j.patcog.2008.08.007>
- [22] Acharya, U.R., Ng, E.Y.K., Tan, J.H., Sree, S.V. (2012). Thermography based breast cancer detection using texture features and support vector machine. *Journal of Medical Systems*, 36: 1503-1510. <https://doi.org/10.1007/s10916-010-9611-z>
- [23] Mookiah, M.R.K., Acharya, U.R., Ng, E.Y.K. (2012). Data mining technique for breast cancer detection in thermograms using hybrid feature extraction strategy. *Quantitative InfraRed Thermography Journal*, 9(2): 151-165. <https://doi.org/10.1080/17686733.2012.738788>
- [24] Milosevic, M., Jankovic, D., Peulic, A. (2014). Thermography based breast cancer detection using texture features and minimum variance quantization. *EXCLI Journal*, 13: 1204.
- [25] Pramanik, S., Bhattacharjee, D., Nasipuri, M. (2015). Wavelet based thermogram analysis for breast cancer detection. In 2015 International Symposium on Advanced Computing and Communication (ISACC), Silchar, India, pp. 205-212. <https://doi.org/10.1109/ISACC.2015.7377343>
- [26] Golestani, N., EtehadTavakol, M., Ng, E.Y.K. (2014). Level set method for segmentation of infrared breast thermograms. *EXCLI Journal*, 13: 241.
- [27] Ghayoumi Zadeh, H., Fayazi, A., Binazir, B., Yargholi, M. (2021). Breast cancer diagnosis based on feature extraction using dynamic models of thermal imaging and deep autoencoder neural networks. *Journal of Testing and Evaluation*, 49(3): 1516-1532. <https://doi.org/10.1520/JTE20200044>
- [28] Fernández-Ovies, F.J., Santiago Alférez-Baquero, E., de Andrés-Galiana, E.J., Cernea, A., Fernández-Muñiz, Z., Fernández-Martínez, J.L. (2019). Detection of breast cancer using infrared thermography and deep neural networks. In *Bioinformatics and Biomedical Engineering: 7th International Work-Conference, IWBBIO 2019, Granada, Spain, Proceedings, Part II*. Springer International Publishing, 7: 514-523. [https://doi.org/10.1007/978-3-030-17935-9\\_46](https://doi.org/10.1007/978-3-030-17935-9_46)
- [29] Tello-Mijares, S., Woo, F., Flores, F. (2019). Breast cancer identification via thermography image segmentation with a gradient vector flow and a convolutional neural network. *Journal of Healthcare Engineering*, 2019(1): 9807619. <https://doi.org/10.1155/2019/9807619>
- [30] Farooq, M.A., Corcoran, P. (2020). Infrared imaging for human thermography and breast tumor classification using thermal images. In 2020 31st Irish Signals and Systems Conference (ISSC), Letterkenny, Ireland, pp. 1-6. <https://doi.org/10.1109/ISSC49989.2020.9180164>
- [31] Civilibal, S., Cevik, K.K., Bozkurt, A. (2023). A deep learning approach for automatic detection, segmentation and classification of breast lesions from thermal images. *Expert Systems with Applications*, 212: 118774. <https://doi.org/10.1016/j.eswa.2022.118774>
- [32] Al-Antari, M.A., Han, S.M., Kim, T.S. (2020).

- Evaluation of deep learning detection and classification towards computer-aided diagnosis of breast lesions in digital X-ray mammograms. *Computer Methods and Programs in Biomedicine*, 196: 105584. <https://doi.org/10.1016/j.cmpb.2020.105584>
- [33] Moreira, I.C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M.J., Cardoso, J.S. (2012). INbreast: toward a full-field digital mammographic database. *Academic Radiology*, 19(2): 236-248. <https://doi.org/10.1016/j.acra.2011.09.014>
- [34] Heath, M., Bowyer, K., Kopans, D., Kegelmeyer Jr, P., Moore, R., Chang, K., Munishkumaran, S. (1998). Current status of the digital database for screening mammography. In *Digital Mammography: Nijmegen*. Dordrecht: Springer Netherlands. Springer, Dordrecht, 1998: 457-460. [https://doi.org/10.1007/978-94-011-5318-8\\_75](https://doi.org/10.1007/978-94-011-5318-8_75)
- [35] Baccouche, A., Garcia-Zapirain, B., Olea, C.C., Elmaghraby, A.S. (2021). Breast lesions detection and classification via YOLO-Based fusion models. *Computers, Materials & Continua*, 69(1): 1407-1425. <https://doi.org/10.32604/cmc.2021.018461>
- [36] Aly, G.H., Marey, M., El-Sayed, S.A., Tolba, M.F. (2021). YOLO based breast masses detection and classification in full-field digital mammograms. *Computer Methods and Programs in Biomedicine*, 200: 105823. <https://doi.org/10.1016/j.cmpb.2020.105823>
- [37] Hamed, G., Marey, M., Amin, S.E., Tolba, M.F. (2021). Automated breast cancer detection and classification in full field digital mammograms using two full and cropped detection paths approach. *IEEE Access*, 9: 116898-116913. <https://doi.org/10.1109/ACCESS.2021.3105924>
- [38] Kolchev, A., Pasyukov, D., Egoshin, I., Kliouchkin, I., Pasyukova, O., Tumakov, D. (2022). YOLOv4-based CNN model versus nested contours algorithm in the suspicious lesion detection on the mammography image: A direct comparison in the real clinical settings. *Journal of Imaging*, 8(4): 88. <https://doi.org/10.3390/jimaging8040088>
- [39] Matworks. (2023). Image Labeler. <https://www.mathworks.com/help/vision/ug/get-started-with-the-image-labeler.html>, accessed on 1 August 2023.
- [40] Skalski, P. (2023). Make Sense. <https://www.makesense.ai/>, accessed on 1 February 2023.
- [41] Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A. (2010). The pascal visual object classes (VOC) challenge. *International Journal of Computer Vision*, 88: 303-338. <https://doi.org/10.1007/s11263-009-0275-4>
- [42] Wu, J.D., Huang, Y.H. (2023). Enhanced identification of internal casting defects in vehicle wheels using YOLO object detection and X-ray inspection. *Traitement du Signal*, 40(5): 1909-1920. <https://doi.org/10.18280/ts.400511>
- [43] Pestana, D., Miranda, P.R., Lopes, J.D., Duarte, R.P., Véstias, M.P., Neto, H.C., De Sousa, J.T. (2021). A full featured configurable accelerator for object detection with YOLO. *IEEE Access*, 9: 75864-75877. <https://doi.org/10.1109/ACCESS.2021.3081818>
- [44] Redmon, J., Farhadi, A. (2017). YOLO9000: Better, faster, stronger. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7263-7271. <https://doi.org/10.48550/arXiv.1612.08242>
- [45] Seong, S., Song, J., Yoon, D., Kim, J., Choi, J. (2019). Determination of vehicle trajectory through optimization of vehicle bounding boxes using a convolutional neural network. *Sensors*, 19(19): 4263. <https://doi.org/10.3390/s19194263>
- [46] Abas, S.M., Abdulazeez, A.M., Zeebaree, D.Q. (2022). A YOLO and convolutional neural network for the detection and classification of leukocytes in leukemia. *Indonesian Journal of Electrical Engineering and Computer Science*, 25(1): 200-213. <http://doi.org/10.11591/ijeecs.v25.i1.pp200-213>
- [47] Kim, S., Kim, H. (2021). Zero-centered fixed-point quantization with iterative retraining for deep convolutional neural network-based object detectors. *IEEE Access*, 9: 20828-20839. <https://doi.org/10.1109/ACCESS.2021.3054879>
- [48] Bi, F., Yang, J. (2019). Target detection system design and FPGA implementation based on YOLOv2 algorithm. In *2019 3rd International Conference on Imaging, Signal Processing and Communication (ICISPC)*, Singapore, pp. 10-14. <https://doi.org/10.1109/ICISPC.2019.8935783>
- [49] Benjdira, B., Khurshed, T., Koubaa, A., Ammar, A., Ouni, K. (2019). Car detection using unmanned aerial vehicles: Comparison between faster R-CNN and yolov3. In *2019 1st International Conference on Unmanned Vehicle Systems-Oman (UVS)*, Muscat, Oman, pp. 1-6. <https://doi.org/10.1109/UVS.2019.8658300>
- [50] Bochkovskiy, A., Wang, C.Y., Liao, H.Y.M. (2020). Yolov4: Optimal speed and accuracy of object detection. *arXiv Preprint arXiv: 2004.10934*. <https://doi.org/10.48550/arXiv.2004.10934>
- [51] Li, Z., Song, J., Qiao, K., Li, C., Zhang, Y., Li, Z. (2022). Research on efficient feature extraction: Improving YOLOv5 backbone for facial expression detection in live streaming scenes. *Frontiers in Computational Neuroscience*, 16: 980063. <https://doi.org/10.3389/fncom.2022.980063>
- [52] Wang, C.Y., Liao, H.Y.M., Wu, Y.H., Chen, P.Y., Hsieh, J.W., Yeh, I.H. (2020). CSPNet: A new backbone that can enhance learning capability of CNN. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 390-391. <https://doi.org/10.48550/arXiv.1911.11929>
- [53] Fang, Y., Guo, X., Chen, K., Zhou, Z., Ye, Q. (2021). Accurate and automated detection of surface knots on sawn timbers using YOLO-V5 model. *BioResources*, 16(3): 5390. <http://doi.org/10.15376/biores.16.3.5390-5406>
- [54] UHeM. (2023). The National Center for High Performance Computing (UHeM). <https://en.uhem.itu.edu.tr/donanim.html>
- [55] Çivilibal, S., Çevik, K.K., Bozkurt, A. (2023). Derin öğrenme yardımıyla aktif termogramlar üzerinden meme lezyonlarının sınıflandırması. *Süleyman Demirel University Faculty of Arts and Science Journal of Science*, 18(2): 140-156. <http://doi.org/10.29233/sdufeffd.1141226>