

Received 26 May 2024, accepted 17 July 2024, date of publication 25 July 2024, date of current version 5 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3433559

## RESEARCH ARTICLE

# Self-Supervised Bi-Pipeline Learning Approach for High Interpretation of Breast Thermal Images

ROSLIDAR ROSLIDAR<sup>1,2</sup>, (Member, IEEE), MUHAMMAD JUREJ ALHAMD<sup>1</sup>,  
AULIA RAHMAN<sup>1,3</sup>, (Member, IEEE), KHAIRUN SADDAMI<sup>1</sup>, (Member, IEEE),  
FITRI ARNIA<sup>1,2</sup>, (Member, IEEE), MAIMUN SYUKRI<sup>4</sup>,  
AND KHAIRUL MUNADI<sup>1,2</sup>, (Member, IEEE)

<sup>1</sup>Department of Electrical and Computer Engineering, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia

<sup>2</sup>Telematics Research Center, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia

<sup>3</sup>Department of Mechanical Engineering, Faculty of Engineering, Universiti Malaya, Kuala Lumpur 50603, Malaysia

<sup>4</sup>Department of Internal Medicine, Faculty of Medicine, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia

Corresponding author: Roslidar Roslidar (roslidar@usk.ac.id)

This work was supported by the Institute for Research and Community Services (LPPM) Universitas Syiah Kuala under Scheme of Penelitian Lektor Kepala under Grant 251/UN11.2.1/PG.01.03/SPK/PTNBH/2024.

**ABSTRACT** The image quality supports a high accuracy rate of medical image diagnosis using computer vision. Digital thermal images resulting from the thermal device usually suffer from many watermarks that may lower the neural network learning performance. Thus, providing only the region of interest (RoI) of the breast area from the breast thermal images for early breast cancer detection is an important task. The goal of our work are to develop a deep learning (DL) model for taking the RoI of the breast thermal images, built a self-supervised DL model to classify the breast thermal images into healthy and cancer categories, and integrated these two models as end-to-end bi-pipeline model for breast thermal image recognition. The segmentation model was built using attention U-Net with residual recurrent network called R2AU-Net, and the classification model was built using self-supervised learning consisting of the Simple Framework for Contrastive Learning of Visual Representations (SimCLR) and ResNet50. These networks were trained using unlabelled limited breast thermal datasets to allow more comprehensive learning. The result shows that proposed self-supervised bi-pipeline model can take the RoI with an accuracy rate of 98.63% and classify the breast thermal images with a top-1 accuracy rate of 84.37% and top-5 accuracy rate of 96.87%. In addition, the bi-pipeline model implementation using a central processing unit shows that the model required only about 4 seconds for segmentation and classification tasks. These findings indicate that the bi-pipeline model can effectively aid the interpretation of unlabeled breast thermal images.

**INDEX TERMS** Deep learning, self-supervised learning, R2AU-net, SimCLR, breast thermal image.

## I. INTRODUCTION

The development of breast cancer screening methods is an important step to facilitate early detection and treatment of breast cancer. Many researches have proposed a deep learning (DL) based algorithm to recognize the indication of breast cancer using a non-invasive breast thermal images [1], [2],

The associate editor coordinating the review of this manuscript and approving it for publication was Abedalrhman Alkhateeb<sup>1</sup>.

[3], [4], [5], [6], but mostly focus on a single task within the DL model, either classification, segmentation, or detection of abnormalities.

In fact, breast thermal images often suffer from issues such as noise, variable size, and watermarks. These factors can negatively affect the performance of DL models, making it essential to improve image quality to ensure effective learning and accurate predictions. Typically, captured thermal images include not just the breast area but also other body parts,

such as the neck and abdomen, which can interfere with classification results.

Extracting the region of interest (RoI) of the breast area from breast thermal images has been the focus of many studies. For example, Pramanik et al. [7] implemented a level set method (LSM) algorithm, but the simulation was complex and achieved an accuracy rate of only 72.18%. Sánchez-Ruiz et al. [8] proposed an automatic method to segment the RoI based on local operations, local analysis, interpolation, and statistical operators. Roslidar et al. [9] proposed an improved second-order polynomial curve fitting technique employing edge detection to effectively capture the RoI of breast thermal images. But both works do not implement direct pipeline from the segmentation algorithm to the classification task. Therefore, in this study, we created a framework of bi-pipeline model to improve breast thermal images interpretation by combining end-to-end segmentation and classification models. The segmentation model focuses on isolating the breast area, while the classification model is used to categorize the images in binary classes of healthy and cancer.

A significant challenge in building breast thermal image classification models is the difficulty in collecting thermal images of breast cancer patients, resulting in an unbalanced dataset with relatively small cancer datasets compared to healthy ones. To address this issue, we used a self-supervised learning approach, which is effective with small datasets. In this paper, we proposed a bi-pipeline deep learning model that utilized the Recurrent Residual Attention U-Net (R2AU-Net) [10] algorithm for the segmentation, and the Simple Framework for Contrastive Learning of Visual Representations (SimCLR) [11] for the classification.

Overall, this article describes our work on developing a DL model that takes the RoI of breast thermal images and classifies the images. This coupled learning approach of segmentation and classification, called bi-pipeline, is needed to allow an automatic, fast, and accurate prediction model in interpreting breast thermal images. The key contributions of our work are as follows:

- 1) Introducing the step-by-step approach of recurrent residual attention U-Net for breast thermal image segmentation.
- 2) Developing a breast thermal image classification model using self-supervised learning consisting of the Simple Framework for Contrastive Learning of Visual Representations (SimCLR) and ResNet50.
- 3) Validating the performance of proposed novel bi-pipeline deep learning model using learning curves, evaluation metrics of accuracy, sensitivity, specificity, precision, F1-score, Jaccard Similarity, and Dice Similarity for the segmentation model, and evaluation metrics of top-1 accuracy and top-5 accuracy for the classifier model.

- 4) Analysing the bi-pipeline model performance on central processing units (CPU) for practical implementation.
- 5) Exposing the proposed bi-pipeline model properties and comparing the research findings with existing works.

The rest of this article is organized as follows. Section II presents a review of related works and the deep learning concept relevant to the approach methods. Section III describes the details of the novel proposed network framework and the learning process. Section IV discusses the proposed bi-pipeline model along with its performance. In this section, we also provide the potential implementation of our proposed bi-pipeline as an early breast cancer detection system for clinical practice. Then, we provide the comparison of our work with similar existing work in Section V. Finally, Section VI highlights the research findings.

## II. RELATED WORKS

Research on developing non-invasive screening methods based on thermography has advanced significantly. In the context of breast cancer screening, the use of breast thermal images to detect early signs of breast cancer has led to numerous proposals for computer-aided algorithms leveraging artificial intelligence. This section reviews related works that have contributed to the development of breast thermal image classifier models based on deep learning over the past decade. Conceptual of the approach methods, along with existing works that applied similar networks, are also provided in the following section.

### A. CONVOLUTIONAL NEURAL NETWORK

Deep learning has gained significant traction due to the convolution operation's ability to extract image features and the deep neural network's capacity to learn these features' characteristics. Convolutional Neural Networks (CNNs) are designed to process input images by assigning learnable weights and biases to map key features distinguishing one image from another. This process enables the generation of classification results. Figure 1 illustrates the general architecture of CNNs used for classifying breast thermograms into categories of healthy and cancer. The classification process involves three major steps: dataset preparation through image preprocessing, feature learning, and classification. The classification can be binary (healthy and cancer) or multi-class, such as healthy, benign, and malignant.

Neural Networks (NNs) are commonly employed for recognizing and detecting objects in image data. In general, CNNs share similarities with other NNs, such as weights, biases, and activation functions when processing inputs. However, CNNs excel in feature extraction, allowing them to learn patterns from high-dimensional inputs. This process, known as convolution, is carried out in a convolutional layer (feature extraction layer). As depicted in Figure 1, CNNs

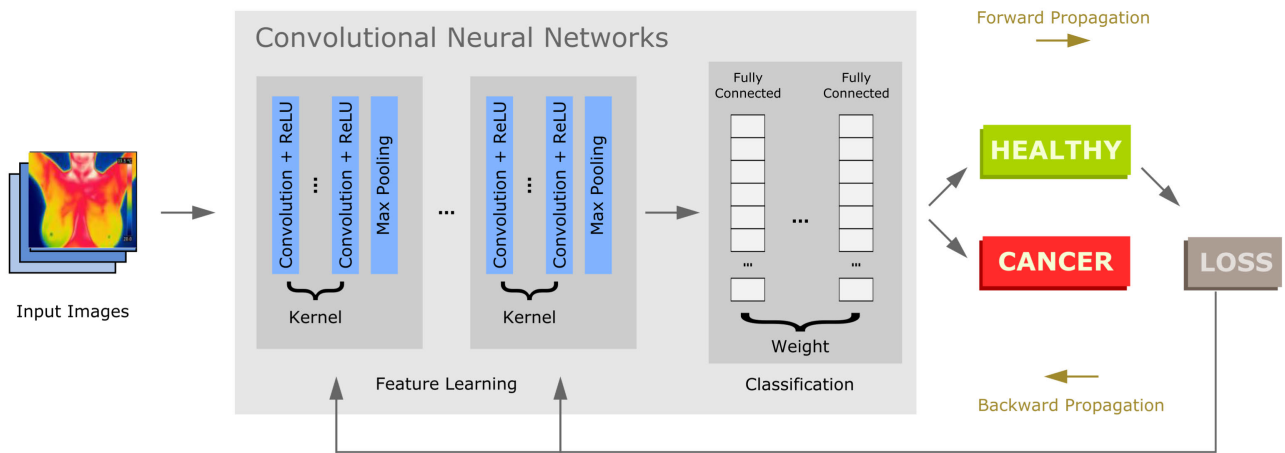


FIGURE 1. Convolutional Neural Network for Breast Thermal Image Classification [12].

comprise two major layers: the feature extraction layer and the fully connected layer.

In 2018, Baffa and Lattari [13] developed a CNN model with two convolutional layers, each with a size of  $5 \times 5$  and 32 outputs, followed by two Max Pooling Layers with  $5 \times 5$  and a threshold of 3. The output layer is fully connected and classifies the data into healthy and unhealthy classes. Roslidar et al. [4] also classified breast thermal images into binary classes but used transfer learning rather than building a network from scratch. This strategy allows for faster model development with high performance. Other studies in the same year implemented transfer learning to identify the best network for learning breast thermal image features [2], [14]. Research in this field continues to develop to this day [15], [16], [17], [18].

### B. U-NET WITH RECURRENT RESIDUAL ATTENTION GATE

U-Net is a CNN architecture used for image segmentation tasks, developed by Ronneberger et al. [19] in 2015. U-Net is widely used in the field of medical image processing, especially in the task of segmenting organs or structures in medical images. U-Net's framework begins with cascading convolutions that combine spatial information at different resolutions. This architecture has two main paths: contraction path (downsampling) and expansion path (upsampling). The downsampling path consists of multiple convolution and max-pooling layers to reduce image resolution. Then, the expansion path uses conventional convolution and transposed convolution layers to increase image resolution and produce final segmentation [19]. In enhancing the breast thermal images classification, works implementing U-Net for segmentation have been developed [6], [21], [22].

Conventional U-Net has the disadvantage that it tends to produce uneven segmentation on objects with complex or blurred boundaries because U-Net cannot remove noise that interferes with images [23]. To eliminate noise or focus only on the important features in the image, U-Net requires an attention mechanism that increases the weight on certain parts

of the feature. This mechanism can improve the segmentation accuracy as only focuses on important features and mitigate noise.

In our work, we use the ability of the attention U-Net [23], [24] to manage the noise of breast thermal images that may interfere with the segmentation process. With this mechanism, the model will only focus on the breast features by giving greater weight than those outside the breast.

The Recurrent Residual Attention U-Net (R2AU-Net) enhances U-Net's capabilities by incorporating recurrent residual convolutional neural networks (RRCNN) and attention mechanisms. This approach improves the network's ability to focus on important regions while ignoring irrelevant information, making it suitable for segmenting complex medical images, including breast thermograms.

Figure 2 shows the architecture of the R2AU-Net mechanism. The recurrent net and residual mechanisms help eliminate the segmentation problem of contours that are not smooth or unclear in U-Net. By adding a residual module, R2AU-Net can reduce the effect of the deep convolution layer on the U-Net, which causes detailed information to be lost by repeating the convolution process several times. This technique is called the recurrent mechanism [25].

Recurrent convolutional neural network (RCNN) contains a stack of residual convolutional layers (RCLs) and interleaved with max pooling layers. The RCL are performed with respect to the discrete time steps that are expressed according to the RCNN. Liang and Hu [25] expressed the output of the nets as,

$$z_{ijk}(t) = (\mathbf{w}_k^f)^T \mathbf{u}^{(i,j)}(t) + (\mathbf{w}_k^r)^T \mathbf{x}^{(i,j)}(t-1) + b_k, \quad (1)$$

where  $\mathbf{u}^{(i,j)}(t)$  and  $\mathbf{x}^{(i,j)}(t-1)$  represents the feed-forward and recurrent input, respectively. Then, the outputs are fed to the standard ReLU activation function  $f$  as follows:

$$x_{ijk}(t) = g(f(z_{ijk}(t))) = \max(z_{ijk}(t), 0). \quad (2)$$

For R2U-Net, the final outputs of RCNN unit ( $x_{l+1}$ ) are passed through the residual unit and calculated as

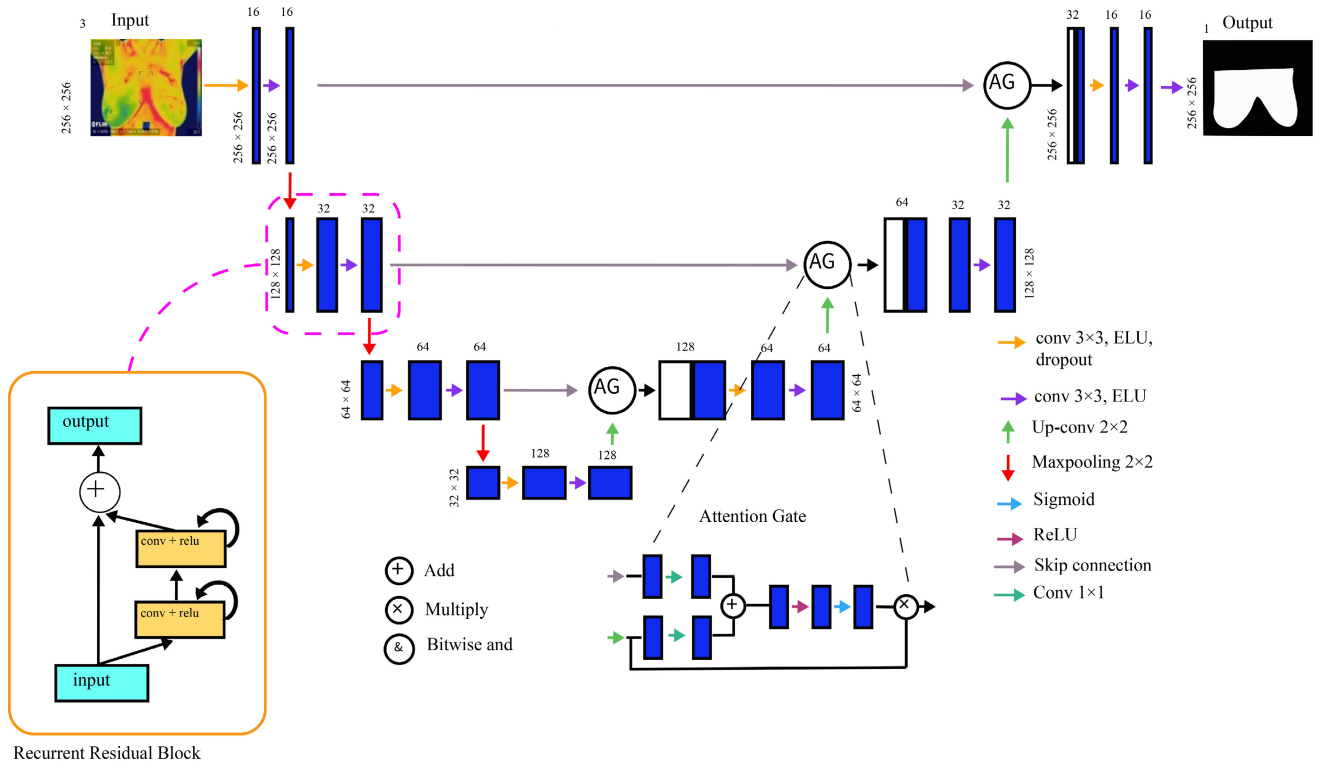


FIGURE 2. R2AU-Net architecture for breast thermal images segmentation.

follows [26]:

$$x_{l+1} = x_l + \mathcal{F}(x_l, w_l) \tag{3}$$

with  $x_l$  as the recurrent input of layer  $l$ .

The attention gates are incorporated into the conventional U-Net architectures to extract salient features that are passed through the skip connection. This operation is performed right before the concatenation operation to merge only relevant activations. Gradients resulting from background regions are down-weighted during the backward pass, thus allowing the model parameters in shallower layers to be updated mostly based on spatial regions that are relevant to a given task. The update rule for convolution parameters in layer  $l - 1$  is formulated in equation 4 [23].

$$\begin{aligned} \frac{\partial (\hat{x}_i^l)}{\partial (\Phi^{l-1})} &= \frac{\partial (\alpha_i^l f(x_i^{l-1}; \Phi^{l-1}))}{\partial (\Phi^{l-1})} \\ &= \alpha_i^l \frac{\partial (f(x_i^{l-1}; \Phi^{l-1}))}{\partial (\Phi^{l-1})} + \frac{\partial (\alpha_i^l)}{\partial (\Phi^{l-1})} x_i^l, \end{aligned} \tag{4}$$

where  $\alpha_i^l$  is the scale of the first gradient and a vector at each grid scale in case of multi-dimensional attention gates.

### C. SELF-SUPERVISED LEARNING

Large-scale labeled data are generally required to train deep neural networks effectively, enhancing performance in visual feature learning from images or videos for computer vision

applications [27], [28]. However, the data collection and annotation of large-scale datasets are time-consuming and costly. For medical images, creating a large, curated dataset to train deep learning algorithms is particularly expensive and time-consuming. Unlike photographic images, which can be sourced online and labeled by non-experts through crowd-sourcing, medical images require expert annotation [29]. To avoid the extensive costs associated with collecting and annotating large-scale datasets, self-supervised learning (SSL) methods, a subset of unsupervised learning, have been proposed to learn from general images without using any human-annotated labels [30], [31]. In application domains such as medical imaging, where high-quality labeled data is scarce and data modalities shift rapidly, self-supervised learning (SSL) has the potential to make a significant impact [32]. Recent advances have demonstrated that SSL methods can learn representations that are competitive with those learned in a fully supervised manner for medical imaging [33], [34].

### D. CONTRASTIVE LEARNING IMPLEMENTING SIMPLE FRAMEWORK

Contrastive Learning (CL) is a deep learning technique for unsupervised representation learning. CL is used to learn the general features of a dataset without labels by teaching the model which data points are similar or different. The goal is to learn such an embedding space in which similar sample

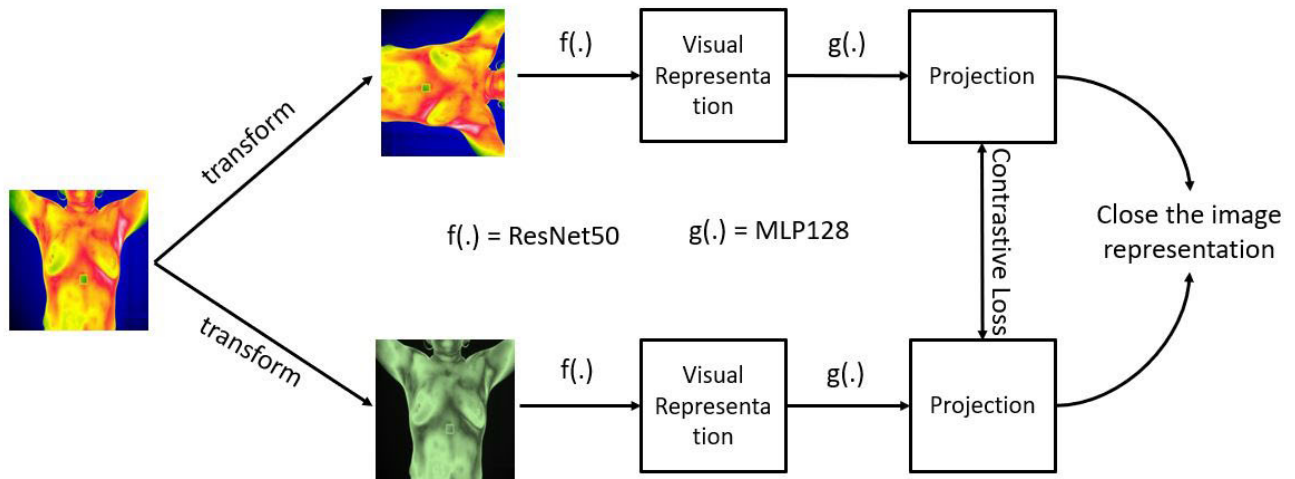


FIGURE 3. SimCLR framework on breast thermal image learning.

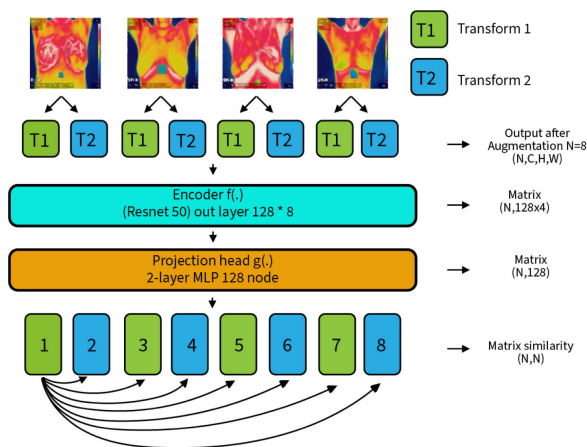


FIGURE 4. Step-by-step learning of SimCLR on breast thermal images.

pairs stay close to each other while dissimilar ones are far apart [35]. Contrastive self-supervised learning methods [11], [36], [37], are currently state of the art in SSL. In these methods, the final layer’s feature representations for positive pairs of images are brought closer, and those for negative pairs are pushed away using contrastive loss functions such as InfoNCE [11].

Several previous studies have applied contrastive self-supervised learning methods in the case of breast cancer classification. Miller et al. [38] implemented several SSL models such as SimCLR, ViT-MAE, BYOL, and SWaV on the breast mammography datasets. The results showed that SSL is better than full supervised, with 81.5% compared to the full supervised, with only a 71% accuracy rate. Another research by Chhipa et al. [39] applied SSL with ResNet50 encoder and EffesientNetB2 on histopathological breast cancer data with different magnification levels [39] achieved a good accuracy value with a small amount of data. Saidnassim et al. [40] used SSL with BYOL method and

transformer base network for breast cancer segmentation on mammography dataset. The mean Intersection over Union (IOU) performance of the U-Net (ResNet34) was enhanced, using SSL, by close to 4%, and the loss function was reduced by 12.5%. These results apparently indicate the efficacy of Self-supervised learning (SSL) over supervised pre-trained models. However, SSL has not been applied to thermography datasets for breast cancer classification.

Simple Framework of Contrastive Learning of Visual Representation (SimCLR) is a CL model that can learn from representations of data without using labels. As demonstrated in Figure 3, SimCLR employs data augmentation techniques and negative sampling techniques to create self-supervised learning tasks. The effectiveness of these techniques is evident in their ability to create multiple variations of each image, forcing the model to learn to recognize the same features in different images. The negative sampling technique ensures that the model learns to discriminate between different pictures, not just match up the same pictures. This robust learning approach prevents overfitting and equips the model to recognize features that are more common and can be applied to different images.

SimCLR uses augmentation to enrich the model’s knowledge of an object to be classified by transforming the original image into several transform types, called “view” [11]. SimCLR uses several types of transforms, such as random cropping, random rotation, random color distortions, and random Gaussian blur. In our work, the original image is converted into two views (transformations).

As shown in Figure 4, the original image is changed into 2 views, namely rotation and color distortion, which are useful for enriching the model’s knowledge of the trained object. The combination of image transformation that has the highest accuracy is crop and color distortion. If the batch size (N) in each training is 4, then the output after going through the batch size augmentation process will be  $(n_{view} \times N)$ , or  $n_{view}$  or the number of views multiplied by the initial batch

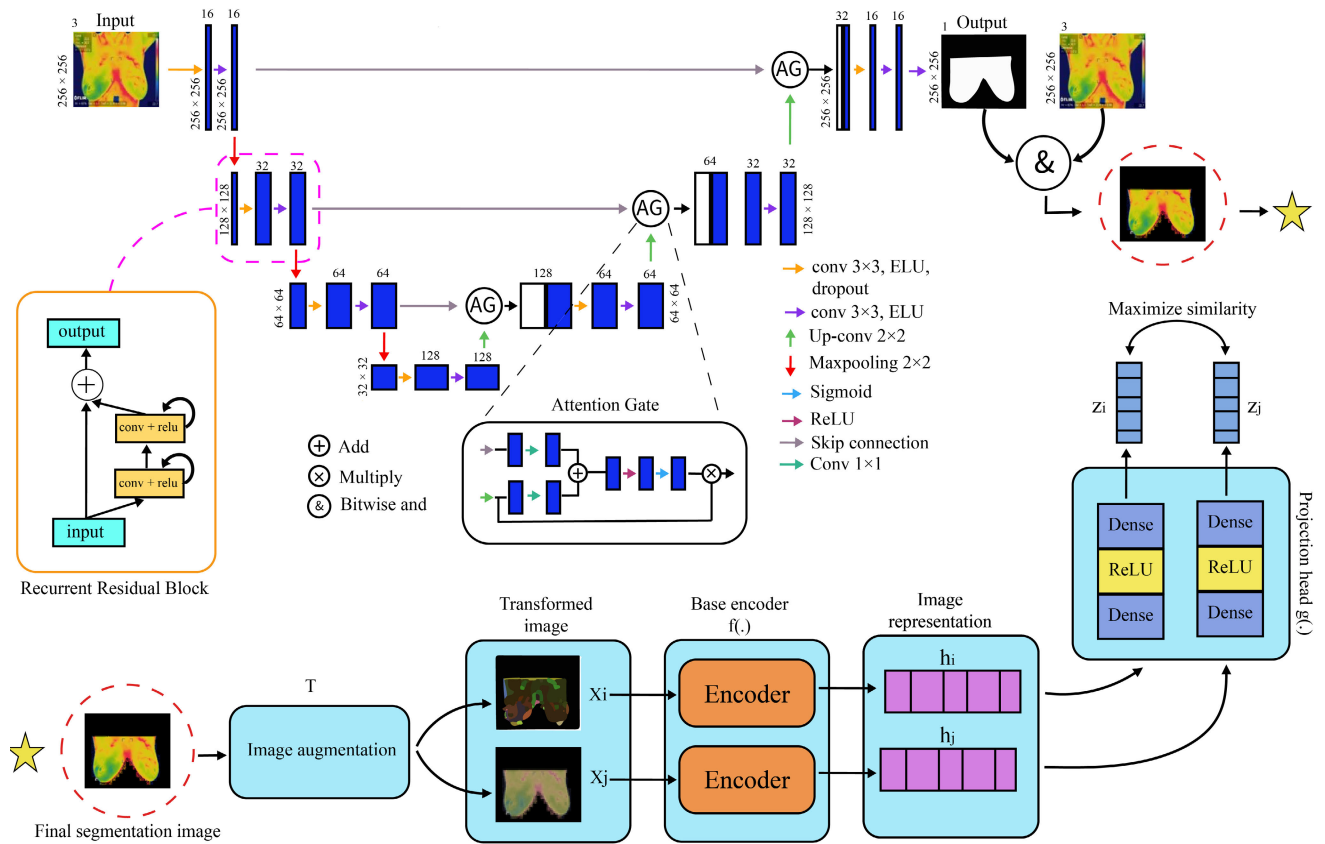


FIGURE 5. The proposed bi-pipeline learning workflow.

size of  $2 \times 4 = 8$ . The two images that come out of the augmentation process are positive pairs or the same image. With SimCLR, these two images will later be brought closer together and placed in the same cluster in latent space. After the augmentation process, the encoder will produce an image representation.

III. METHOD

In this section, the processes of dataset preparation and model development are described. The proposed bi-pipeline model is explained, covering the concepts and algorithms for both segmentation and classification tasks. Finally, the discussion ends with model performance evaluation using learning curves and evaluation metrics.

A. DATASET

In this study, we used breast thermal images grouped into three datasets: training with 400 images, validation with 100 images, and testing with 150 images. The images were downloaded from the Database for Mastology Research (DMR) [41] which provides breast thermal images with five views with more than 20 images per patient. We sorted the images with similar size of  $256 \times 256$  and resized into  $224 \times 224$  as required by the network input specification. We maintained a balanced dataset with an equal number

of images from healthy and cancer classes. The R2AU-Net segmentation model requires the same number of mask images as the training, testing, and validation data. The segmentation network uses the mask image to guide the segmentation process to match the target segment. To evaluate the classifier model’s performance, the dataset for training SimCLR was divided into two subsets: one resulting from the segmentation process of the R2AU-Net model and one without segmentation (full body dataset).

B. PROPOSED ALGORITHM

Proposed bi-pipeline framework is divided into two parts: segmentation and classification. The segmentation model focuses on extracting the region of interest (RoI) of breast thermal images obtained from DMR [41], primarily encompassing the area from the head to the waist. This work is important in the deep learning process; as Roslidar et al. [9] mentioned, feeding only the input’s RoI into the learning network will improve the learning accuracy. Thus, the RoI of the breast area was segmented using the R2AU-Net and fed into the Simple Framework for Contrastive Learning of Visual Representations (SimCLR) for the classification task.

Figure 5 shows our proposed model workflow, where a thermal image with an image size of  $224 \times 224$  is fed into the R2AU-Net segmentation model to take only the breast

areas. In the R2AU-Net model, the thermal image passes through several convolution layers with a recurrent residual CNN architecture. In the recurrent residual CNN there is a skip connection from the input to the output layer of the CNN.

The algorithm comprises a segmentation process implemented R2AU-Net using Algorithm 1 and a classification process implemented SimCLR and ResNet50 using Algorithm 2. To minimize the duration of the training, we applied the pre-trained networks. Thus, initial weights were taken from the network when trained on the Imagenet.

---

#### Algorithm 1 Segmentation Using R2AU-Net

---

**Input:** Pre-trained R2AU-Net, threshold, breast thermal images ( $I_0$ )

**Output:** ROI of breast thermal images,  $I_S$

- 1: **for** each image in dataset **do**
  - 2:      $I_0 \leftarrow$  read image
  - 3:      $I_r \leftarrow$  resize of  $I_0$  to  $224 \times 224$
  - 4:      $I_m \leftarrow$  apply pre-trained R2AU-net to  $I_r$
  - 5:      $M_0 \leftarrow$  0 if  $I_m$  less than threshold, else 1
  - 6:      $I_s \leftarrow$  apply bitwise-and-operator to  $I_0$  and  $M_0$
  - 7: **end for.**
  - 8: **return**  $I_S$
- 

---

#### Algorithm 2 SimCLR Training Process Using Segmented Breast Thermal Images

---

**Input:** Segmented breast thermal images, batch size ( $N$ ), temperature ( $\tau$ ), encoder  $f(\cdot)$ , projection head  $g(\cdot)$ , augmented function

**Output:** Trained SimCLR encoder,  $f(\cdot)$

- 1: **for** sample minibatch from segmented breast thermal dataset **do**
  - 2:     **for** all  $k \in \{1 \dots, N\}$  **do**
  - 3:          $I_s \leftarrow$  segmented images
  - 4:          $x_i, x_j \leftarrow$  apply augmented function to  $I_m$
  - 5:          $h_i, h_j \leftarrow$  apply  $x_i$  and  $x_j$  to  $f(\cdot)$
  - 6:          $z_i, z_j \leftarrow$  apply  $h_i, h_j$  to  $g(\cdot)$
  - 7:     **end for.**
  - 8:     **for** all  $i \in \{1 \dots, 2N\}$  and  $j \in \{1 \dots, 2N\}$  **do**
  - 9:          $S_{i,j} \leftarrow$  apply pairwise similarity function to  $z_i$ , and  $z_j$
  - 10:          $h_i, h_j \leftarrow$  apply  $x_i$  and  $x_j$
  - 11:     **end for.**
  - 12:      $L \leftarrow$  apply  $S_{i,j}$  to NT-Xent loss with constant  $\tau$
  - 13:     Update network  $f$  and  $g$  to minimize  $L$
  - 14: **end for.**
  - 15: **return** Trained SimCLR encoder  $f(\cdot)$
- 

As shown in Algorithm 1, the segmentation process is started with providing the initial weight of the pretrained network R2AU-Net and the threshold value, following with the images along with the ground truths. Next, the network read the images, resize, and feeding the images and the ground truths into the pretrained R2AU-Net. Then, the segmented images are fed into the classification network.

Afterward, the classification network of SimCLR was trained using pretrained ResNet50 for feature extraction of the segmented images fed from the R2AU-Net. The image resulting from the segmentation of the breast becomes the input to SimCLR, where this image will be transformed into 2 views ( $x_i, x_j$ ), which are chosen randomly. In this research, the types of transformations include random cropping, random rotation, random color distortions, and random Gaussian blur. This transformation is used to enrich the feature patterns recognized by the model so that it is forced to recognize the same data sample but from a different view. The two data views that have been transformed become input to the SimCLR encoder  $f(\cdot)$  section. The encoder in SimCLR is part of the model architecture, which is responsible for converting images into representations. The main function of the encoder in SimCLR is to extract important features from the image and compress them into a smaller representation. In this research, we use ResNet50 as an encoder and added two fully connected layer as the classification head. The data representation resulting from the encoder  $f(\cdot)$  is then changed to be more compact and concise by the projection head  $g(\cdot)$ . The projection head of this research consists of two MLP layers, each consisting of 128 nodes with a ReLU activation function. The output from the projection head  $g(\cdot)$  is a representation vector with shape  $1 \times 128$ . This representation vector will then calculate the loss  $l_{(i,j)}$  using NT-Xent loss [42] as shown in the Algorithm 2 using equation 5.

$$l_{(i,j)} = -\log \frac{e^{\text{sim}(\mathbf{z}_i, \mathbf{z}_j)/\tau}}{\sum_{k=1}^{2N} 1_{k \neq i} e^{\text{sim}(\mathbf{z}_i, \mathbf{z}_k)/\tau}} \quad (5)$$

where  $\text{sim}(\mathbf{z}_i, \mathbf{z}_j)$  denotes the cosine similarity between two vectors  $\mathbf{z}_i$  and  $\mathbf{z}_j$ .

### C. MODEL DEVELOPMENT

As we have two pipelines, the model development was conducted in two parts: developing the segmentation model and building the classification model. For the first part, We applied R2AU-Net to take the region of interest (RoI) of the breast thermal image to increase the feature mapping focus only on the breast area. The segmentation model was obtained by fine-tuning the R2AU-Net hyperparameter. We fed the ground truth into the network and set the hyperparameter. During the learning process, we observed the learning performance and fine-tuned the hyperparameter to obtain the model with stable learning and a high accuracy rate. The segmentation algorithm best learnt the input with a learning rate of  $2 \times 10^{-4}$ . The residual convolutional layer was expanded to three time steps,  $t = 3$ .

The best network performance was achieved with the fined-tuned hyperparameter, as mentioned in the Subsection R2AU-Net Hyperparameter. Following by the fined-tuned hyperparameter for best SimCLR performance is provided in Subsection SimCLR+ResNet50 Hyperparameter.

1) R2AU-NET HYPERPARAMETER

- Batch Size = 4
- Epoch = 100
- Loss function = Binary Cross Entropy
- Optimizer = Adam
- Learning Rate =  $2 \times 10^{-4}$
- Recurrent process (t) = 3

2) SIMCLR+RESNET50 HYPERPARAMETER

- Batch Size = 4
- Epoch = 300 and 1000
- Loss function = NT-Xent loss
- Optimizer = AdamW
- Learning Rate =  $5 \times 10^{-4}$ ,  $2 \times 10^{-4}$ ,  $10^{-4}$
- Weight decay =  $10^{-4}$
- Temperature = 0.07
- Learning rate scheduler = CosineAnnealingLR
- Encoder = ResNet50

The proposed unsupervised classifier of SimCLR+ResNet50 was trained twice, using epoch 300 and epoch 1000 with similar hyperparameter values. Three learning rates of  $5 \times 10^{-4}$ ,  $2 \times 10^{-4}$ , and  $10^{-4}$  were applied to figure out the most stable learning of the classifier model.

Finally, the process of fine-tuning or taking the pre-trained model in this work is by freezing the parameters on the encoder trained using SimCLR and segmented dataset as shown in Figure 6. Here, fine-tuning proceeded to utilize the results from training with SSL SimCLR; fine-tuning must be done because creating a downstream task classification system requires a classification head in the form of a linear layer with a number of nodes 2. This fine-tuning process uses a learning rate of  $10^{-3}$  with a loss function called CrossEntropyLoss. The fine-tuning strategy used in this research is as follows:

- 1) Freezing the parameters on the ResNet50 encoder trained with SimCLR and segmented image data.
- 2) Training the classification head with 100 epochs with a learning rate of  $10^{-3}$ .

D. MODEL EVALUATION

To build the bi-pipeline model, we trained each of the networks, segmentation and classification, separately. We evaluated the learning processes using learning curves, accuracy rates, and loss rates. After the network training resulted in a stable learning curve, we recorded the model and tested it using evaluation metrics.

Taha and Hanbury [43] defined the medical volume by the point set  $X = \{x_1, \dots, x_n\}$  with  $|X| = w \times h \times d = n$  where  $w$ ,  $h$  and  $d$  respectively are the width, height, and depth of the grid on which the volume is defined. As we segment the images into two partitions, breast area and non breast area (background), there are four common cardinalities that reflect the overlap between the two partitions, true positive (TP), false positive, (FP), false negative (FN), and true negative

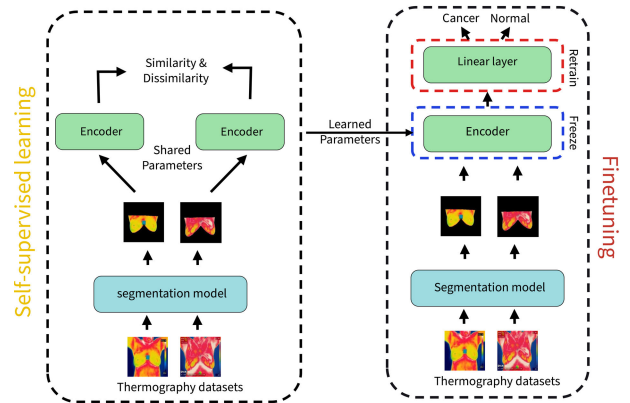


FIGURE 6. The Bi-Pipeline end to end.

(TN), which the values are determined using equation 6.

$$m_{(i,j)} = \sum_{r=1}^{|X|} f_g^i(x_r) f_t^r(x_r) \tag{6}$$

where  $f_g$  and  $f_t$  represent the ground truth segmentation and the evaluated segmentation respectively. These cardinalities provide for each pair of subsets  $i \in S_g$  (ground segmentation) and  $j \in S_t$  (evaluated segmentation), then the sum of agreement  $m_{ij}$  between them are as follows: when  $i$  is 1 or  $m_{11} = TP$  as the first class of interest; and  $i$  is 0 or  $m_{00} = TN$  for the second class or the background. Then,  $m_{10} = FP$  and  $m_{01} = FN$  are the false segmented areas.

For R2AU-Net, we used segmentation evaluation metrics of accuracy (ACC), sensitivity (SE), specificity (SP), precision (PC), F1-score, Jaccard Similarity (JS), and Dice Similarity (DC) as follows [43].

$$ACC = \frac{TP + TN}{TP + FP + FN + TN} \tag{7}$$

$$PC = \frac{TP}{TP + FP} \tag{8}$$

$$SE = \frac{TP}{TP + FN} \tag{9}$$

$$SP = \frac{TN}{TN + FP} \tag{10}$$

$$F1_{score} = \frac{2 \times Recall \times Precision}{Recall + Precision} \tag{11}$$

$$JS = \frac{TP}{TP + FP + FN} \tag{12}$$

$$DC = \frac{2TP}{2TP + FP + FN} \tag{13}$$

We also trained U-Net and Segnet with similar hyperparameter tuning and recorded the result. Then, R2AU-Net performance is compared to vanilla UNet and Segnet to see how well each model segments the breast region.

Meanwhile, for the SimCLR network, we evaluated the model with the top-1 accuracy and top-5 accuracy metrics. The top-1 accuracy metric measures the percentage of conformity between the predicted class and the actual label



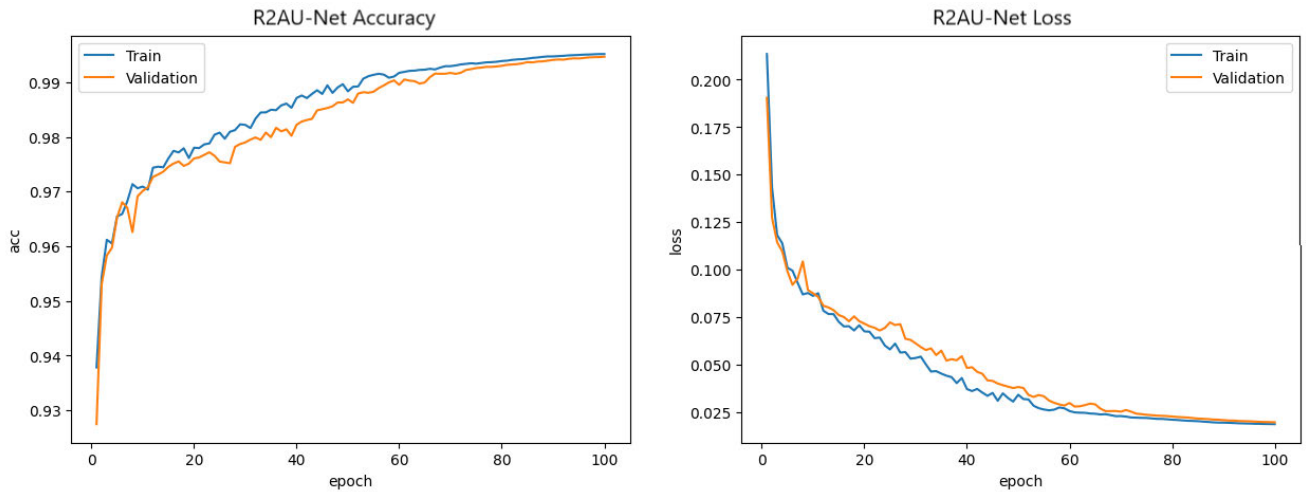


FIGURE 7. Learning curve of segmentation model using R2AU-Net.

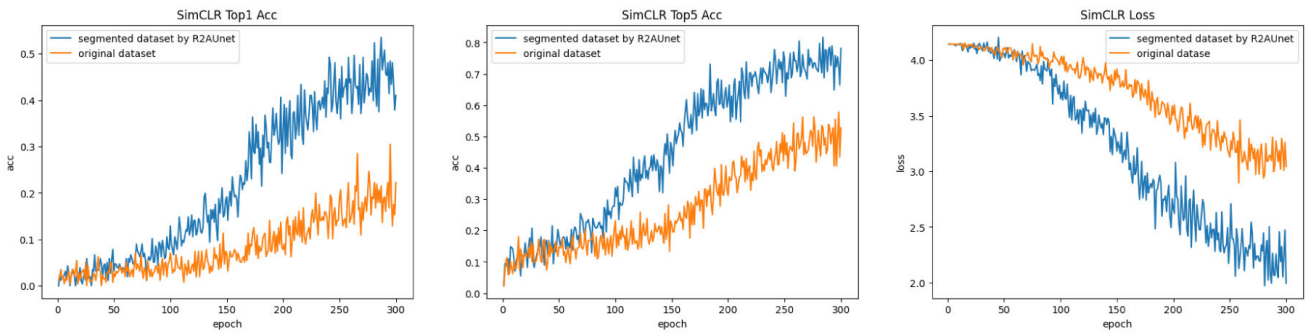


FIGURE 8. Learning curve of SimCLR with 300 epochs.

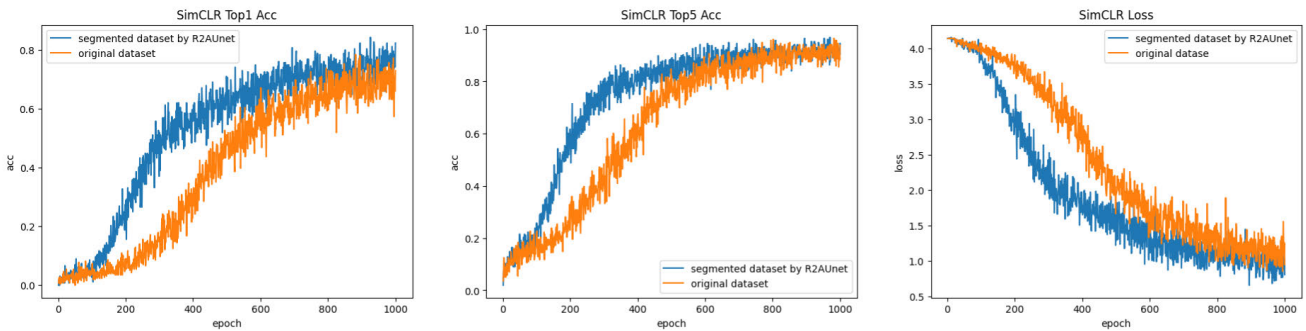


FIGURE 9. Learning curve of SimCLR with 1000 epochs.

on the test data sample. This metric only considers the highest probability predictions, i.e., the top-1 predictions. In other words, top-1 accuracy calculates what percentage of data samples have correct predictions among all samples, and top-5 accuracy measures the percentage of conformity between the predicted class and the actual label in the test data sample. Still, in this case, it considers predictions with the highest probability up to the top five predictions (top-5).

Afterward, We tested the SimCLR model using the segmented and raw datasets. Then, we integrated both R2AU-Net and SimCLR models as a bi-pipeline model and tested the end-to-end process.

#### IV. RESULT AND DISCUSSION

In this section, we evaluated and analyzed the performance of the segmentation and classification model. The learning performances were evaluated using graphs of the learning

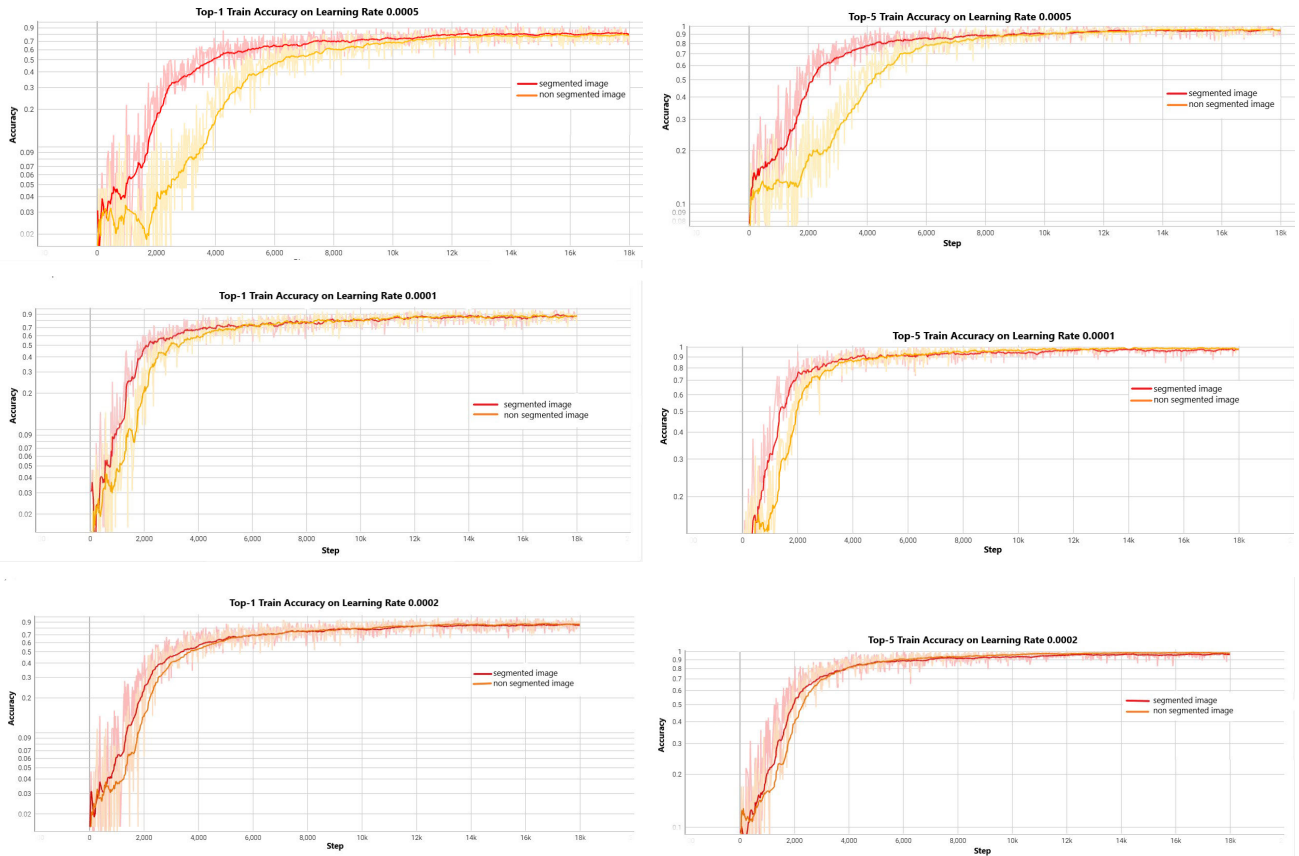


FIGURE 10. Top-1 and Top-5 learning accuracy.

TABLE 1. Segmentation model testing result.

Model	Acc	SE	SP	PC	F1	JS	DC
R2AU-Net	98.63	98.63	99.68	99.24	98.89	97.92	98.89
Unet	98.21	97.11	98.65	96.93	96.94	94.29	96.94
SegNet	93.97	89.18	96.17	90.06	89.56	82.07	89.56

curves resulting from segmentation and classification network training. Then, the resulting model interpretation ability was justified using the metric evaluation. Finally, we exposed the proposed bi-pipeline implementation performance along with its properties.

**A. LEARNING PERFORMANCE**

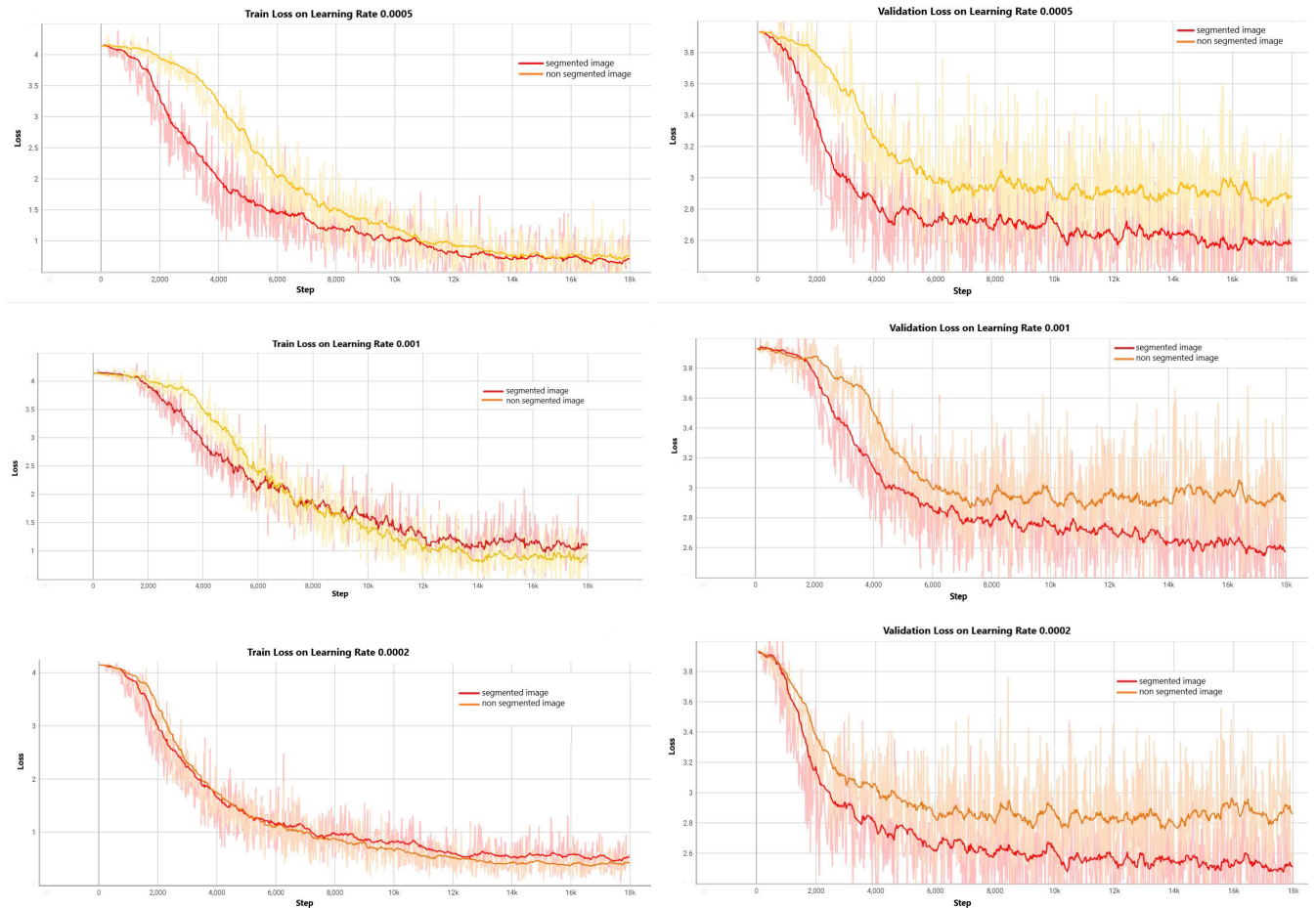
Figure 7 shows the learning curve of R2AU-Net, depicting the accuracy and loss resulting during the training. The learning approached the highest accuracy after 100 epochs of 98% and loss of 0.0195. The learning curve of training accuracy and validation accuracy confirmed that the learning is a good fit, showing a stable and consistent training performance. The proposed hyperparameter values are appropriate and in accordance with the task for breast segmentation. Our

TABLE 2. Proposed classifier model accuracy.

Dataset	Epoch 300		Epoch 1000	
	Top-1	Top-5	Top-1	Top-5
Non segmented	30.46	57.81	78.51	96.05
Segmented	53.51	81.64	<b>84.37</b>	<b>96.87</b>

work supports the stability and satisfactory results of the R2AU-Net model that was trained on breast thermal images.

Meanwhile, the unsupervised classifier model built using the SimCLR network has shown an increasing learning performance along with an increasing number of epochs, as shown in Figure 8 and Figure 9. Figure 8 shows that the learning performance significantly improved when the network was trained using the segmented images. Thus, we added epochs until 1000 to obtain higher accuracy. The learning curves confirmed that SimCLR required a large number of epochs to obtain high accuracy. Moreover, the learning curves are not stable; thus, more work on hyperparameter tuning is needed to ensure the learning is convergent. However, the learning performance consistently increased along with the epochs.



**FIGURE 11.** Training and Validation loss using various learning rates.

To improve the model performance, the classifier network was trained using various learning rates, resulting in learning performance as shown in Figure 10 for the learning accuracy rate and Figure 11 for the learning loss rate. In our investigation of the learning curves depicting the top-1 train accuracy for a deep learning model trained with three different learning rates ( $10^{-4}$ ,  $2 \times 10^{-4}$ , and  $5 \times 10^{-4}$ ), several intriguing patterns and insights emerged. These observations shed light on the intricate relationship between learning rates and model convergence and performance.

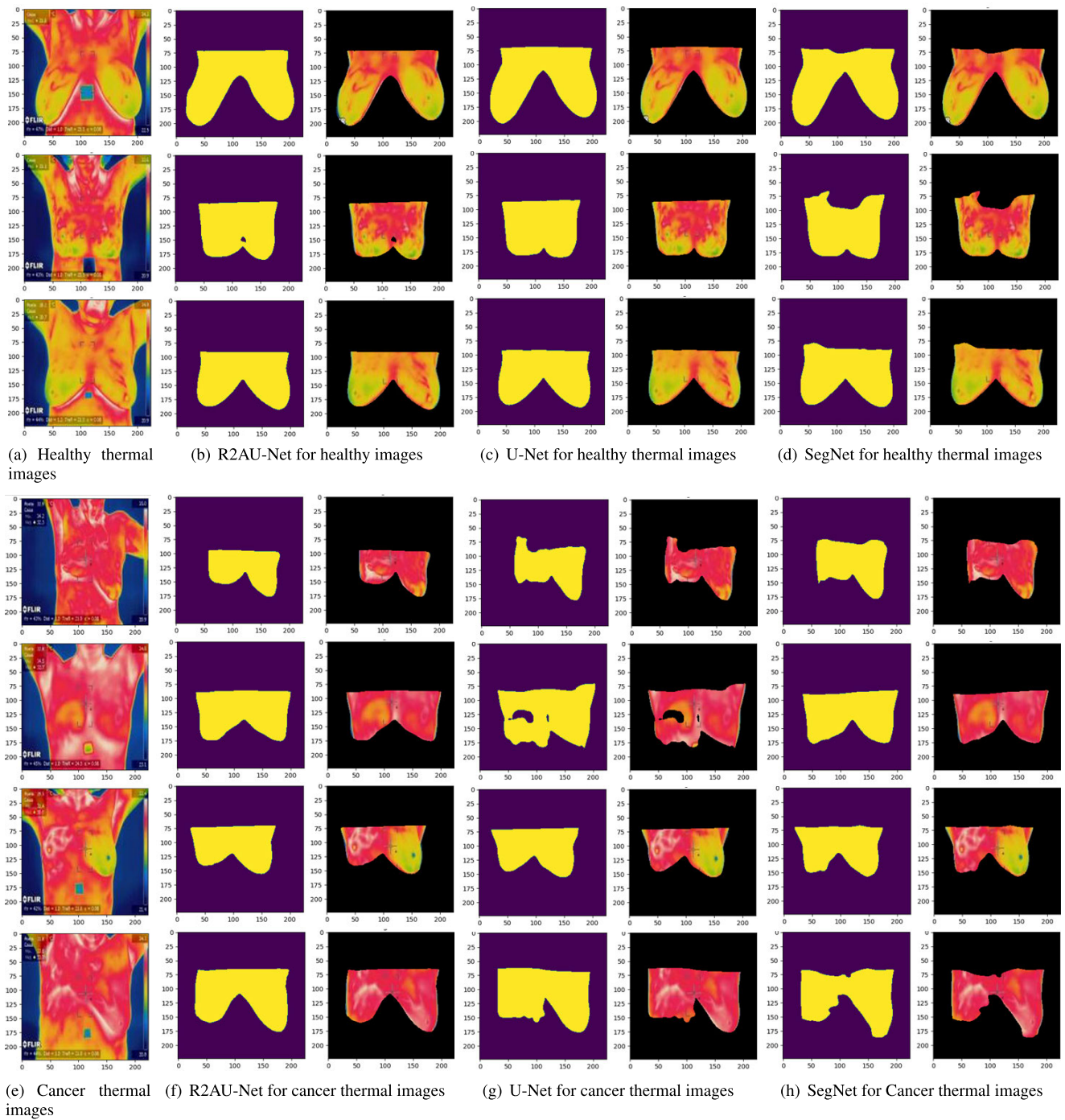
The initial stages of training across all three learning rates exhibited a common trend: a steady increase in top-1 train accuracy. This phenomenon is a typical hallmark of successful model learning, indicating that the model effectively assimilates information from the training data. As training progresses, however, we observed that the rate of increase in accuracy gradually decelerates, suggesting that the model approaches convergence.

When scrutinizing the performance of the model at different learning rates, distinct behaviors became apparent. The learning rate of  $5 \times 10^{-4}$  showcased the highest top-1 train accuracy among the three rates considered. However, this apparent advantage was accompanied by notable

volatility in the learning curve. Fluctuations in accuracy were conspicuous throughout the training process, indicating that the model's optimization might be overly aggressive. Such volatility is indicative of the learning rate overshooting the minimum loss value during gradient descent, which can impede the model's ability to converge to an optimal solution.

In contrast, the learning rate of  $2 \times 10^{-4}$  demonstrated a smoother trajectory in terms of accuracy progression. Despite not reaching the pinnacle of accuracy achieved by the  $5 \times 10^{-5}$  rate, this learning rate exhibited comparable performance while maintaining a more stable training process. The smoother increase in accuracy suggests that this learning rate strikes a balance between rapid learning and stability, potentially leading to more reliable convergence.

Conversely, the learning rate of  $10^{-4}$  displayed the slowest ascent in top-1 train accuracy among the three rates. Despite the prolonged training duration, the attained accuracy remained the lowest. This sluggish progress suggests that the learning rate might be too conservative, causing the model to learn at a leisurely pace. Consequently, the model may struggle to capture intricate patterns in the data, hindering its ability to achieve higher levels of accuracy.



**FIGURE 12.** Segmentation results of healthy and cancer thermal images using R2AU-Net, U-Net, and SegNet.

In conclusion, our analysis underscores the critical role of learning rates in shaping the convergence and performance of deep learning models. While higher learning rates may offer expedited learning, they also pose challenges related to stability and convergence. Conversely, lower learning rates prioritize stability but may compromise learning efficiency. Through meticulous selection and fine-tuning of learning rates, practitioners can navigate this delicate trade-off to optimize the performance of deep learning models.

**B. METRIC EVALUATION**

R2AU-Net has a good ability to segment breast parts by looking at the model accuracy (ACC), Sensitivity (SE), Specification (SP), Precision (PC), F1, Jaccard Similarity (JS), Dice Similarity (DS). In Table 1, we provided the training and testing results of the segmentation model of R2AU-Net and the comparison with 2 other segmentation models of SegNet and U-Net. From the testing results, it can be seen that R2AU-Net excels in training and testing for all

metrics. For training and testing results, the commonly seen dice similarity and Jaccard similarity values for segmentation models resulted in by R2AU-Net with values for training JS 98.31, DS 99.14, and testing JS 97.92, DC 98.89.

In Figure 12, we visualized three healthy images and four cancer images that were segmented using R2AU-Net, conventional U-Net, and SegNet. We chose those images as we found out that they were hardly segmented by the model resulted in SegNet dan U-Net. The segmentation results show that the conventional U-Net model and the Segnet model have difficulty in segmenting cancer images, where many parts of the breast that were not successfully segmented properly. Even the SegNet segmentation model struggles to accurately segment some healthy images. But not for the R2AU-Net model, this model manages to segment all cancer and healthy images well. Thus, the R2AU-Net model is good to be implemented for breast segmentation which will later be combined with SimCLR to classify breast cancer thermal images.

Moreover, We evaluated the proposed classification model performance of SimCLR for top-1 and top-5 learning accuracy. In Table 2, we recorded the top-1 and top-5 accuracy learning of 300 and 1000 epochs. The results show that the classifier model fed with R2AU-Net segmentation is better than the model without segmentation. For top-1 accuracy with 300 and 1000 epochs, the proposed bi-pipeline model achieves 23% and 5.8% accuracy higher compared to the one without a segmentation network, respectively. Our work shows an increase in the performance of the SimCLR model when combined with R2AU-Net for breast cancer image classification.

### C. PROPOSED BI-PIPELINE PROPERTIES

Finally, we recorded the properties of our proposed self-supervised bi-pipeline model as follows.

#### 1) PROPOSED SEGMENTATION MODEL PROPERTIES

- Total parameters: 39,442,925
- Trainable parameters: 39,442,925
- Non-trainable parameters: 0
- Total multiply-adds (G): 150.99
- Input size (MB): 0.60
- Forward/backward pass size (MB): 1859.18
- Model size (MB): 157.77
- Estimated Total Size (MB): 2017.56

#### 2) PROPOSED CLASSIFICATION MODEL PROPERTIES

- Total parameters: 24,622,784
- Trainable parameters: 24,622,784
- Non-trainable parameters: 0
- Input size (MB): 0.57
- Forward/backward pass size (MB): 286.56
- Parameters size (MB): 93.93
- Estimated Total Size (MB): 381.06

The segmentation model required 39,442,925 parameters with a parameter size of 2017.56 MB. At the same time, the

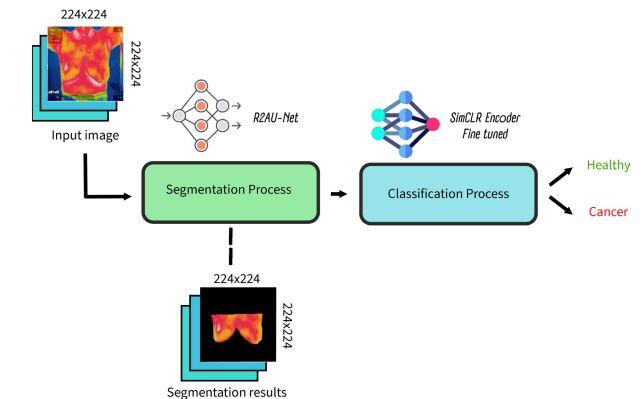


FIGURE 13. Bi-Pipeline implementation scheme.

total learning parameter of the classifier model is 24,622,728, with a model size of 381 MB. This model costs more compared to BreaCNet [9], which only has 6.1 million parameters and 22 MB in size. The limitations of the proposed method, which are computational cost and large size, make it difficult to embed in mobile devices.

### D. POTENTIAL PRACTICAL IMPLEMENTATION OF PROPOSED BI-PIPELINE SYSTEM

In Figure 13, we visualized the proposed bi-pipeline end-to-end process and some results in Table 3. As this finding is intended to be used by medical experts, We used CPU instead of GPU (Graphical Processing Unit) for the performance evaluation, with the specification of AMD FX-9830P and 8GB of RAM. Four images were fed into the proposed bi-pipeline model, and the interpretation results were recorded along with the simulation time.

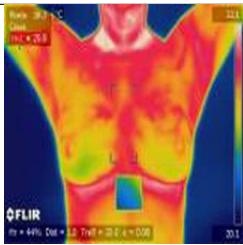
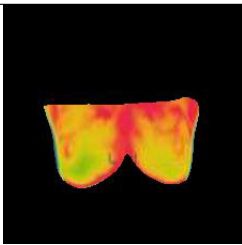

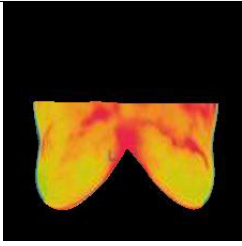

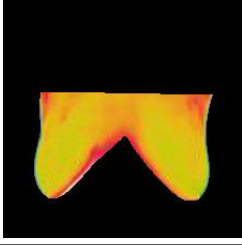
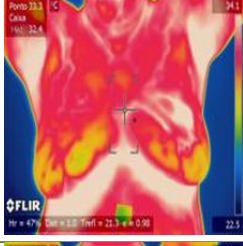
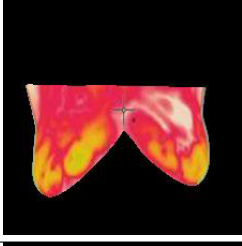
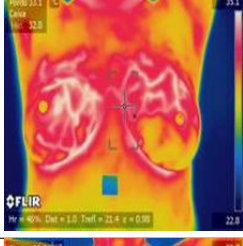
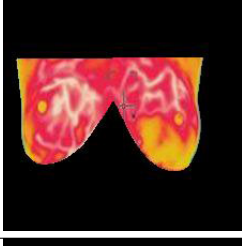

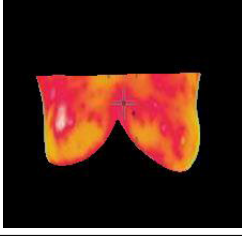
The bi-pipeline model can take the RoI of the input images and classify them all correctly. The segmentation process takes more time, with an average time of 3.508 seconds, but less in the classification process, with only 0.531 seconds on average. The implementation result confirmed that the model can be implemented as a real-time application.

On the other hand, a self-supervised learning model can lead to more robust and generalizable models. Since this model was trained on a broader spectrum of data, it tends to develop a more comprehensive understanding of the underlying patterns, making them less prone to overfitting. This aspect is particularly crucial in medical image recognition, where the model needs to perform well across diverse and unpredictable real-world breast thermal image datasets. With its robust and large size, the proposed bi-pipeline model should be implemented in a website application for medical experts.

### V. COMPARISON WITH SIMILAR WORKS

We have investigated similar work in the last ten years using the keywords of segmentation, classification, bi-pipeline, fully automated, deep learning, breast thermal image, thermogram, and breast cancer using scholar engine searches.

TABLE 3. The proposed bi-pipeline implementation result.

Input Thermal Images	Segmentation Result	Time for Segmentation	Classification Result	Ground Truth	Time for Classification
		3.22 second	Healthy	Healthy	0.56 second
		3.39 second	Healthy	Healthy	0.49 second
		3.74 second	Healthy	Healthy	0.52 second
		3.63 second	Cancer	Cancer	0.49 second
		3.52 second	Cancer	Cancer	0.5 second
		3.55 second	Cancer	Cancer	0.63 second

We found one similar work by Mohamed et al. [20] as shown in Table 4. They proposed a fully automated breast cancer

detection using U-Net for segmentation and a two-class deep learning classification. Mohammed et al. proposed a

**TABLE 4. The comparison of proposed model with similar work.**

Works	Dataset	Segmentation Model	Classification Model	Evaluation metrics	Justification
Fully Automated by Mohammed et al. [20]	Labeled dataset 1000 thermal images	U-Net	CNN with six convolutional layers and two fully-connected layers with 1024 neurons	Accuracy 99.33% Sensitivity 100% Specificity 98.67%	the model shows high accuracy rate with supervised learning
<b>Proposed Bi-Pipeline</b>	Unlabeled dataset 600 thermal images	R2AU-Net	SimCLR + ResNet50	Top-1 accuracy 84.37% Top-5 accuracy 96.87%	the model trained with limited dataset and self-supervised learning but can achieve high accuracy rate

classification model using CNN that was built from scratch. With only 60 epochs, their model can achieve an accuracy rate of 99.33% for two classes. From the learning curve, their work exhibits a stable and high accuracy rate of the classifier, but the accuracy rate of the segmentation should have been discussed. Nevertheless, the properties of the proposed model were not exposed.

## VI. CONCLUSION

In this paper, we propose a self-supervised bi-pipeline deep learning model to automatically recognize breast thermal images. The model consists of a deeply trained and fine-tuned hyperparameter of R2AU-Net for the segmentation task and SimCLR connected to ResNet50 for the classification task. All networks were trained on unlabelled breast thermal images to obtain a set of weights that enable the extraction of the region of interest (RoI) of the breast area and classify the images into healthy or cancer categories. By applying transfer learning, the model achieved high accuracy rates in both segmentation and classification tasks, with 98.63% and 96.87%, respectively. Moreover, the proposed bi-pipeline model exhibits an effective use with only around 4 seconds to proceed with segmentation and classification tasks. These findings confirmed that the model can be implemented as a real-time application. Despite the high accuracy rate, the classifier model requires further fine-tuning of hyperparameters to ensure more stable learning.

## REFERENCES

- [1] S. J. Mambou, P. Maresova, O. Krejcar, A. Selamat, and K. Kuca, "Breast cancer detection using infrared thermal imaging and a deep learning model," *Sensors*, vol. 18, no. 9, p. 2799, Aug. 2018.
- [2] M. Abdel-Nasser, A. Moreno, and D. Puig, "Breast cancer detection in thermal infrared images using representation learning and texture analysis methods," *Electronics*, vol. 8, no. 1, p. 100, Jan. 2019.
- [3] F. J. Fernández-Ovies, E. S. Alférez-Baquero, E. J. de Andrés-Galiana, A. Cernea, Z. Fernández-Muñiz, and J. L. Fernández-Martínez, "Detection of breast cancer using infrared thermography and deep neural networks," in *Proc. Int. Work-Confer. Bioinf. Biomed. Eng.* Cham, Switzerland: Springer, 2019, pp. 514–523.
- [4] R. Roslidar, K. Saddami, F. Arnia, M. Syukri, and K. Munadi, "A study of fine-tuning CNN models based on thermal imaging for breast cancer classification," in *Proc. IEEE Int. Conf. Cybern. Comput. Intell. (CyberneticsCom)*, Aug. 2019, pp. 77–81.
- [5] A. Lou, S. Guan, N. Kamona, and M. Loew, "Segmentation of infrared breast images using MultiResUnet neural networks," in *Proc. IEEE Appl. Imag. Pattern Recognit. Workshop (AIPR)*, Oct. 2019, pp. 1–6.
- [6] R. Roslidar, K. Saddami, M. Irhamsyah, F. Arnia, M. Syukri, and K. Munadi, "Effective loss function for unbalanced breast thermal image segmentation," in *Proc. Int. Conf. Comput. Syst., Inf. Technol., Electr. Eng. (COSITE)*, Oct. 2021, pp. 107–111.
- [7] S. Pramanik, D. Banik, D. Bhattacharjee, M. Nasipuri, M. K. Bhowmik, and G. Majumdar, "Suspicious-region segmentation from breast thermogram using DLPE-based level set method," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 572–584, Feb. 2019.
- [8] D. Sánchez-Ruiz, I. Olmos-Pineda, and J. A. Olvera-López, "Automatic region of interest segmentation for breast thermogram image classification," *Pattern Recognit. Lett.*, vol. 135, pp. 72–81, Jul. 2020.
- [9] R. Roslidar, M. Syaryadhi, K. Saddami, B. Pradhan, F. Arnia, M. Syukri, and K. Munadi, "BreacNet: A high-accuracy breast thermogram classifier based on mobile convolutional neural network," *Math. Biosci. Eng.*, vol. 19, no. 2, pp. 1304–1331, 2021.
- [10] Q. Zuo, S. Chen, and Z. Wang, "R2AU-Net: Attention recurrent residual convolutional neural network for multimodal medical image segmentation," *Secur. Commun. Netw.*, vol. 2021, pp. 1–10, Jun. 2021.
- [11] T. Chen, S. Kornblith, M. Norouzi, and G. E. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. 37th Int. Conf. Mach. Learn.*, vol. 119, 2020, pp. 1597–1607.
- [12] R. Roslidar, A. Rahman, R. Muharar, M. R. Syahputra, F. Arnia, M. Syukri, B. Pradhan, and K. Munadi, "A review on recent progress in thermal imaging and deep learning approaches for breast cancer detection," *IEEE Access*, vol. 8, pp. 116176–116194, 2020.
- [13] M. de Freitas Oliveira Baffa and L. G. Lattari, "Convolutional neural networks for static and dynamic breast infrared imaging classification," in *Proc. 31st SIBGRAPI Conf. Graph., Patterns Images (SIBGRAPI)*, Oct. 2018, pp. 174–181.
- [14] R. R. Devi and G. S. Anandhamala, "Analysis of breast thermograms using asymmetry in infra-mammary curves," *J. Med. Syst.*, vol. 43, no. 6, pp. 1–9, Apr. 2019.
- [15] R. Sánchez-Cauce, J. Pérez-Martín, and M. Luque, "Multi-input convolutional neural network for breast cancer detection using thermal images and clinical data," *Comput. Methods Programs Biomed.*, vol. 204, Jun. 2021, Art. no. 106045.
- [16] M. J. Mammoottil, L. J. Kulangara, A. S. Cherian, P. Mohandas, K. Hasikin, and M. Mahmud, "Detection of breast cancer from five-view thermal images using convolutional neural networks," *J. Healthcare Eng.*, vol. 2022, pp. 1–15, Feb. 2022.
- [17] C. B. Gonçalves, J. R. Souza, and H. Fernandes, "CNN architecture optimization using bio-inspired algorithms for breast cancer detection in infrared images," *Comput. Biol. Med.*, vol. 142, Mar. 2022, Art. no. 105205.
- [18] M. A. S. Al Husaini, M. H. Habaebi, T. S. Gunawan, M. R. Islam, E. A. A. Elsheikh, and F. M. Suliman, "Thermal-based early breast cancer detection using inception v3, inception V4 and modified inception MV4," *Neural Comput. Appl.*, vol. 34, no. 1, pp. 333–348, Jan. 2022.
- [19] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351. Cham, Switzerland: Springer, 2015, pp. 234–241.
- [20] E. A. Mohamed, E. A. Rashed, T. Gaber, and O. Karam, "Deep learning model for fully automated breast cancer detection system from thermograms," *PLoS ONE*, vol. 17, no. 1, Jan. 2022, Art. no. e0262349.

- [21] A. Iqbal and M. Sharif, "PDF-UNet: A semi-supervised method for segmentation of breast tumor images using a U-shaped pyramid-dilated network," *Expert Syst. Appl.*, vol. 221, Jul. 2023, Art. no. 119718.
- [22] P. Gomathi, C. Muniraj, and P. S. Periasamy, "Digital infrared thermal imaging system based breast cancer diagnosis using 4D U-Net segmentation," *Biomed. Signal Process. Control*, vol. 85, Aug. 2023, Art. no. 104792.
- [23] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*.
- [24] J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, and D. Rueckert, "Attention gated networks: Learning to leverage salient regions in medical images," *Med. Image Anal.*, vol. 53, pp. 197–207, Apr. 2019.
- [25] M. Liang and X. Hu, "Recurrent convolutional neural network for object recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3367–3375.
- [26] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation," 2018, *arXiv:1802.06955*.
- [27] L. Jing and Y. Tian, "Self-supervised visual feature learning with deep neural networks: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 11, pp. 4037–4058, Nov. 2021.
- [28] Y. Roh, G. Heo, and S. E. Whang, "A survey on data collection for machine learning: A big data–AI integration perspective," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 4, pp. 1328–1347, Apr. 2021.
- [29] E. Chinn, R. Arora, R. Arnaout, and R. Arnaout, "ENRICHing medical imaging training sets enables more efficient machine learning," *J. Amer. Med. Inform. Assoc.*, vol. 30, no. 6, pp. 1079–1090, May 2023.
- [30] A. Jaiswal, A. R. Babu, M. Z. Zadeh, D. Banerjee, and F. Makedon, "A survey on contrastive self-supervised learning," *Technologies*, vol. 9, no. 1, p. 2, Dec. 2020.
- [31] S. Shurrab and R. Duwairi, "Self-supervised learning methods and applications in medical imaging analysis: A survey," *PeerJ Comput. Sci.*, vol. 8, p. e1045, Jul. 2022.
- [32] A. Kaku, S. Upadhyay, and N. Razavian, "Intermediate layers matter in momentum contrastive self supervised learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 24063–24074.
- [33] J. Zbontar, L. Jing, I. Misra, Y. LeCun, and S. Deny, "Barlow twins: Self-supervised learning via redundancy reduction," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 12310–12320.
- [34] M. Caron, I. Misra, J. Mairal, P. Goyal, P. Bojanowski, and A. Joulin, "Unsupervised learning of visual features by contrasting cluster assignments," in *Proc. NIPS*, Dec. 2020, pp. 9912–9924.
- [35] H. Wu, Y. Qu, S. Lin, J. Zhou, R. Qiao, Z. Zhang, Y. Xie, and L. Ma, "Contrastive learning for compact single image dehazing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10546–10555.
- [36] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 9726–9735.
- [37] M. Caron, H. Touvron, I. Misra, H. Jegou, J. Mairal, P. Bojanowski, and A. Joulin, "Emerging properties in self-supervised vision transformers," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9630–9640.
- [38] J. D. Miller, V. A. Arasu, A. X. Pu, L. R. Margolies, W. Sieh, and L. Shen, "Self-supervised deep learning to enhance breast cancer detection on screening mammography," 2022, *arXiv:2203.08812*.
- [39] P. C. Chhipa, R. Upadhyay, G. G. Pihlgren, R. Saini, S. Uchida, and M. Liwicki, "Magnification prior: A self-supervised method for learning representations on breast cancer histopathological images," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2023, pp. 2716–2726.
- [40] N. Saidnassim, B. Abdikenov, R. Kelesbekov, M. T. Akhtar, and P. Jamwal, "Self-supervised visual transformers for breast cancer diagnosis," in *Proc. Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC)*, Dec. 2021, pp. 423–427.
- [41] L. F. Silva, D. C. M. Saade, G. O. Sequeiros, A. C. Silva, A. C. Paiva, R. S. Bravo, and A. Conci, "A new database for breast research with infrared image," *J. Med. Imag. Health Informat.*, vol. 4, no. 1, pp. 92–100, Mar. 2014.
- [42] K. Sohn, "Improved deep metric learning with multi-class N-pair loss objective," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 1–7.
- [43] A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool," *BMC Med. Imag.*, vol. 15, no. 1, pp. 1–28, Dec. 2015.



**ROSLIDAR ROSLIDAR** (Member, IEEE) received the bachelor's degree in electrical engineering from Universitas Syiah Kuala, in 2002, the master's degree in telecommunication engineering from the University of Arkansas, USA, in 2009, and the Ph.D. degree, in January 2022. Since 2003, she has been a Lecturer with Universitas Syiah Kuala. She has been actively doing research in science and technology. She was a Visiting Scholar with the Faculty of System Design, Tokyo Metropolitan University, Japan, in May 2014. She dedicated her time to develop the e-health monitoring system based on a non-invasive technique. She was granted a Fulbright Scholarship for the master's degree from the University of Arkansas.



**MUHAMMAD JUREJ ALHAMDI** received the bachelor's degree in electrical engineering from the Department of Electrical and Computer Engineering, Faculty of Engineering, Universitas Syiah Kuala, Banda Aceh, Indonesia, in 2022, where he is currently pursuing the master's degree in electrical engineering. He was a Research Assistant with the Control System Laboratory, from 2020 to 2021. His research interests include designing and developing autonomous car using deep learning and application of deep learning in healthcare.



**AULIA RAHMAN** (Member, IEEE) received the bachelor's degree from the Institut Teknologi Bandung (ITB), Indonesia, in 2005, and the master's degree in automation and robotics from Technische Universität Dortmund (TU Dortmund), Germany, in 2011. He is currently pursuing the Ph.D. degree with Universiti Malaya, Malaysia. He has been with the Department of Electrical Engineering, Universitas Syiah Kuala, since 2012. He affiliated with the Control System Laboratory and the Embedded System Laboratory, Universitas Syiah Kuala. His research interests include mobile robotics, SLAM, control systems, computer vision, deep learning, and visual localization.





**KHAIRUN SADDAMI** (Member, IEEE) received the bachelor's degree and the Ph.D. degree in electrical and computer engineering from Universitas Syiah Kuala, Indonesia, in 2015 and 2020, respectively. He has been with the Department of Electrical and Computer Engineering, Universitas Syiah Kuala, since 2020. His research interests include document image analysis, deep learning, biometric and biomedical image processing, and pattern recognition. He is a member of the

International Association for Pattern Recognition (IAPR). He was awarded a scholarship from the Ministry of Research, Technology, and Higher Education, Indonesia, under the Scheme of Pendidikan Magister Menuju Doktor Untuk Sarjana Unggul (PMDSU) from Universitas Syiah Kuala. He also acts as a reviewer of reputable journals, such as *Nature*, *Scientific Reports*, *IEEE ACCESS*, and *ACM Computing Surveys*.



**MAIMUN SYUKRI** received the Medical degree from the University of Airlangga, Surabaya, Indonesia, in 1988, and the Ph.D. degree in medicine from Gadjah Mada University, Yogyakarta, Indonesia, in 2014. From 2009 to 2016, he was the Head of the Internal Medicine Department, Medical Faculty, Universitas Syiah Kuala, Banda Aceh, Indonesia. He is currently a Lecturer with the Medical Faculty, Universitas Syiah Kuala. He is training as an internist with the University of Airlangga, and subsequently trained in nephrology. He is also a member of Indonesian Medical Association, Indonesian Society of Nephrology, Indonesian Society of Internal Medicine, Indonesian Society of Hypertension, Asian Pacific Society of Nephrology, and the International Society of Nephrology.



**FITRI ARNIA** (Member, IEEE) received the B.Eng. degree from Universitas Sumatera Utara (USU), Medan, in 1997, the master's degree from the University of New South Wales (UNSW), Sydney, Australia, in 2004, and the Ph.D. degree from Tokyo Metropolitan University (TMU), Tokyo, Japan, in 2008. She has been with the Department of Electrical and Computer Engineering, Faculty of Engineering, Universitas Syiah Kuala (Unsyiah), since 1999. In November 2019, she

becomes a Professor with Unsyiah. She was a Visiting Scholar with TMU, in 2013, and Suleyman Demirel University (SDU), Isparta, Türkiye, in 2017. Her research interests include signal, image, and multimedia information processing. She is a member of ACM and APSIPA.



**KHAIRUL MUNADI** (Member, IEEE) received the B.Eng. degree from the Sepuluh Nopember Institute of Technology, Surabaya, Indonesia, in 1996, and the M.Eng. and Ph.D. degrees from Tokyo Metropolitan University (TMU), Japan, in 2004 and 2007, respectively, all in electrical engineering. From 1996 to 1999, he was with Alcatel Indonesia, as a System Engineer. Since 1999, he has been with the Electrical and Computer Engineering Department, Universitas Syiah Kuala (Unsyiah), Banda Aceh, Indonesia, as a Lecturer, where he has been a Professor, since August 2019. He was a Visiting Researcher of information and communication systems engineering with the Faculty of System Design, TMU, from March 2007 to March 2008. He was also a Visiting Scholar with the Department of Computer Engineering, Suleyman Demirel University (SDU), Isparta, Türkiye, in 2016. His research interests include multimedia signal processing, knowledge-based management, and disaster management. He is a member of APSIPA.

• • •