

SINGH, V.K., RASHWAN, H.A., ROMANI, S., AKRAM, F., PANDEY, N., SARKER, M.M.K., SALEH, A., ARENAS, M., ARQUEZ, M., PUIG, D. and TORRENTS-BARRENA, J. 2020. Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network. *Expert systems with applications* [online], 139, article number 112855. Available from: <https://doi.org/10.1016/j.eswa.2019.112855>

Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network.

SINGH, V.K., RASHWAN, H.A., ROMANI, S., AKRAM, F., PANDEY, N., SARKER, M.M.K., SALEH, A., ARENAS, M., ARQUEZ, M., PUIG, D. and TORRENTS-BARRENA, J.

2020

Breast Tumor Segmentation and Shape Classification in Mammograms using Generative Adversarial and Convolutional Neural Network

Vivek Kumar Singh^{a,*}, Hatem A. Rashwan^a, Santiago Romani^a, Farhan Akram^b, Nidhi Pandey^d, Md. Mostafa Kamal Sarker^a, Adel Saleh^a, Meritxell Arenas^c, Miguel Arquez^c, Domenec Puig^a, Jordina Torrents-Barrena^a

^a*Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili, Tarragona, Spain.*

^b*Imaging Informatics Division, Bioinformatics Institute, A*STAR, Singapore, Singapore.*

^c*Hospital Universitari Sant Joan, Reus, Spain.*

^d*Department of Medicine and Health Sciences, Universitat Rovira i Virgili, Tarragona, Spain.*

Abstract

Mammogram inspection in search of breast tumors is a tough assignment that radiologists must carry out frequently. Therefore, image analysis methods are needed for the detection and delineation of breast tumors, which portray crucial morphological information that will support reliable diagnosis. In this paper, we proposed a conditional Generative Adversarial Network (cGAN) devised to segment a breast tumor within a region of interest (ROI) in a mammogram. The generative network learns to recognize the tumor area and to create the binary mask that outlines it. In turn, the adversarial network learns to distinguish between real (ground truth) and synthetic segmentations, thus enforcing the generative network to create binary masks as realistic as possible. The cGAN works well even when the number of training samples are limited. As a consequence, the proposed method outperforms several state-of-the-art approaches. Our

*This work was supported by the Spanish project: DPI2016-77415-R and the Beatriu de Pinós program: 2016-BP-00063.

*Corresponding author

Email addresses: vivekkumar.singh@urv.cat (Vivek Kumar Singh), hatem.abdellatif@urv.cat (Hatem A. Rashwan), santiago.romani@urv.cat (Santiago Romani), farhana@bii.a-star.edu.sg (Farhan Akram), nidhi.pandey@estudiants.urv.cat (Nidhi Pandey), mdmostafakamal.sarker@urv.cat (Md. Mostafa Kamal Sarker), adelsalehali.alraimi@urv.cat (Adel Saleh), meritxell.arenas@gmail.com (Meritxell Arenas), miguelarquez@outlook.com (Miguel Arquez), domenec.puig@urv.cat (Domenec Puig), jordina.torrents@urv.cat (Jordina Torrents-Barrena)

working hypothesis is corroborated by diverse segmentation experiments performed on INbreast and a private in-house dataset. The proposed segmentation model, working on an image crop containing the tumor as well as a significant surrounding area of healthy tissue (loose frame ROI), provides a high Dice coefficient and Intersection over Union (IoU) of 94% and 87%, respectively. In addition, a shape descriptor based on a Convolutional Neural Network (CNN) is proposed to classify the generated masks into four tumor shapes: irregular, lobular, oval and round. The proposed shape descriptor was trained on DDSM, since it provides shape ground truth (while the other two datasets does not), yielding an overall accuracy of 80%, which outperforms the current state-of-the-art.

Keywords: Mammograms, conditional generative adversarial network, convolutional neural network, tumor segmentation and shape classification.

1. Introduction

Breast cancer is the most common diagnosed cause of death from cancer in women in the world Siegel et al. (2017). Mammography is a world recognized tool that has been proven effective to reduce the mortality rate, since it allows early detection of breast diseases Lauby-Secretan et al. (2015).

Breast masses are the most important findings among diverse types of breast abnormalities, such as micro-calcification and architectural distortion. All these findings may point out the presence of carcinomas Rangayyan et al. (2010). Moreover, morphological information of tumor shape (irregular, lobular, oval and round) and margin type (circumscribed, ill defined, spiculated and obscured) also play crucial roles in the diagnosis of tumor malignancy Tang et al. (2009).

Computer aided diagnosis (CAD) systems are highly recommended to assist radiologists in detecting breast tumors and outlining their borders. However, breast tumor segmentation and classification are still challenges due to low signal-to-noise ratio and variability of tumors in shape, size, appearance, texture and location. Recently, many studies based on deep representation of breast images and combining features have been proposed to improve performance on breast mass classification Jiao et al. (2018).

In addition, based on mammographic images, it is very complicated for an expert radiologist to discern the molecular subtypes, i.e., Luminal-A, Luminal-B, HER-2 (Human Epidermal growth factor receptor 2) and Basal-like (triple negative), which are key for prescribing the best oncological treatment Cho (2016), Liu et al. (2016a), Tamaki et al. (2011). However, recent studies point out some loose correlations between visual tumor features (e.g., texture and shape) and molecular subtypes. Recently, a Convolutional Neural Network (CNN) was used to classify molecular subtypes using texture patches extracted from mammography Singh et al. (2017), which yielded an overall accuracy of 67%. However, depending only on texture feature is not sufficient to classify the breast cancer molecular subtypes from mammograms Tamaki et al. (2011). Thus, some studies attempt to use morphological information of tumor shape in classifying breast cancer molecular subtypes.

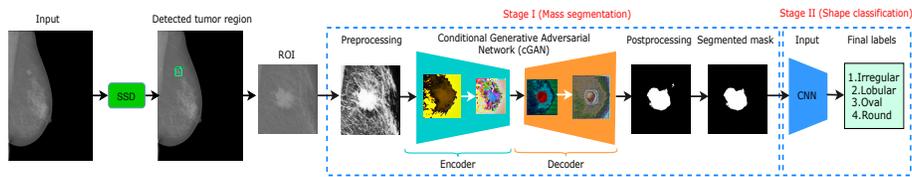


Figure 1: Automatic workflow for our breast tumor segmentation and shape classification system.

In this paper, we propose a method based on two main stages, one for breast tumor segmentation and another for tumor shape classification, as shown in Figure 1. Before applying our segmentation approach, the Single Shot Detector (SSD) Liu et al. (2016b) is used to locate the tumor and then our method computes the proper coordinates to crop the ROI. Afterwards, the first stage segments the breast tumor, contained in the ROI, as a binary mask. In the second stage, the binary mask is classified to a shape type (irregular, lobular, oval and round). Unlike traditional object classifiers Kisilev et al. (2015), Kim et al. (2018) that use texture, intensity or edge information, our method is forced to learn only morphological features from the binary masks. The current proposal is a thorough improvement of our previous work Singh et al. (2018b). The major contributions of this paper are as follows:

1. We believe this is the first adaptation of cGAN in the area of breast tumor seg-

mentation in mammograms. The adversarial network yields more reliable learning than other state-of-the-art algorithms since training data is scarce (*i.e.*, mammograms with labeled breast tumor boundaries), while it does not increase the computational complexity at prediction time.

2. The application of a multi-class CNN architecture to predict the four breast tumor shapes (*i.e.*, irregular, lobular, oval and round) using the binary mask segmented in the previous stage (cGAN output).
3. An in-depth evaluation of our system’s performance using two public (1,274 images) and one private (300 images) databases. The obtained results outperform current state-of-the-art in both tumor segmentation and shape classification.
4. A study of the correlation between the tumor shape predicted by our automatic method with respect to the ground-truth molecular subtypes of breast cancer, which reasonably matches with other clinical studies like Boisserie-Lacroix et al. (2013).

This paper is organized as follows. Section II provides the related work of both tumor segmentation and shape classification. The proposed architectures for tumor segmentation (using cGAN) and shape classification (using CNN) are described in Section III. In Section IV, extensive experiments are performed on the two stages of the proposed method and the obtained results are compared with the state-of-the-art results. In addition, the limitations of the proposed models are explained in Section IV. Finally, Section V concludes our work and suggests some future lines of research.

2. Related Work

In the following paragraphs we point out some works mainly focused on breast tumor segmentation and shape classification in mammography, as well as generic image analysis methods highly related with our field of interest.

2.1. Tumor Segmentation Background

Convolutional Neural Networks (CNNs) can automatically learn features from the given images to represent objects at different scales and orientations. By increasing the

70 number of layers (depth of CNN model) more detailed features can be obtained, which play crucial part in solving different computer vision problems, such as object detection, classification and segmentation. Thus, numerous methods has been proposed to solve the image segmentation problem based on deep learning approaches Schmidhuber (2015).

75 One of the well-known architectures for semantic segmentation is the Fully Convolutional Network (FCN) Long et al. (2015), which is based on encoding (convolutional) and decoding (deconvolutional) layers. This approach gets rid of the fully connected layers of CNNs to convert the image classification networks into image filtering networks. An improvement of this scheme was proposed by the U-Net architecture Ronneberger et al. (2015), where skip connections between encoding and decoding layers are added to retain significant information from the input features. Later on, a new variation of FCN was proposed Badrinarayanan et al. (2017) named SegNet, which consists of hierarchy of decoders, each one corresponding to each encoder. The decoder network uses the max-pooling indices received from the corresponding encoder
80 to perform non-linear upsampling of their input feature maps.
85

Since semantic segmentation has achieved great progress with deep learning, there is recent popularity in applying such models to medical imaging, such as for skin lesions segmentation (Litjens et al. (2017), Sarker et al. (2018)), and for fundus photography of the rear of an eye (Fu et al. (2018), Singh et al. (2018a)).

90 For breast tumor detection, segmentation and classification, many medical image analysis methods have been proposed so far, such as Yassin et al. (2018) and Hamidinekoo et al. (2018).

A tumor classification and segmentation method was proposed Rouhi et al. (2015) using an automated region growing algorithm whose threshold was obtained by a trained Artificial Neural Network (ANN) and Cellular Neural Network (CeNN). In
95 turn, to reduce the computational complexity and increase the robustness, a quantized and non-linear CeNN for breast tumor segmentation was proposed in Liu et al. (2018). After segmenting the breast tumor region, a Multilayer Perceptron Classifier was used for tumor classification as benign or malignant.

100 Furthermore, Dhungel et al. (2015b) segmented breast tumors using Structured

Support Vector Machines (SSVM) and Conditional Random Fields (CRF). Both graphical models minimize a loss function build on pixel probabilities provided by a CNN and Deep Belief Network, a Gaussian Mixture Model (GMM) and shape prior. The SSVM is based on graph cuts and the CRF relies on tree re-weighted belief propagation with truncated fitting training Dhungel et al. (2015a). Cardoso et al. (2015, 2017) tackled the same problem by employing a closed contour fitting in the mammogram and minimizing a cost function depending on the radial derivative of the tumor contour. A measure of regularity of the gray pixel values inside and outside the tumor was also included in Cardoso et al. (2017).

In turn, Zhu et al. (2018) proposed an FCN concatenated to a CRF layer to impose the compactness of the segmentation output taking into account pixel position. This approach was trained end-to-end, since the CRF and FCN can exchange data in the forward-backward propagation. An adversarial term was introduced to prevent the samples with the worst perturbation in the loss function, which reduced the overfitting and provided a robust learning with few training samples. In addition, Al-antari et al. (2018) proposed a CAD system consisting of three deep learning stages for detecting, segmenting and classifying the tumors in mammographic images. To locate tumors in a full mammogram, the YOLO network proposed in Redmon et al. (2016) was used. A Full resolution Convolutional Network (FrCN) was then used for segmenting the located tumor region. Finally, a CNN network was used for classifying segmented ROI as either benign or malignant.

We believe that Yang et al. (2017) is the first work that exploits GAN Goodfellow et al. (2014) for medical image segmentation. In particular, they performed three-dimensional (3D) liver segmentations using abdominal Computerized Tomography (CT) scans. In Singh et al. (2018b), we adapted a cGAN image-to-image translation algorithm Isola et al. (2017) to address the tumor segmentation in two-dimensional (2D) mammograms. With that method, we achieved state-of-the-art performance on both public and private databases.

2.2. Shape Classification Background

130 In the literature, many approaches used traditional computer vision techniques to extract hand-craft features and subsequently classify them. For instance, Matos et al. (2018) applied the Bag of Features (BoF) paradigm on local feature descriptors, such as Scale-Invariant Feature Transform (SIFT), Speed Up Robust Feature (SURF) or Local Binary Patterns (LBP), achieving very high accuracy of 99% in classifying tumors as
135 malignant or benign.

Recently, deep learning architectures have been designed for 2D and 3D shape classification Kurnianggoro et al. (2018). For example, topological data analysis (TDA) using deep learning was proposed in Hofer et al. (2017) to extract relevant 2D/3D topological and geometrical information. In turn, a CNN model was formulated, which
140 used spectral graph wavelets in conjunction with the BoF paradigm to target the shape classification problem Masoumi and Hamza (2017).

In addition, the authors in Fang et al. (2015) proposed a CNN based shape descriptor for retrieving the 3D shapes. A deep neural network named PointNet was proposed Qi et al. (2017), which directly consumes point cloud for object classification, localized
145 and global semantic segmentation. Moreover, a deep learning framework for efficient 3D shape classification Luciano and Hamza (2018) used geodesic moments by inheriting various properties from the geodesic distance, such as the intrinsic geometric structure of 3D shapes and the invariance to isometric deformations.

To date, numerous shape classification methods are applied for medical image analysis Singh et al. (2018b), and Kim et al. (2018). An automated method for textual
150 description of anatomical breast tumor lesions was proposed by Kisilev et al. (2015), which performs joint semantic estimation from image measurements to classify the tumor shape. In addition, Kisilev et al. (2016) also presented a multi-task fast region-based CNN Ren et al. (2015) to classify three tumor shapes: irregular, oval and round. Furthermore, the work in Kim et al. (2018) utilized a GAN to diagnose and classify tumors in mammograms into four shapes: irregular, lobular, oval and round. Previously,
155 Singh et al. (2018b) proposed a multi-class CNN to categorize the tumor shapes into

four classes as in Kim et al. (2018) from the public dataset DDSM¹.

3. Proposed Methodology

160 The proposed CAD system shown in Fig. 1 is divided into two stages: breast tumor segmentation and shape classification.

3.1. Obtaining and Processing ROIs

Before feeding an image to the first stage, our optimal workflow applies the Single Shot Detector (SSD) Liu et al. (2016b) to locate the tumor position and fit a bounding box around it. Based on these bounding coordinates, our method computes new coordinates containing the tumor (*vide infra*), and then uses these new coordinates to crop the mammogram, thus obtaining the Region of Interest (ROI). We evaluated different detectors based on deep learning models, such as SSD Liu et al. (2016b), YOLO Redmon et al. (2016) and Faster R-CNN Ren et al. (2015). Empirically, the SSD detector yields the best results since it is able to detect small tumor regions and provides an overall accuracy of 97%. We are not targeting object sizes less than 7×7 pixels because those objects are really hard to be identified as tumors. Indeed, they may correspond to other types of findings, such as calcifications. We are considering only mammograms with tumors, since our main goal is tumor shape classification following tumor segmentation. Therefore, we have not applied SSD on normal mammograms (no tumor), although the SSD method is capable of dealing with this case as well.

To obtain the proper cropping area, our best framing method, so-called "loose frame", expands the original bounding box coordinates by adding extra space around, so that the cropped ROI always encompasses the tumor as well as some surrounding area containing healthy tissue (30% and 70% for tumor and healthy tissues, respectively). The computed coordinates are shifted to make the ROI frame fit inside the mammogram image. Besides, both sides of the frame are set equal in order to preserve the original aspect ratio of tumors. Last adjustments required to make the image square sometimes cause the tumor be out of the ROI center. However, this does not

¹<https://wiki.cancerimagingarchive.net/display/Public/CBIS-DDSM>

185 preclude the segmentation and classification due to the position-independent nature of convolutional filters.

Moreover, ROI images are scaled to 256×256 pixels, which is the optimal cGAN input size found experimentally. After scaling, they are pre-processed for noise removal as proposed in Kshema et al. (2017) (Gaussian filter with $\sigma = 0.5$ yields the
190 best segmentation results) and then contrast is enhanced using histogram equalization, similarly to Cheng et al. (2003). Then, we apply a normalization for rescaling the pixel values between $[0..1]$.

The prepared data is then fed to the cGAN to obtain a binary mask of the breast tumor, which is post-processed using morphological operations (we used filter sizes of
195 3×3 for closing, 2×2 for erosion, and 3×3 for dilation) to remove small speckles, as proposed in Hazarika and Mahanta (2018). Fig. 2 shows a couple of examples of these small speckles, enclosed in red boxes, which are filtered out after post-processing.

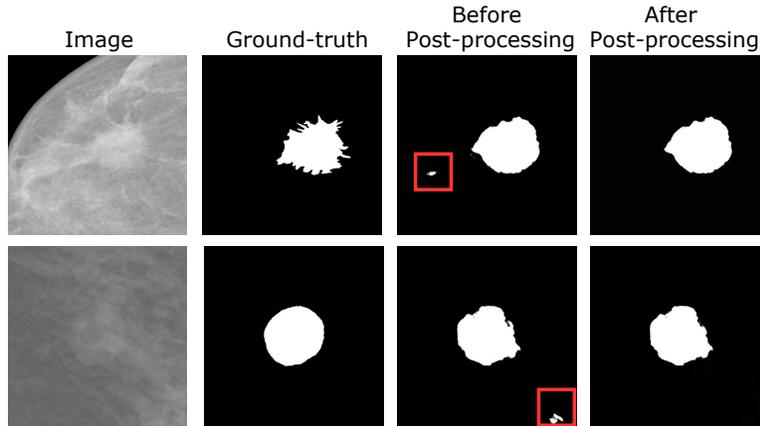


Figure 2: Two examples of the effect of morphological post processing after the segmentation.

In the second stage, the output binary mask is downsampled into 64×64 pixels, which is then fed to a multi-class CNN shape descriptor to categorize it into four
200 classes: irregular, lobular, oval and round. The reason of this downsampling is that our shape classification CNN does not need a high resolution image to extract the core morphological features for each class, since the tumors are represented with flat white areas in front of a black background. Hence, the changes in the image present very low

frequencies.

205 **3.2. Tumor Segmentation Model (cGAN)**

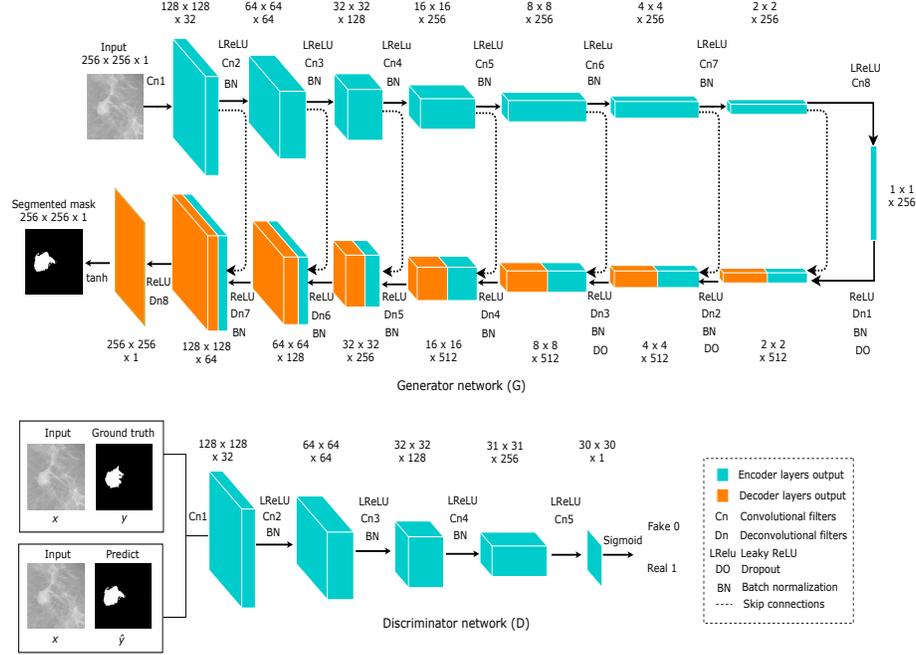


Figure 3: Proposed cGAN architecture: generator G (top), and discriminator D (down).

Our previous work Singh et al. (2018b) demonstrated the feasibility of applying the cGAN image-to-image translation approach Isola et al. (2017) to breast tumor segmentation, since it can be adapted to our problem in the following senses:

1. The Generator G network of the cGAN is an FCN composed of encoding and decoding layers, which learn the intrinsic features (gray-level, texture, gradients, edges, shape, etc.) of healthy and unhealthy (tumor) breast tissue, and generate a binary mask according to these features.
2. The Discriminative D network of the cGAN assesses if a given binary mask is likely to be a realistic segmentation or not. Therefore, including the adversarial score in the computation of the generator loss strengthens its capability to provide a correct segmentation.

The combination of G and D networks allows robust learning with few training samples. Since the ROI image is a conditioning input for both G and D , the segmentation result is better fitted to the tumor appearance. Otherwise, regular (unconditional) GAN Goodfellow et al. (2014) will infer the segmentation just from random noise, which will require more training iterations compared to the cGAN to obtain an acceptable segmentation result.

Fig. 3 represents the suggested architectures for G and D . The former consists of several encoding and decoding layers (see Fig. 3-top). Encoding layers are composed of a set of convolutional filters followed by batch normalization and the leaky ReLU (slope 0.2) activation function. Similarly, decoding layers are composed of a set of deconvolutional filters followed by batch normalization, dropout and ReLU.

Convolutional and deconvolutional filters are defined with a kernel of 4×4 and stride of 2×2 , which respectively downsample and upsample the activation maps by a factor of 2. Batch normalization is not applied after the first and the last convolutional filters (Cn_1 and Cn_8). After Cn_8 , the ReLU activation function is applied instead of leaky ReLU. Dropout is applied only at the first three decoding layers (Dn_1 , Dn_2 and Dn_3). There is no skip connection in the last decoding layer (Dn_8), after which the \tanh activation function is applied to generate a binary mask of the breast tumor.

The architecture of D shown in Fig. 3-down consists of five encoding layers with convolutional filters with a kernel of 4×4 , stride 2×2 at the first three layers and stride 1×1 at 4^{th} and 5^{th} layers. Batch normalization is applied after Cn_2 , Cn_3 and Cn_4 and a leaky ReLU (slope 0.2) is applied after each layer except for the last one. The sigmoid activation function is used after the last convolutional filter (Cn_5). The network input is the concatenation of the ROI and the binary mask to be evaluated (ground truth or predicted). The output segmentation is an array of 30×30 values, each one from 0.0 (completely fake) to 1.0 (perfectly plausible or real). Each output value is the degree of proper segmentation likelihood of a crop of the binary mask and the input image, which corresponds to a 70×70 receptive field for each value.

Let x be a tumor ROI, y the ground truth mask, z a random variable, λ an empirical weighting factor, $G(x, z)$ and $D(x, G(x, z))$ the outputs of G and D , respectively. Then,

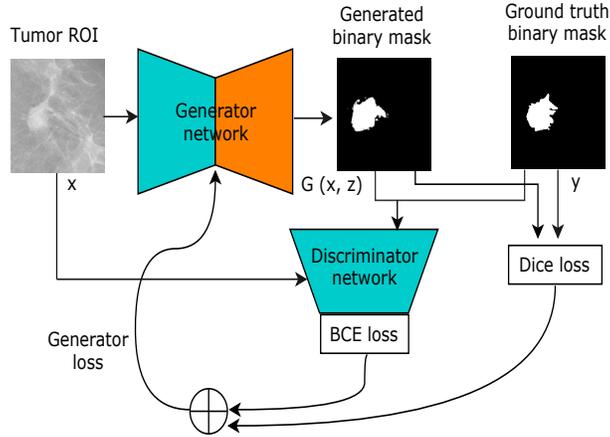


Figure 4: Proposed cGAN framework based on dice and BCE losses.

the loss function of G is defined as:

$$\ell_{Gen}(G, D) = \mathbb{E}_{x,y,z}(-\log(D(x, G(x, z)))) + \lambda \mathbb{E}_{x,y,z}(\ell_{Dice}(y, G(x, z))), \quad (1)$$

245 where z is introduced as dropout in the decoding layers Dn_1 , Dn_2 and Dn_3 at both training and testing phases, which provides stochasticity to generalize the learning processes and avoid overfitting.

The optimization process of G will try to minimize both expected values, *i.e.*, the D values should approach to 1.0 (correct tumor segmentations), and the dice loss ℓ_{Dice} 250 should approach to 0.0 (generated masks are equal to ground truth). Both terms of generator loss enforce the proper optimization of G : the dice loss term fosters a rough prediction of the mask shape (central tumor area) while the adversarial term fosters an accurate prediction of the mask outline (tumor borders). Neglecting one of the two terms may lead to either very poor segmentation results or slow learning speed.

In addition, $\ell_{Dice}(y, G(x, z))$ is the dice loss of the predicted mask with respect to ground truth, which is defined as:

$$\ell_{Dice}(y, z) = 1 - \frac{2|y \circ G(x, z)|}{|y| + |G(x, z)|}, \quad (2)$$

255 where \circ is the pixel wise multiplication of the two images and $|\cdot|$ is the total sum of pixel values of a given image. If inputs are binary images, then each pixel can be considered

as a boolean value (white is 1 / black is 0). The formulation in (2) is equivalent to the dice coefficient *i.e.*, $2 \times \frac{TP}{TP+FN+TP+FP}$, but it must be subtracted from 1.0 because the loss function will be minimized. Let A be the ground truth of the ROI and B the segmented region. Then the true positive degree (TP) is defined as $TP = A \cap B$, which is the area of the segmented region common in both A and B . The false positive degree (FP) is defined as $\bar{A} \cap B$, which is the segmented area not belonging to A . Similarly, the false negative degree (FN) is defined as $A \cap \bar{B}$, which is the true area missed by the proposed segmentation method.

In our previous work Singh et al. (2018b), the generator network loss was formulated by combining the logistic Binary Cross Entropy (BCE) loss and the $L1$ -norm. In this work, we replace the $L1$ -norm loss with the dice loss as shown in Fig. 4. $L1$ -norm loss minimizes the sum of absolute differences between the ground truth label y and estimated binary mask $G(x, z)$ obtained from the generator network, which takes all pixels into account. In turn, dice loss is highly dependent on TP predictions, which is the most influential term in foreground segmentation. Fig. 5 shows that the dice loss achieves lower values (more optimal) than the $L1$ -norm loss.

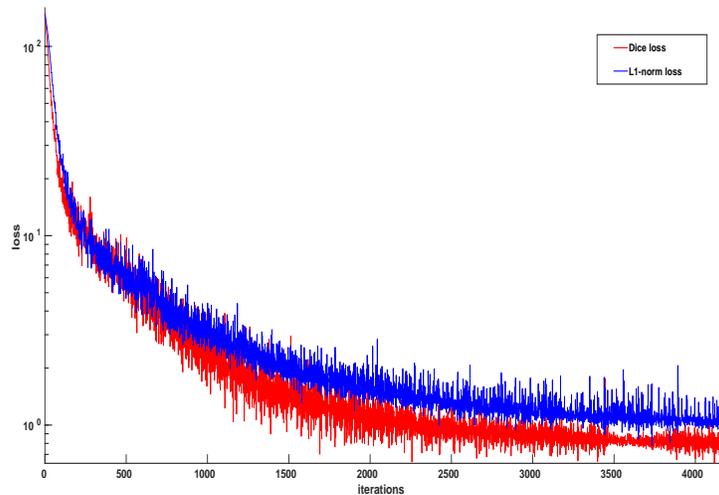


Figure 5: Dice and $L1$ -norm loss comparison over iterations.

Moreover, the loss function of D is defined in (3):

$$\ell_{Dis}(G, D) = \mathbb{E}_{x,y,z}(-\log(D(x, y))) + \mathbb{E}_{x,y,z}(-\log(1 - D(x, G(x, z)))) \quad (3)$$

The optimizer will fit D to maximize the loss values for ground truth masks (by minimizing $-\log(D(x, y))$) and minimize the loss values for generated masks (by minimizing $-\log(1 - D(x, G(x, z)))$). These two terms compute BCE loss using both masks, assuming that the expected class for ground truth and generated masks is 1 and 0, respectively.

The optimization of G and D is done concurrently, *i.e.*, one optimization step for both networks at each iteration, where G learns how to compute a valid tumor segmentations and D learns how to differentiate between synthetic and real segmentations.

In this work, we experimented on different hyper-parameters to improve the segmentation accuracy of our previous contribution in Singh et al. (2018b). Besides introducing the dice loss, we have reduced the number of filters of each network from 64 to 32. We also explored different learning rates and loss optimizers (SGD, AdaGrad, Adadelta, RMSProp and Adam), finding Adam with $\beta_1 = 0.5$, $\beta_2 = 0.999$ and initial learning rate = 0.0002 with batch size 8 the best combination. In (1), the dice loss weighting factor $\lambda = 150$ was found to be the best choice. Finally, the best results were achieved by training both G and D from scratch for 150 epochs.

3.3. Shape Classification Model (CNN)

In the literature, various approaches for tumor shape classification have found that texture and intensity features are relevant for their proposals. However, in this proposal we attempt to use only shape context to classify the tumor shapes. Specifically, we propose a multi-class CNN architecture for breast tumor shape classification (*i.e.*, irregular, lobular, oval and round) using the binary masks obtained from the cGAN. In the literature, most methods attempted to directly categorize the shape using breast tumor intensity, texture, boundary, etc. (Kisilev et al. (2015, 2016); Ren et al. (2015); Kim et al. (2018)), which increase computational complexity. We simplify the problem by extracting morphological features from binary masks.

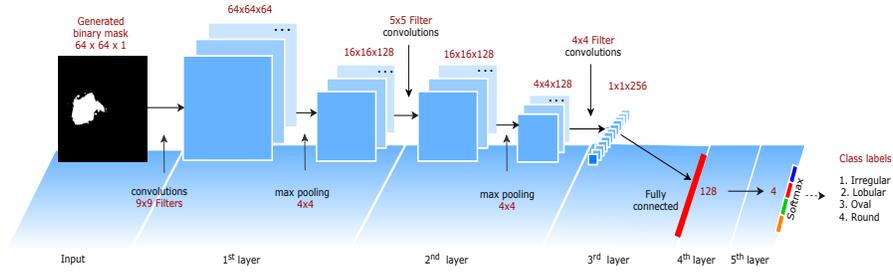


Figure 6: CNN architecture for tumor shape classification.

As shown in Fig. 6, our model consists of three convolutional layers with kernel sizes 9×9 , 5×5 and 4×4 , respectively, and two fully connected (FC) layers. The first two convolutional layers are followed by 4×4 max-pooling with stride 4×4 . The output of the last convolutional layer is flattened and then fed into the first FC layer with 128 neurons. These four layers use ReLU as activation function. A dropout of 0.5 is used to reduce overfitting in the first FC layer. Finally, the last FC layer with 4 neurons applies the softmax function to generate the final membership degree of the input binary mask to each class. A weighted categorical cross-entropy loss is used to avoid the problem of unbalanced dataset. The class weight is one minus the ratio of samples per class to the total number of samples.

The RMSProp is employed for optimizing the model with learning rate = 0.001, momentum = 0.9 and batch size = 16. The network is trained from scratch and the weights of five layers are randomly initialized. During training, we experimentally found the best architecture, number of layers, filters per layer, and number of neurons in FC layers.

4. Experiments and Discussion

We have evaluated the performance of proposed models on two public mammography datasets and one private dataset:

INbreast dataset²

It is a publicly available database containing a total of 115 cases (410 mammo-
grams), which include: masses, calcifications, asymmetries and distortions. However,
320 only 106 out of 410 mammograms have their corresponding ground truth of binary
masks. Thus, we only used this 106 mammograms to test our detection and segmenta-
tion model.

DDSM dataset

It is a publicly available digital database for screening mammography containing
325 2,620 mammography studies. In this work, 1,168 cases of breast tumors with their
corresponding ground truths are used for shape classification, where 504, 473, 115 and
76 tumors are labeled as irregular, lobular, oval and round, respectively. The remaining
images are excluded since they do not provide shape labels. We have used 75% of the
images for training and rest for testing the tumor shape classification model.

330 *Hospital Sant Joan de Reus dataset*

It is our private dataset that contains 300 malignant tumors (123 Luminal-A, 107
Luminal-B, 33 Her-2 and 37 Basal-like) with their respective ground truth binary masks
obtained by radiologists. The SSD detector and proposed cGAN segmentation model
is trained and tested using 220 and 80 images, respectively. The duty of confidentiality
335 and security measures were fully complied, in accordance with the current legislation
on the Protection of Personal Data (article 7.1 of the Organic Law 15/1999, 13th of
December).

The proposed method was implemented using python with Pytorch³ running on a
64-bit Ubuntu operating system using a 3.4 GHz Intel Core-i7 with 16 GB of RAM
340 and Nvidia GTX 1070 GPU with 8 GB of video RAM.

4.1. Tumor Detection Experiments

²http://medicalresearch.inescporto.pt/breastcancer/index.php/Get_INbreast_Database/

³<https://pytorch.org/>

In order to localize the tumor in the input mammographies, we compared different common deep learning detectors, such as Dhungel et al. (2017), Kozegar et al. (2013), Faster R-CNN Ren et al. (2015) , YOLO Redmon et al. (2016), and SSD Liu et al. (2016b). The tested detectors were trained with the Hospital Sant Joan de Reus dataset and tested with the INbreast dataset. Table 1 presents a quantitative comparison in terms of True Positive Rate (TPR) and False Positive Rate (FPR) with respect to the degree of overlapping between predicted and ground truth bounding boxes containing the tumor. To consider a true positive prediction, we require at least 60% of area overlapping.

Table 1: Mass detection accuracy of proposed method compared with the existing state-of-the-art methods.

Dataset	Method	TPR (%)	FPR (%)
INbreast	Dhungel et al. (2017)	96.00	1.20
	Kozegar et al.(2013)	87.00	3.67
	Faster R-CNN Ren et al. (2015)	96.00	2.94
	YOLO Redmon et al. (2016)	96.35	2.40
	SSD Liu et al. (2016b)	97.00	1.10

The SSD method yields the best results, with the highest TPR and lowest FPR. In turn, YOLO, Faster R-CNN and Dhungel et al. (2017) models have properly detected masses in the input mammograms, but with slightly worse quantitative results. Consequently, we have chosen the SSD model in order to locate tumors in mammograms.

4.2. Tumor Segmentation Experiments

The proposed breast tumor segmentation method is compared with the state-of-the-art methods and evaluated both quantitatively and qualitatively. For the quantitative analysis, segmentation accuracy is computed using Dice coefficient (F1 score) and Jaccard index (IoU). In turn, for the qualitative analysis, segmentation results with their respective ground truth binary masks are compared visually.

These experiments have been carried using three different framing of the tumor ROI: full mammogram, loose and tight frames (see Fig. 7). The ideal CAD system should be able to automatically segment the breast tumor from a full mammogram.

However, this is a very difficult task due to high similarity between gray level pixel
 365 distributions of healthy and tumorous tissue. Therefore, removing most of non-ROI
 portions of the image logically helps the model on learning the visual features that dif-
 ferentiate breast tumor from non-tumor areas. As mentioned in the methodology, for
 computing the loose and tight frames we rely on the initial tumor delimitation provided
 by the SSD method Liu et al. (2016b). The loose frame provides a convenient propor-
 370 tion between healthy and tumorous pixels. The tight frame is a square shrunk on the
 tumor, as it is intended to evaluate the behavior of the segmentation model when the
 majority of ROI contains tumor pixels. (see Fig. 9).

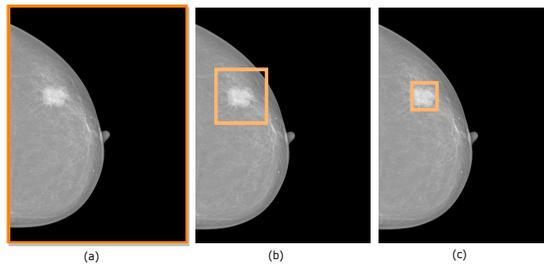


Figure 7: Three cropping strategies: (a) full mammogram, (b) loose frame, (c) tight frame.

The three cropping strategies are evaluated on our cGAN and eleven baseline seg-
 mentation models, referred as FCN, FCN-ResNet101, UNet, UNet-VGG16, SegNet,
 375 SegNet-VGG16, CRFCNN, SLSDeep, cGAN-ResNet101, cGAN-ResNet101 (Dice
 Loss) and proposed cGAN (without post-processing). FCN, UNet, SegNet, CNNCRF
 and proposed cGAN are trained from scratch. FCN-ResNet101, UNet-VGG16, SegNet-
 VGG16 and cGAN-ResNet101 (with and without Dice loss) are modifications of the
 original models, where the filters of the starting encoding layers are replaced by the
 380 starting convolutional layers of the well-known VGG (16 layers) and ResNet (101 lay-
 ers) models, which were pre-trained on the ImageNet database. Thus, we loaded the
 pre-trained weights and fine tuned the network. When using cGAN-ResNet101 Isola
 et al. (2017), we replaced the $L1$ -norm loss with the Dice loss in the generator loss
 function to see how the base line model will behave under such change. We called
 385 this model cGAN-ResNet101 (Dice loss) to compare the segmentation results with our
 proposal.

Table 2: Dice and IoU metrics obtained with the proposed model with/without post-processing and ten alternatives evaluated on the testing sets of our private and INbreast datasets, for the three cropping strategies. Best results are marked in bold. Dashes (-) indicate that results are not reported in referred papers.

Dataset	Methods	Dice (%)			IoU (%)		
		Full	Loose	Tight	Full	Loose	Tight
Private	FCN	59.06	74.94	80.20	39.92	62.21	78.89
	FCN-ResNet101	59.21	77.42	82.78	40.26	68.16	77.32
	UNet	63.69	78.03	83.15	46.73	68.36	78.81
	UNet-VGG16	59.27	78.57	83.71	42.13	69.71	79.42
	SegNet	59.87	80.26	82.33	42.79	70.07	76.17
	SegNet-VGG16	61.59	81.09	81.41	41.61	68.19	77.82
	CRFCNN	53.21	71.33	63.52	41.38	65.24	54.28
	SLSDeep	59.64	71.10	84.28	43.89	60.16	79.93
	cGAN-ResNet101	58.37	80.11	86.22	42.12	71.91	76.62
	cGAN-ResNet101 (Dice Loss)	61.49	86.57	86.37	45.90	76.32	77.26
	Proposed cGAN (without post-processing)	65.17	88.42	87.77	48.45	80.67	78.22
	Proposed cGAN (with post-processing)	66.38	89.99	88.12	49.68	81.81	79.87
INbreast	FCN	54.36	66.12	81.74	36.88	49.38	77.33
	FCN-ResNet101	51.76	83.80	82.38	38.49	74.12	78.09
	UNet	55.58	77.92	80.76	38.46	70.83	77.97
	UNet-VGG16	56.79	78.02	80.89	39.65	68.32	78.13
	SegNet	53.33	79.06	81.11	36.36	65.37	77.02
	SegNet-VGG16	56.27	80.17	81.75	39.46	69.79	78.68
	CRFCNN	52.96	73.25	65.41	40.41	67.14	57.69
	SLSDeep	60.35	75.90	85.53	44.63	65.16	80.26
	cGAN-ResNet101	54.69	87.19	89.17	37.94	77.51	82.26
	cGAN-ResNet101 (Dice Loss)	59.72	88.89	90.42	44.89	82.58	82.95
	Proposed cGAN (without post-processing)	67.55	93.64	91.47	50.05	86.29	83.58
	Proposed cGAN (with post-processing)	68.69	94.07	92.11	52.31	87.03	84.55
	Dhungel et al. (2015b)	-	-	90.00	-	-	-
	Cardoso et al. (2017)	-	-	90.00	-	-	-
Zhu et al. (2018)	-	-	90.97	-	-	-	
Al-antari et al. (2018)	-	-	92.69	-	-	86.37	

The results depicted in Table 2 are divided in two sections, one for our private dataset and another for the INbreast dataset. Note that all models are trained on the private dataset, and then tested using our private dataset as well as the INbreast dataset
390 without fine tuning. The results of all tested methods are after the post processing explained in sub-section 3.1.

According to the results, our method outperforms the compared state-of-the-art methods in all cases except for the IoU computed on tight crops of our private dataset. The SLSDeep approach yielded the best IoU (79.93%), whereas our method yielded
395 the second best result (79.87%) with a very small difference of 0.06%. The post-processing improved the results of our proposed model by 1% with the three framing inputs.

All models yielded their worst segmentation results with full mammograms compared to other framing inputs, which is logical taking into account the difficulties stated
400 earlier in this section. Most of the models have obtained their best results for the tight frame crops except for CRFCNN and our proposal, which yielded their best results for loose frame crops. However, the good results for tight crops may be due to the imbalance of tumor/non-tumor pixels, since the former class is present in more than 90% of the image area. The learning can be biased towards this class, which makes rough so-
405 lutions (almost everything is tumor) to provide very high ranks of performance. Loose frame crops, on the contrary, have a more balanced proportion of pixels for both classes, which makes them ideal to learn and evaluate the model on a realistic situation: it is more convenient for radiologists to provide a fast frame drawing around the breast tumor rather than a tight frame.

410 Comparing the general results for both datasets, most methods performed better on INbreast rather than on private dataset with loose and tight framing. This effect can be explained by the fact that INbreast provides more detailed ground truths, which leads to better testing results, despite all network training has been conducted on our private dataset.

415 In general, our proposal, with and without post-processing, has performed well in terms of both Dice and IoU metrics. For private dataset, in Dice/Loose frame column, our model with post-processing's score (89.99%) is almost 9% above the second

best model, SegNet-VGG16 (81.09%). In the IoU/Loose frame column, our model's percentage (81.81%) is almost 10% above the second best model, cGAN-ResNet101 (71.91%). For INbreast dataset, our loose frame results for Dice and IoU are again the best (94.07%, 87.03%), where cGAN-ResNet101 is the second best model for both metrics (87.19%, 77.51%). Thus, our model provides an improving of 7% and 10%, respectively. The fact that the second best results are obtained by the cGAN-ResNet101 model indicates that the adversarial network really helps in training the generative network. In turn, the results obtained by the cGAN-ResNet101 (Dice Loss) mixture model are in-between the cGAN-ResNet101 and our proposal, since the Dice loss term substitution improves the accuracy of tumor segmentations.

For the INbreast dataset, we have included the results mentioned in four related papers Dhungel et al. (2015b), Cardoso et al. (2017), Zhu et al. (2018) and Al-antari et al. (2018). For these methods, we could not compute the metrics for all columns, since they have not released their source code. Our method outperformed the first three papers under similar framework conditions. However, Al-antari et al. (2018) yielded better results for dice (92.69%) and IoU (86.37%) than our model in the Tight frame columns. Our results in the Loose frame columns surpass their results. For a fair comparison, however, it should be checked how the referenced methods would perform on loose frame crops.

The box-plot in Figure 8 shows Dice and IoU values obtained for the 106 testing samples from INbreast dataset with loose frames using FCN-ResNet101, Unet-VGG16, SegNet-VGG16, SLSDeep, cGAN-ResNet101 and proposed cGAN. The two models based on cGAN provide small ranges of Dice and IoU values. For instance, the proposed cGAN is in the range 0.89 to 0.93 for Dice coefficient and 0.80 to 0.91 for IoU values, while other deep segmentation methods, SLSDeep, Unet-VGG16 and FCN-ResNet101, show a wider range of values. Moreover, there are many outliers in the results for the segmentation based on the cGAN using pre-trained ResNet101 layers, while using our cGAN trained from scratch there are few number of outliers.

The high Dice and IoU metrics obtained by our model empirically support our hypothesis that it achieves accurate tumor segmentation. In Fig. 9, we show some examples of our model's segmentations using two tumors from the INbreast dataset by

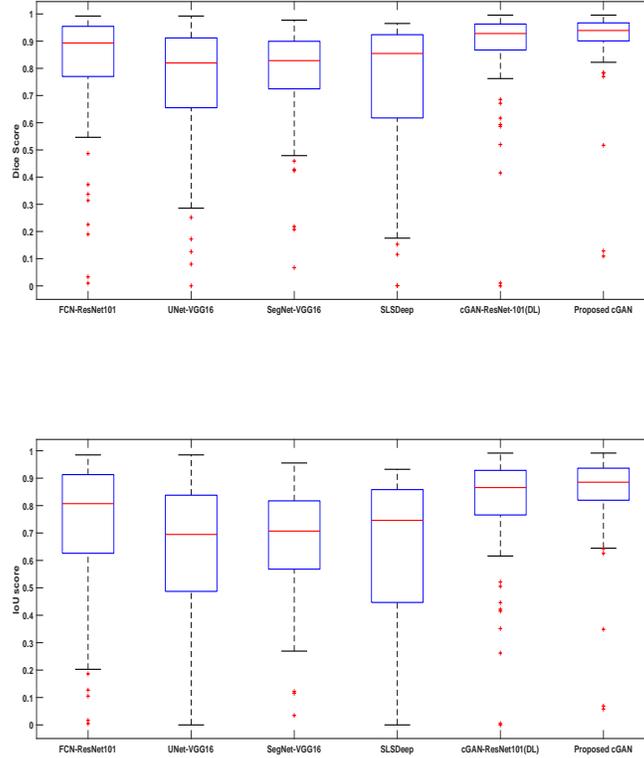


Figure 8: Boxplot of dice (Top) and IoU (Bottom) score over five models compared to our method on loose frames of the test subset of INbreast dataset (106 samples). Blue boxes indicate the interquartile range (Q3-Q1) of the metrics distribution, the red line inside each box represents the median value, the whiskers extend 1.5 times the length of Q1 and Q3, and (+) indicate outlier values, i.e. metrics out of the whiskers.

applying all three cropping strategies. For each experiment, we show the original ROI
 450 image and the comparison of predicted and ground truth mask, color coded to mark up
 the true positives (TP:yellow), false negatives (FN:red), false positives (FP:green) and
 true negatives (TN:black). For the full mammogram, the ROI image (1) is an example
 of good segmentation, since yellow and black pixels depict a high degree of confidence
 between predicted and real masks. On the contrary, the ROI image (2) is an example
 455 of poor segmentation, since red pixels mark up a high portion of the breast tumor area

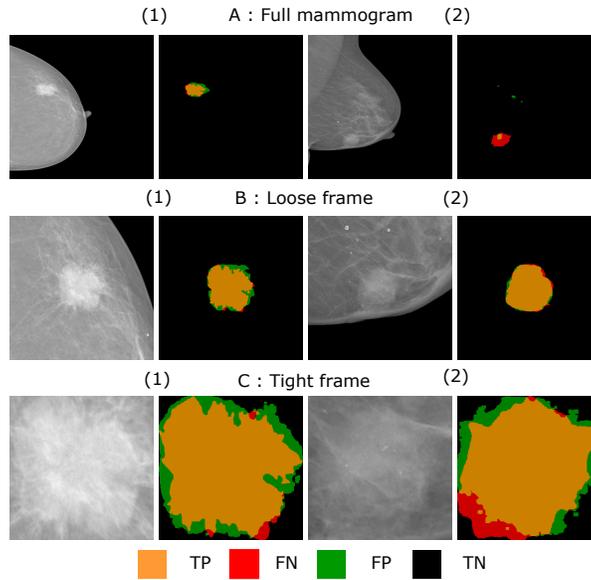


Figure 9: Segmentation results of two testing samples extracted from the INbreast dataset with the three cropping strategies.

that has been misclassified as healthy area (FN). At the same time, a tiny region of green pixels shows the misclassification of healthy tissue as breast tumor area (FP). Nevertheless, even in this second segmentation, there is a very high rate of black pixels (TN), which indicates that the model easily recognizes non-tumor areas.

460 In the loose frame segmentations (middle row), specially with example (2), the results contain very few FN and FP pixels. For example (1), a modest amount of green pixels indicate that our model expands the tumor segmentation beyond its respective ground truth. In the tight frame crops (bottom row), besides the green areas, our model also has missed some tumor areas *i.e.*, the red pixels (FN). The mistaken areas (red and
 465 green) are mostly around the tumor borders, since these areas have a mixture of healthy and unhealthy cells. At the same time, the inner part of the tumor as well as the image regions outside of tumors are properly classified, which indicates the stability of our model.

470 Fig. 10 shows a comparison between our and other six segmentation models, which worked on loose and tight frame crops using four tumors from the INbreast dataset. For

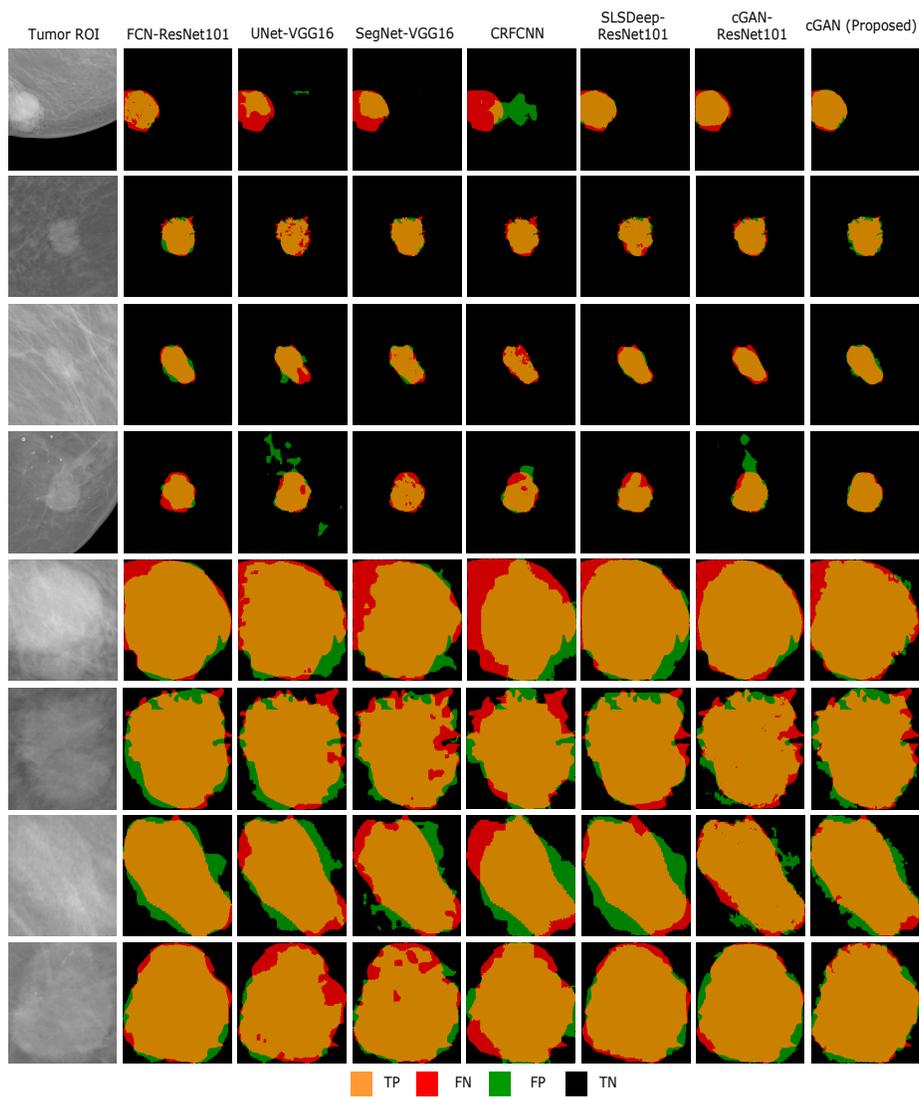


Figure 10: Segmentation results of seven models with the INbreast dataset and two cropping strategies: loose frame (the first four rows) and tight frame (the last four rows). (Col 1) original images, (Col 2) FCN-ResNet101, (Col 3) UNet-VGG16, (Col 4) SegNet-VGG16, (Col 5) CRFCNN, (Col 6) SLSDeep, (Col 7) cGAN-ResNet101, and (Col 8) proposed cGAN.

the loose frame cases (four top rows), our method clearly outperforms the rest for all tumors except for the second one, where the majority of models provided a similar degree of accuracy. In these four tumors, UNet-VGG16 and CRFCNN provided the worst results. Moreover, cGAN-ResNet101 also performed bad in the fourth example.

475 For the tight frame cases (four bottom rows), our method also provides the lowest degrees of FN and FP compared to the rest of the models. Our cGAN and the cGAN-ResNet101 model yield irregular borders compared to FCN-ResNet101 and SLSDeep, since GAN models strive for higher accuracy on edges. However, in the third tight frame sample (seventh row), both cGAN-ResNet101 and our proposal generated an
480 irregular border that slightly differs from the smooth ground truth border, which results in lower segmentation accuracy around the edges. Although the rest of the models generate smoother borders, the resulting segmentations may differ from the ground truth significantly.

From the experimental results, it can be concluded that the proposed breast tumor
485 segmentation method is the most effective to date compared to the currently available state-of-the-art methods. However, our method needs a loose crop around the tumor to obtain a proper segmentation, which can be done by the SSD model. Our segmentation model contains about 13,607,043 parameters for tuning the generator part in the cGAN network. In addition, our method is fast in both training *i.e.*, around 30 seconds per
490 epoch (220 loose frames) and predicting, around 7 images per second. That is 7 to 8 times faster than the segmentation method proposed in Al-antari et al. (2018) and 10 to 15 times faster than the FCN model.

4.3. Shape Classification Experiments

For validating the tumor shape classification performance, we computed the confu-
495 sion matrix and the overall classification accuracy on the test set of the DDSM dataset. This set contains 292 images divided into 126, 117, 31 and 18 for irregular, lobular, oval and round classes, respectively.

However, The DDSM dataset does not have the ground truth binary masks for the breast tumor segmentation. Thus, we applied active contours Akram et al. (2015),
500 which was also used in our previous work Singh et al. (2018b), to generate the ground

Table 3: Confusion matrix of the tumor shape classification of testing samples of the DDSM dataset.

Prediction / Ground Truth	Irregular	Lobular	Oval	Round	Total
Irregular	96 (76%)	30	0	0	126
Lobular	33	83 (71%)	1	0	117
Oval	0	1	26 (84%)	4	31
Round	0	1	1	16 (89%)	18

truths of the breast tumor regions. Previously, Kisilev et al. (2015) also used active contours Lankton and Tannenbaum (2008) to generate the ground truths in a similar fashion. These ground truth masks are verified by expert radiologists of the hospital of Sant Joan de Reus. In addition, for reliable performance results, we used a stratified 5 fold cross validation with 50 epochs per fold.

In Table 3, the proposed method yielded around 73% of classification accuracy for irregular and lobular classes. This result is logical, since both lobular and irregular shapes have similar irregular boundaries. In turn, our model yielded classification accuracies of 84% and 89% for oval and round shape classes, respectively.

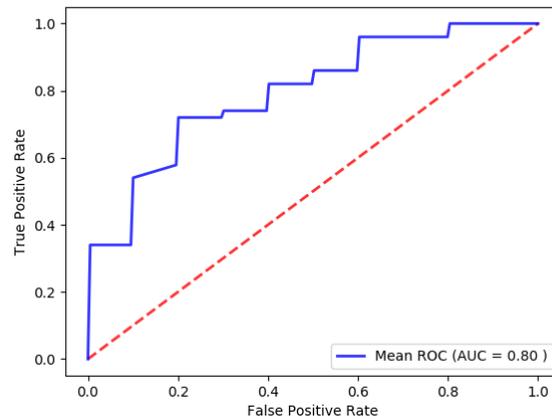


Figure 11: Mean ROC curve of 5 folds, for TPR and FPR from shape classification result of 292 test images from DDSM dataset.

510 For a quantitative comparison, we compared three state-of-the-art tumor shape classification methods Singh et al. (2018b); Kisilev et al. (2015); Kim et al. (2018) with three variations of our shape classification model: one is fed with a binary mask with the ground-truth, the second is fed with a binary mask generated by our segmentation stage and the third is fed with the original ROI image masked with the segmented
515 area (with pixel-wise multiplication). The five methods were evaluated on the DDSM dataset. We have computed the overall accuracy of each method by averaging the correct predictions (i.e., true positive) of the four classes, weighted with respect to the number of samples per class. As shown in Table 4, our classifier based only on binary masks yields an overall accuracy of 80%, outperforming the second best results Kim
520 et al. (2018); Singh et al. (2018b) by 8%. The 83% obtained with our method fed with the original ground truth cannot be considered as a valid result for comparison, since it is the training data accuracy. We provide this result only to show the low degree of overfitting achieved by our network. In turn, the proposed method fed with the masked ROI images provided 70% of overall accuracy. This experiment indicates that
525 gray-level variations inside the segmented area is somehow confusing our shape classification network. In another hand, the multi-task CNN proposed in Kim et al. (2018) based on a pre-trained VGG-16 yielded the worst overall accuracy (66%), probably because the input mammograms are gray-scale images, while the VGG-16 network was trained on color-scale images. In addition, Fig. 11 shows ROC curve illustrating that
530 our model attained AUC about 0.8.

Furthermore, the proposed shape descriptor contains 767,684 parameters, which can be trained in less than a second per epoch, and predict in about 6 milliseconds per image.

4.4. Shape Features Correlation to Breast Cancer Molecular Subtypes

535 Tumor shape could play an important role to predict the breast cancer molecular subtypes Tamaki et al. (2011). Thus, we have computed the correlation between breast cancer molecular subtypes classes of our in-house private dataset with the four shape classes. As shown in Table 5, most of Luminal-A and -B samples (i.e., 96/123 and 82/107 for Luminal-A and -B, respectively) are mostly assigned to irregular and

Table 4: Shape classification overall accuracy with the DDSM dataset resulting from Kisilev et al. (2015); Kim et al. (2018); Singh et al. (2018b) and our model. Best result is marked in bold.

Methods	Test samples	Accuracy (%)
(SSVM) Kisilev et al. (2015)	515	71
(Multi-task CNN) Kim et al. (2018)	218	66
(ICADx) Kim et al. (2018)	218	72
Singh et al. (2018b)	113	72
Proposed (with ground-truth masks)	292	83
Proposed (generated masks)	292	80
Proposed (masked ROI images)	292	70

Table 5: Distribution of breast cancer molecular subtypes samples from the hospital dataset with respect to its predicted mask shape.

Shape Classes / Molecular Subtypes	Irregular	Lobular	Oval	Round	Total
Luminal-A	67	29	10	17	123
Luminal-B	58	24	14	11	107
Her-2	6	4	8	15	33
Basal-like	5	10	9	13	37

540 lobular shape classes. In turn, oval and round tumors give indications to the Her-2
and Basal-like samples, (i.e., 23/33 and 22/37 for Her-2 and Basal-like, respectively).
Moreover, some images related to Basal-like are moderately assigned to the lobular
class. Afterwards, from the visual inspection, if the tumor shape is irregular or lobular
then radiologist can suspect that it belongs to the Luminal group. In turn, if the tumor
545 shape is round or oval then it is more probable that the tumor is a Her-2 or Basal-
like Tamaki et al. (2011). Therefore, this study shows the importance of tumor shape,
which can be considered as a key feature to distinguish between different malignancies
of breast cancer.

4.5. Limitations

550 For the segmentation stage, our model has only one limitation. If there are two tu-
mors (i.e., one is with complete shape and the other is incomplete) in the loose framing,
the proposed segmentation methods will be able to properly segment the complete tu-
mor and it will fail to segment the incomplete one. As shown in Fig. 12, we found three
samples that are mis-segmented because they contained two tumors, the one in the cen-
555 ter, which is properly segmented, and another that is shown partially in the left-down
border of the image, which is wrongly ignored as non-tumor region (FN). When the
bigger tumor is located in the center of the crop, nevertheless, it is correctly segmented.

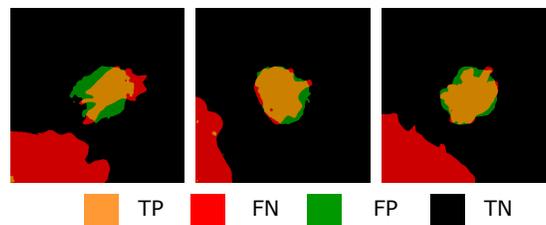


Figure 12: Three mis-segmented of non-full tumor shapes with INbreast dataset. The red part in the down-left border.

To classify the tumor shape, we depend only on the DDSM dataset to train our
model, since it is the only public dataset that has the shape classification information.
560 Thus, more databases containing more samples are required to improve the classifica-
tion accuracy of four shape classes.

To study the molecular subtypes of breast cancer, Her-2 and Basal-like classes have less samples compared to the other two classes, Luminal-A and Luminal-B. Indeed, we used a weighted loss function to train our shape classification model in order to make
565 a balance between the four classes. However, we anticipate that, by increasing the samples related to the Her-2 and Basal-like classes, we will improve the prediction of molecular subtypes from tumor shape information.

5. Conclusion

In this paper, we propose a two stage breast tumor segmentation and classification
570 method, which first segments the breast tumor ROI using a cGAN and then classify its binary mask using a CNN based shape descriptor.

The segmentation results reveal the importance of the adversarial network in the optimization of the generative network. cGAN-ResNet101 shows an improvement of about 1% to 3% in both Dice and IoU metrics in comparison to the other non-GAN
575 methods. In turn, the proposed method yields an increment of about 10% over the results of cGAN-ResNet101 by training our model from scratch, and replacing the $L1$ -norm with the dice loss using loose frame crop on the given datasets. The breast tumor segmentation from full-mammograms yields low segmentation accuracy for all models including the proposed cGAN. For the tight frame crop, the proposed cGAN yields
580 similar or better segmentation accuracy compared to the other methods.

The classification results show that our second stage properly infers the tumor shape from the binary mask of the breast tumor, which was obtained from the first stage (cGAN segmentation). Hence, we have empirically shown that our CNN is focusing its learning on the morphological structure of the breast tumor, while the rest
585 of approaches (Kisilev et al. (2015), Kim et al. (2018), Kisilev et al. (2016), Ren et al. (2015)) rely on the original pixel variations of the input mammogram to make the same inference. Moreover, in Al-antari et al. (2018) they used a hybrid strategy in which they include the pixel variability within the mask of breast tumor region to retain the intensity and texture information. However, the higher performance obtained by our method
590 supports our initial idea that the second stage CNN can reliably recognize the tumor

shape based only on morphological information.

Furthermore, this paper provided a study of correlation between the tumor shape and the molecular subtypes of the breast cancer. Most samples of the Luminal-A and -B group are assigned to irregular shapes. In turn, the majority of Her-2 and Basal-like
595 samples are assigned to regular shapes (e.g., oval and round shapes). That gives an indication that the tumor shape can be considered for inferring the molecular subtype of the tumor.

Future work aims at refining our multi-stage framework to detect other breast tumor features (*i.e.*, margin type, micro-calcifications), which will be integrated into a more
600 comprehensive diagnostic to compute the degree of malignancy of the breast tumors.

Conflict of Interest

The authors declare that there is no conflict of interest.

References

- Akram, F., Kim, J., Lee, C., Choi, K.N., 2015. Segmentation of regions of interest
605 using active contours with SPF function. *Comp. Math. Methods in Medicine* 2015, 710326:1–710326:14. doi:10.1155/2015/710326.
- Al-antari, M.A., Al-masni, M.A., Choi, M.T., Han, S.M., Kim, T.S., 2018. A fully integrated computer-aided diagnosis system for digital x-ray mammograms via deep learning detection, segmentation, and classification. *International Journal of Medical
610 Informatics* 117, 44–54.
- Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39, 2481–2495.
- Boisserie-Lacroix, M., Hurtevent-Labrot, G., Ferron, S., Lippa, N., Bonnefoi, H.,
615 Mac Grogan, G., 2013. Correlation between imaging and molecular classification of breast cancers. *Diagnostic and interventional imaging* 94, 1069–1080.

- Cardoso, J.S., Domingues, I., Oliveira, H.P., 2015. Closed shortest path in the original coordinates with an application to breast cancer. *International Journal of Pattern Recognition and Artificial Intelligence* 29.
- 620 Cardoso, J.S., Marques, N., Dhungel, N., Carneiro, G., Bradley, A., 2017. Mass segmentation in mammograms: A cross-sensor comparison of deep and tailored features, in: *Proceeding of the IEEE International Conference on Image Processing (ICIP)*, pp. 1737–1741.
- Cheng, H., Cai, X., Chen, X., Hu, L., Lou, X., 2003. Computer-aided de-
625 tection and classification of microcalcifications in mammograms: a survey. *Pattern Recognition* 36, 2967 – 2991. URL: <http://www.sciencedirect.com/science/article/pii/S0031320303001924>, doi:[https://doi.org/10.1016/S0031-3203\(03\)00192-4](https://doi.org/10.1016/S0031-3203(03)00192-4).
- Cho, N., 2016. Molecular subtypes and imaging phenotypes of breast cancer. *Ultra-
630 sonography* 35, 281.
- Dhungel, N., Carneiro, G., Bradley, A., 2015a. Tree re-weighted belief propagation using deep learning potentials for mass segmentation from mammograms, in: *12th IEEE International Symposium on Biomedical Imaging, ISBI 2015, Brooklyn, NY, USA, April 16-19, 2015*, pp. 760–763. doi:10.1109/ISBI.2015.7163983.
- 635 Dhungel, N., Carneiro, G., Bradley, A.P., 2015b. Deep learning and structured prediction for the segmentation of mass in mammograms, in: *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 605–612.
- Dhungel, N., Carneiro, G., Bradley, A.P., 2017. A deep learning approach for the
640 analysis of masses in mammograms with minimal user intervention. *Medical image analysis* 37, 114–128.
- Fang, Y., Xie, J., Dai, G., Wang, M., Zhu, F., Xu, T., Wong, E., 2015. 3d deep shape descriptor, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2319–2328.

- 645 Fu, H., Cheng, J., Xu, Y., Wong, D.W.K., Liu, J., Cao, X., 2018. Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. arXiv preprint arXiv:1801.00926 .
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets, in: Advances in neural
650 information processing systems, pp. 2672–2680.
- Hamidinekoo, A., Denton, E., Rampun, A., Honnor, K., Zwigelaar, R., 2018. Deep learning in mammography and breast histology, an overview and future trends. Medical image analysis 47, 45–67.
- Hazarika, M., Mahanta, L.B., 2018. A new breast border extraction and contrast en-
655 hancement technique with digital mammogram images for improved detection of breast cancer. Asian Pacific journal of cancer prevention: APJCP 19, 2141.
- Hofer, C., Kwitt, R., Niethammer, M., Uhl, A., 2017. Deep learning with topological signatures, in: Advances in Neural Information Processing Systems, pp. 1634–1644.
- Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A., 2017. Image-to-image translation with con-
660 ditional adversarial networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5967–5976.
- Jiao, Z., Gao, X., Wang, Y., Li, J., 2018. A parasitic metric learning net for breast mass classification based on mammography. Pattern Recognition 75, 292–301.
- Kim, S.T., Lee, H., Kim, H.G., Ro, Y.M., 2018. ICADx: Interpretable computer aided
665 diagnosis of breast masses, in: Proceedings of the SPIE - Medical Imaging 2018: Computer-Aided Diagnosis.
- Kisilev, P., Sason, E., Barkan, E., Hashoul, S., 2016. Medical image description using multi-task-loss cnn, in: Deep Learning and Data Labeling for Medical Applications. Springer, pp. 121–129.
- 670 Kisilev, P., Walach, E., Hashoul, S.Y., Barkan, E., Ophir, B., Alpert, S., 2015. Semantic description of medical image findings: structured learning approach, in: Proceedings of the British Machine Vision Conference (BMVC), pp. 171.1–171.11.

- Kozegar, E., Soryani, M., Minaei, B., Domingues, I., et al., . Assessment of a novel mass detection algorithm in mammograms .
- 675 Kshema, George, M.J., Dhas, D.A.S., 2017. Preprocessing filters for mammogram images: A review, in: 2017 Conference on Emerging Devices and Smart Systems (ICEDSS), pp. 1–7. doi:10.1109/ICEDSS.2017.8073694.
- Kurnianggoro, L., Jo, K.H., et al., 2018. A survey of 2d shape representation: Methods, evaluations, and future research directions. *Neurocomputing* 300, 1–16.
- 680 Lankton, S., Tannenbaum, A., 2008. Localizing region-based active contours. *IEEE transactions on image processing* 17, 2029–2039.
- Lauby-Secretan, B., Scoccianti, C., Loomis, D., Benbrahim-Tallaa, L., Bouvard, V., Bianchini, F., Straif, K., 2015. Breast-cancer screening-viewpoint of the IARC working group. *New England Journal of Medicine* 372, 2353–2358.
- 685 Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A., Van Ginneken, B., Sánchez, C.I., 2017. A survey on deep learning in medical image analysis. *Medical image analysis* 42, 60–88.
- Liu, S., Wu, X.D., Xu, W.J., Lin, Q., Liu, X.J., Li, Y., 2016a. Is there a correlation between the presence of a spiculated mass on mammogram and luminal a subtype breast cancer? *Korean journal of radiology* 17, 846–852.
- 690 Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C., 2016b. Ssd: Single shot multibox detector, in: *European conference on computer vision*, Springer. pp. 21–37.
- Liu, Z., Zhuo, C., Xu, X., 2018. Efficient segmentation method using quantised and non-linear cenn for breast tumour classification. *Electronics Letters* .
- 695 Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation, in: *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pp. 3431–3440.

- Luciano, L., Hamza, A.B., 2018. Deep learning with geodesic moments for 3d shape
700 classification. *Pattern Recognition Letters* 105, 182–190.
- Masoumi, M., Hamza, A.B., 2017. Spectral shape classification: A deep learning
approach. *Journal of Visual Communication and Image Representation* 43, 198–
211.
- Matos, C.E.F., Souza, J.C., Diniz, J.O.B., Junior, G.B., de Paiva, A.C., de Almeida,
705 J.D.S., da Rocha, S.V., Silva, A.C., 2018. Diagnosis of breast tissue in mammog-
raphy images based local feature descriptors. *Multimedia Tools and Applications* ,
1–26.
- Qi, C.R., Su, H., Mo, K., Guibas, L.J., 2017. Pointnet: Deep learning on point sets for
3d classification and segmentation. *Proc. Computer Vision and Pattern Recognition*
710 (CVPR), IEEE 1, 4.
- Rangayyan, R.M., Banik, S., Desautels, J.L., 2010. Computer-aided detection of archi-
tectural distortion in prior mammograms of interval cancer. *Journal of Digital*
Imaging 23, 611–631.
- Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified,
715 real-time object detection, in: *Proceedings of the IEEE conference on computer*
vision and pattern recognition, pp. 779–788.
- Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: Towards real-time object
detection with region proposal networks, in: *Advances in neural information pro-*
cessing systems, pp. 91–99.
- 720 Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for
biomedical image segmentation, in: *Proceedings of the International Conference on*
Medical image computing and computer-assisted intervention (MICCAI), pp. 234–
241.
- Rouhi, R., Jafari, M., Kasaei, S., Keshavarzian, P., 2015. Benign and malignant breast
725 tumors classification based on region growing and cnn segmentation. *Expert Sys-*
tems with Applications 42, 990–1002.

- Sarker, M.M.K., Rashwan, H.A., Akram, F., Banu, S.F., Saleh, A., Singh, V.K., Chowdhury, F.U.H., Abdulwahab, S., Romani, S., Radeva, P., Puig, D., 2018. SLSDeep: Skin lesion segmentation based on dilated residual and pyramid pooling networks, in: Medical Image Computing and Computer Assisted Intervention - MICCAI 2018 - 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II, pp. 21–29. URL: https://doi.org/10.1007/978-3-030-00934-2_3, doi:10.1007/978-3-030-00934-2_3.
- Schmidhuber, J., 2015. Deep learning in neural networks: An overview. *Neural networks* 61, 85–117.
- Siegel, R.L., Miller, K.D., Jemal, A., 2017. Cancer statistics, 2017. *CA: a cancer journal for clinicians* 67, 7–30.
- Singh, V.K., Rashwan, H., Akram, F., Pandey, N., Sarker, M., Kamal, M., Saleh, A., Abdulwahab, S., Maarroof, N., Romani, S., et al., 2018a. Retinal optic disc segmentation using conditional generative adversarial network. *arXiv preprint arXiv:1806.03905*.
- Singh, V.K., Romani, S., Rashwan, H.A., Akram, F., Pandey, N., Sarker, M.M.K., Abdulwahab, S., Torrents-Barrena, J., Saleh, A., Arquez, M., Arenas, M., Puig, D., 2018b. Conditional generative adversarial and convolutional networks for x-ray breast mass segmentation and shape classification, in: Medical Image Computing and Computer Assisted Intervention - MICCAI 2018 - 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II, pp. 833–840. URL: https://doi.org/10.1007/978-3-030-00934-2_92, doi:10.1007/978-3-030-00934-2_92.
- Singh, V.K., Romani, S., Torrents-Barrena, J., Akram, F., Pandey, N., Sarker, M.M.K., Saleh, A., Arenas, M., Arquez, M., Puig, D., 2017. Classification of breast cancer molecular subtypes from their micro-texture in mammograms using a vggnet-based convolutional neural network, in: Recent Advances in Artificial Intelligence Research and Development: Proceedings of the 20th International Conference of the

- 755 Catalan Association for Artificial Intelligence, Deltebre, Terres de L'Ebre, Spain,
October 25-27, 2017, IOS Press. p. 76.
- Tamaki, K., Ishida, T., Miyashita, M., Amari, M., Ohuchi, N., Tamaki, N., Sasano, H.,
2011. Correlation between mammographic findings and corresponding histopathol-
ogy: potential predictors for biological characteristics of breast diseases. *Cancer*
760 *science* 102, 2179–2185.
- Tang, J., Rangayyan, R.M., Xu, J., El Naqa, I., Yang, Y., 2009. Computer-aided de-
tection and diagnosis of breast cancer with mammography: recent advances. *IEEE*
Transactions on Information Technology in Biomedicine 13, 236–251.
- Yang, D., Xu, D., Zhou, S.K., Georgescu, B., Chen, M., Grbic, S., Metaxas, D., Co-
765 maniciu, D., 2017. Automatic liver segmentation using an adversarial image-to-
image network, in: *Proceedings of the International Conference on Medical image*
computing and computer-assisted intervention (MICCAI), pp. 507–515.
- Yassin, N.I., Omran, S., El Houbay, E.M., Allam, H., 2018. Machine learning tech-
niques for breast cancer computer aided diagnosis using different image modalities:
770 A systematic review. *Computer methods and programs in biomedicine* 156, 25–45.
- Zhu, W., Xiang, X., Tran, T.D., Hager, G.D., Xie, X., 2018. Adversarial deep structured
nets for mass segmentation from mammograms, in: *Proceedings of the IEEE 15th*
International Symposium on Biomedical Imaging (ISBI), pp. 847–850.