A performance comparison between shallow and deeper neural networks supervised classification of tomosynthesis breast lesions images

A Performance Comparison between Shallow and Deeper Neural Networks Supervised Classification of Tomosynthesis Breast Lesions Images

Vitoantonio Bevilacqua[a], Antonio Brunetti[a], Andrea Guerriero[a], Gianpaolo Francesco Trotta[b], Michele Telegrafo[c], Marco Moschetta[c]

[a] *Department of Electrical and Information Engineering (DEI), Polytechnic University of Bari. Via Orabona 4, 70126 - Bari, Italy.*
[b] *Department of Mechanics, Mathematics and Management Engineering (DMMM), Polytechnic University of Bari. Via Orabona 4, 70126 - Bari, Italy.*
[c] *D.E.T.O. University of Bari Medical School. Piazza Giulio Cesare 11, 70124 - Bari, Italy.*

Abstract

Computer Aided Decision (CAD) systems, based on 3D tomosynthesis imaging, could support radiologists in classifying different kinds of breast lesions and then improve the diagnosis of breast cancer (BC) with a lower X-ray dose than in Computer Tomography (CT) systems.

In previous work, several Convolutional Neural Network (CNN) architectures were evaluated to discriminate four different classes of lesions considering high-resolution images automatically segmented: a) irregular opacity lesions, b) regular opacity lesions, c) stellar opacity lesions and d) no-lesions. In this paper, instead, we use the same previously extracted relevant Regions of Interest (ROIs) containing the lesions, but we propose and evaluate two different approaches to better discriminate among the four classes.

In this work, we evaluate and compare the performance of two different frameworks both considering supervised classifiers topologies. The first framework is feature-based, and consider morphological and textural handcrafted features, extracted from each ROI, as input to optimised Artificial Neural Network (ANN) classifiers. The second framework, instead, considers non-neural classifiers based on automatically computed features evaluating the classification performance extracting several sets of features using different Convolutional Neural Network models.

Final results show that the second framework, based on features computed

automatically by CNN architectures performs better than the first approach, in terms of accuracy, specificity, and sensitivity.

Keywords: Breast Cancer, Tomosynthesis, Image Processing, Hand-crafted features, Shallow and Deep Artificial Neural Networks, Convolutional Neural Networks.

1. Introduction

Breast cancer is the second most prevalent cause of cancer death, which is turning into a global public health problem due to its complex aetiology and poor response to the treatment [1, 2]. In recent years, several studies have dealt with innovative treatments to retard breast cancer progression, such as melatonin [3, 4, 5, 6].

The increasing women life expectancy and the higher incidence of breast cancer in the general population require an accurate assessment of the breast glands with imaging techniques, and mammography still represents the gold standard imaging tool in this field [7].

Mammographic examinations, in fact, are used in several screening programs as they could lead to very early detection of the pathology, whereas other kinds of diagnostic tests, such as Magnetic Resonance (MR), Computed Tomography (CT) or Digital Breast Tomosynthesis (DBT) techniques, are necessary to perform a more in-depth analysis of risky cases, or for the follow-up of treated patients [8].

DBT has been recently introduced for breast cancer screening and detection; it consists in a promising innovative radiological technique for early diagnosis and staging. In details, DBT produces a limited angle cone beam tomosynthesis of the breast glands and has demonstrated to have a higher accuracy if compared to the most commonly used bi-dimensional imaging techniques, such as the previously introduced mammography, CT or MR [7, 9, 10, 11]. After the acquisition of multiple thin and high-resolution images, the DBT system produces a quasi-three-dimensional format of the reconstructed breast images aiming to reduce the effect of tissue superimposition, a typical side effect of the conventional planar digital mammography [12]. Moreover, since the edges of the breast lesions are better defined, DBT also improves the visualisation of masses and architectural distortions, thus leading to an improvement of the final diagnostic performance [13].

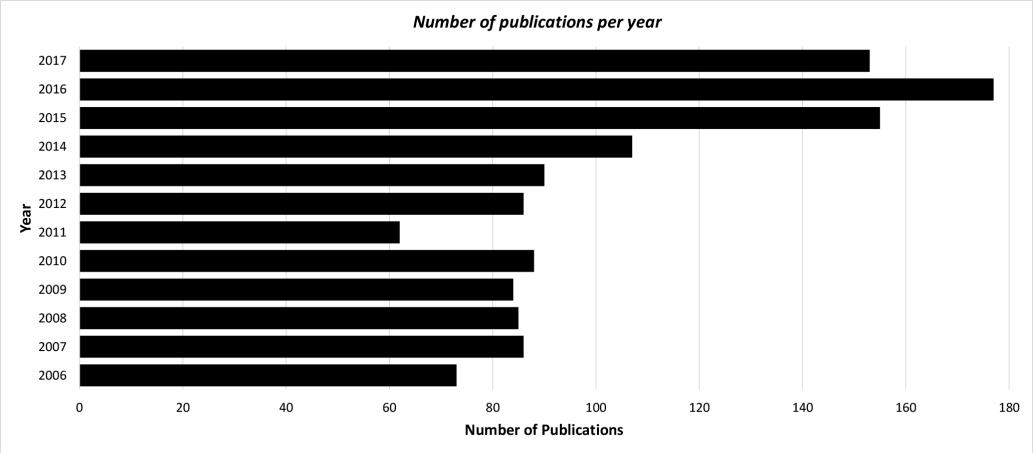


Figure 1: Number of publications per year from 2006 to 2017. Topic: Computer Aided Diagnosis & Medical Imaging. Indexes: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI.

In the last decades, Computer-Aided Decision (CAD) systems have become very popular in literature, with more than 1000 papers, indexed by the most common academic search engines from 2006 to 2017, dealing with computer-assisted diagnosis in the fields of medical imaging and biosignal processing, as depicted in Fig. 1. In fact, always more frequently in real-world applications, CAD systems support clinicians in everyday diagnostic practice offering a cheap and suitable alternative to double reading as a mean for reducing errors [14].

The design of traditional CAD systems usually follows a linear pipeline, as represented in Fig. 2, usually based on different steps, which are: (1) Image Acquisition, (2) Image Processing, (3) ROIs Extraction, (4) Feature Extraction, (5) Classification and finally (6) Validation [15, 16]. Although all the steps are crucial for the correct detection and classification of the pathological areas, the extraction good sets of features describing candidate areas for classification is of fundamental importance since the classification performances strongly depend on the description capability of the features [17].

In details, acquired images are generally pre-processed to eliminate noise and artefacts by using optimized filters [18, 19, 20]; after the pre-processing phase, images are segmented by using different algorithms based on the methodology used for pixels classification, such as the absolute grey-level values of pixels, the gradient, or even combined with geometrical distribution or

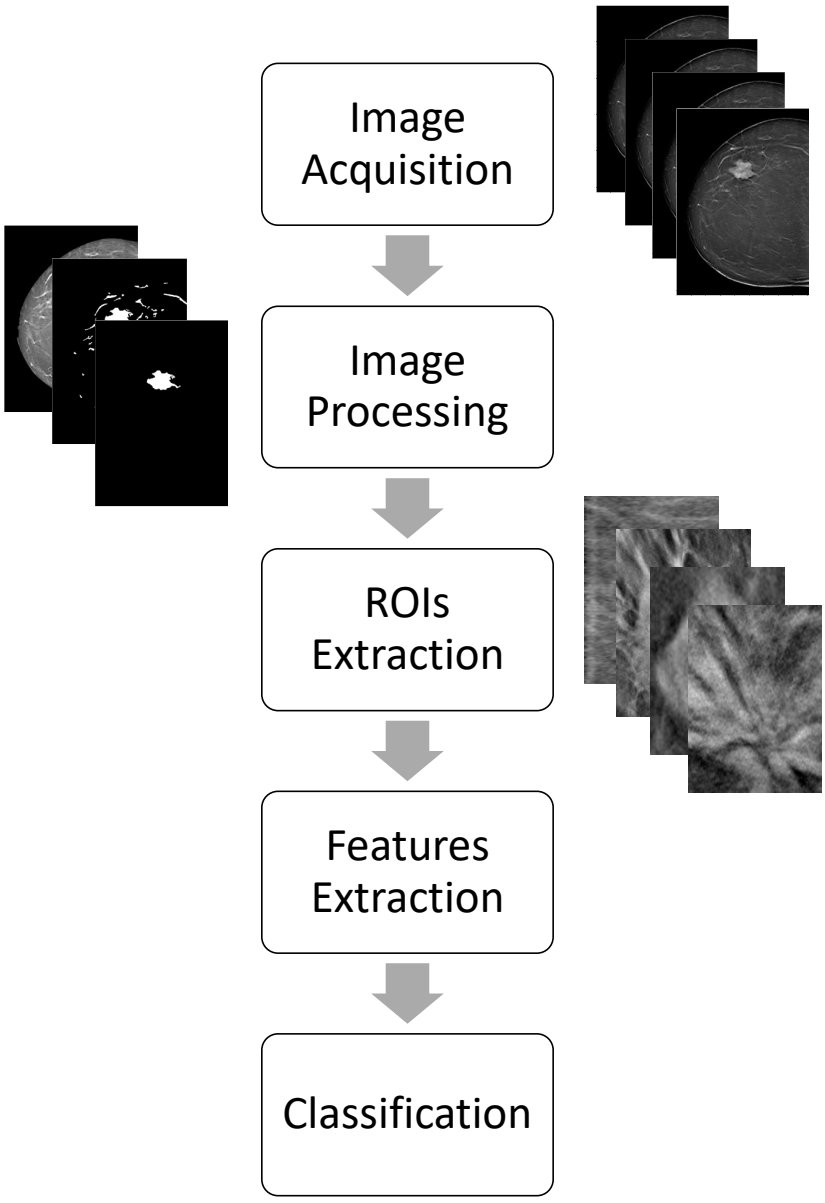


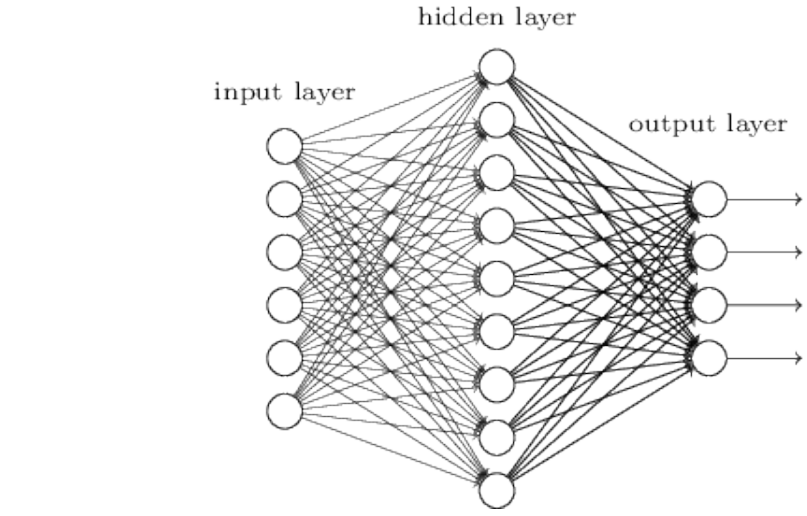
Figure 2: The steps of the pipeline usually implemented by traditional CAD systems. Input devices acquire the images which are then stored on a supporting device (generally in DICOM format). Subsequently, image processing algorithms are performed to obtain an enhanced representation, allowing an easier extraction of Regions of Interest (ROIs). The features describing the selected ROIs are computed and used as input to a classifier.

orientation properties [21, 22, 23, 24]. The segmentation step allows extracting ROIs containing candidate suspicious lesions which are then represented computing several descriptors, thus creating specific patterns of features. In the most of cases, the considered sets are built considering the same parameters used for the image processing step (e.g., gradient-based, intensity-based, or geometric), or may also group features of different kinds [25].

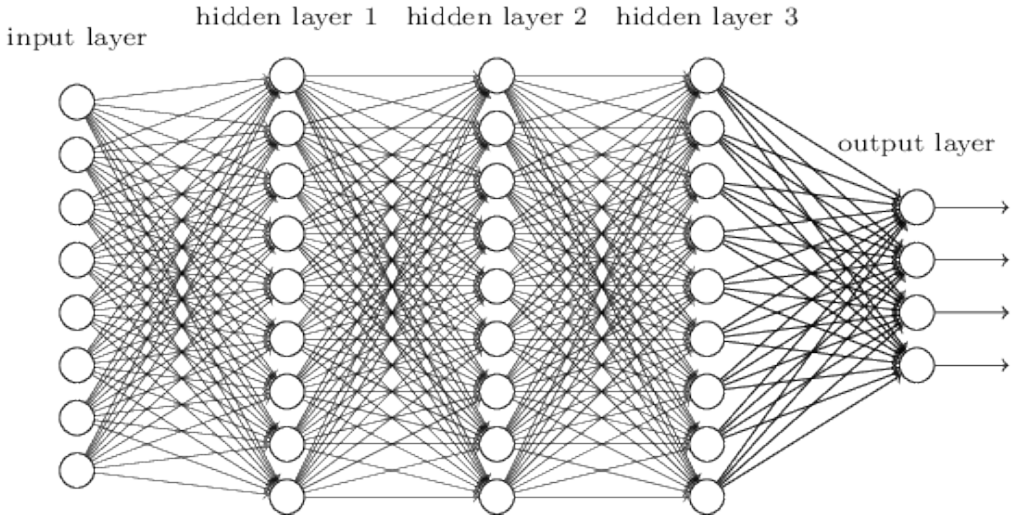
Moreover, an accurate designing of the clinical study, as a pattern recognition task, is crucial to obtain good results concerning classification performance. For this reason, in literature, several algorithms and procedures could be found which are used to design the optimal sets of features, from the extraction approaches to their dimensionality reduction, such as based on relative correlation ranking, principal or independent components extraction [26, 27]. In the last years, a relevant number of studies have been proposed for the classification of breast lesions with classifiers based on radiologists' gold standard labelling. In most cases, they were based on supervised learning approaches using Artificial Neural Networks (ANNs) or Support Vector Machines (SVMs), as well as Swarm Intelligence, simpler Linear Discriminant Analysis (LDA), Decision Tree or Bayes classifiers which are well known mathematical models used to perform classification for different aims [28, 29, 30, 31, 32, 33, 34, 35, 5, 36, 37, 38, 39, 40, 41, 42, 6, 43, 44, 45, 46, 47].

Regarding the classification approach using ANNs, two main classes of Artificial Neural Networks could be identified in the literature, generally depending on the number of hidden layers, which are *Shallow* and *Deep* Neural Networks. According to the commonly accepted differentiation, ANNs with a single layer are named shallow neural networks, whereas a deep architecture has a number of hidden layers greater than one [48]. A representation of the two topologies is represented in Fig. 3. This discrimination in the ANNs topology is particularly important since the design of the optimal architecture for Artificial Neural Networks, in terms of the number of hidden layers and the number of neurons per layer, is still an open problem today, which strongly affect the classification performance [49].

Recent studies have evaluated the performance differences between shallow and deeper neural networks by highlighting the strengths and weaknesses of both architectures [50, 51]. According to literature, there is not an objective motivation to prefer Deep or Shallow Neural Networks. In fact, both the neural architecture could approximate any (reasonable) function, where the quality of the final generalisation properties strictly depend on the signif-



(a)



(b)

Figure 3: Architectural differences between (a) shallow and (b) deep neural networks. (Adapted from Nielsen [48] under Creative Commons Attribution-Non Commercial 3.0 Unported License)

.

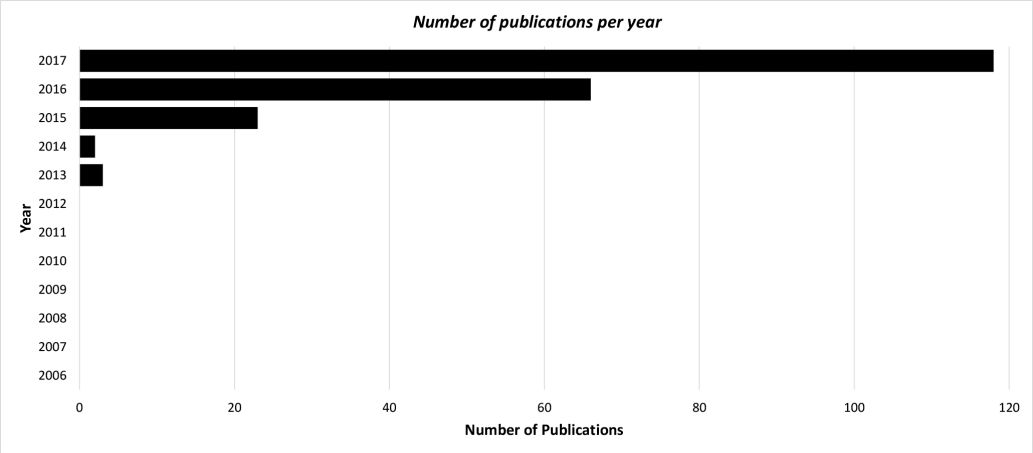


Figure 4: Number of publications per year from 2006 to 2017. Topic: Convolutional Neural Network & Medical Imaging. Indexes: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI.

icance and classes-balance of the available training data. However, shallow neural networks could reach an extremely high number of neurons in their hidden layer, leading to very wide ANNs, thus making the number of parameters to be tuned during the training phase considerably high, with the risk of data overfitting [52]. On the other side, the introduction of multiple layers makes ANNs able to learn features at different levels of abstraction, based on the number of hidden layers, leading to stronger capabilities of generalisation if compared to a shallow architecture with the same computing power, in terms of number of neurons and connections.

Regarding deep neural networks, different learning strategies and architectures have been introduced in the literature so far; these automatic learning systems are having a huge success mainly thanks to the use of (deep) Convolutional Neural Networks (CNNs) and Deep Learning algorithms in the field of image processing and classification. In fact, these kind of architectures are able to make a decision (i.e. classify) working directly on a raw image used as input to the network [53, 54, 55]. A CNN, in fact, has the capability to automatically extract some descriptors from an image, the so-called *feature learning capability*, thus eliminating the need to develop image processing algorithms aiming to the extraction of *hand-crafted* features necessary to a traditional feature-based classifier, such as ANN or SVM [54, 56, 57].

In recent years, Deep Learning (DL) and Convolutional Neural Networks

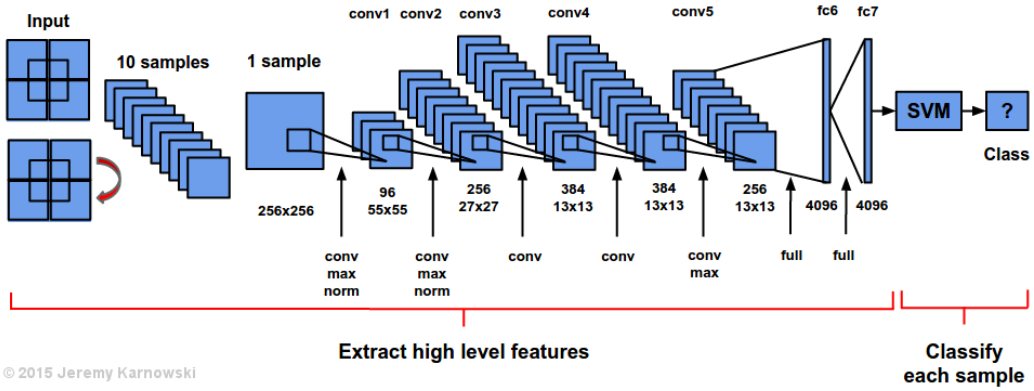


Figure 5: The implementation of Alexnet CNN combined with SVM classifier [60]. This architecture is composed of 5 convolutional layer, some of them coupled with max-pooling and normalization layers, and 2 fully-connected layers. Finally, the SVM classifiers is used for class discrimination.

(CNNs) have been used in many applications for the segmentation and classification of images, including the medical field, with more than 200 works dealing with the topics of interest published in the last 5 years (Fig. 4), having half of these works focused on breast cancer.

For example, Samala *et al.* designed a DL-CNN architecture for breast micro-calcification classification [58]. The authors found the optimal DL-CNN architecture by varying among 216 combinations of parameters in the network (e.g., the number of filters and the filter kernels) and analysing the effects of their variation in the parameter space. The comparison between the selected DL-CNN and their previously designed CNN, which was a non-deep learning Artificial Neural Network performing convolution on the input images, showed a statistically significant improvement since the Areas Under the Curve (AUCs) of the CNN and DL-CNN were 0.89 and 0.93 respectively.

Another interesting work in the breast cancer prevention field is the one performed in [59]. In fact, Kallenberg *et al.* presented an innovative method capable of learning from features at multiple scales hierarchy from not labelled data addressing two different tasks: (i) breast density segmentation and (ii) scoring of mammographic texture. The authors reported that the scores obtained by using the proposed method, which was based on automatic learning, had a high correlation with the ones obtained with the manual approach. Furthermore, the learned texture scores were predictive of breast cancer.

Unlike traditional neural architectures, a Convolutional Neural Network may have different kinds of layers, which are generally combined in different ways depending on the specific implementation [56]. Specifically, the three main types of layers to build these architectures are Convolutional Layer, Pooling Layer, and Fully-Connected Layer. An example of CNN architecture is reported in Fig. 5, which show the AlexNet implementation combined with a SVM classifier [60].

According to literature, Convolutional Neural Networks are powerful architectures that may be used in three different ways [61, 56, 54]:

- **Training from scratch**: as for ANNs, Convolutional Neural Networks may be created from scratch, designing the overall architecture and providing enough samples as input for training. Generally, this process takes a lot of time and computational resources using large datasets with several classes.

- **Transfer Learning or Fine-Tuning**: this approach allows using an available pre-trained model for classification purposes different from the original target classes. In details, it is possible to fine-tune the classification layer of a CNN to predict new classes given as input. Thanks to the power and versatility of these architectures, it is possible to fine-tune all the layers of the network, or just some of them maintaining the weights of the other layers. In particular, some authors suggest to fine-tune the higher levels of networks, due to the higher generality of the features computed in the lower levels of the networks [56].

- **Features Extractors**: in addition to the previous ways for CNN implementation, it is possible to get the output at a specific level of the CNN and use it as automatic descriptors of the input data of the network itself. Since this process is iterative, it is possible to intercept the output at the desired level, based on the desired level of abstraction of the features.

In a previous work, several pre-trained deep models coupled with non-neural classifiers were investigated for classifying breast lesions, considering images from digital tomosynthesis [62].

In this paper, two different approaches have been considered for the classification of breast lesions; in particular, after image pre-processing and extraction of ROIs containing lesions, two neural architectures were used for

classification. In the first approach, morphological and textural features have been extracted from each ROI. An Artificial Neural Network was used to discriminate among the classes of lesion considering the hand-crafted previously mentioned features. The obtained results are then compared with those obtained following a Deep Learning approach based on Convolutional Neural Networks, where well known pre-trained CNN models have been evaluated as feature extractors by comparing the classification performance of several non-neural classifiers cascading to the CNN.

The paper is organized as follows: patients, image acquisition system and the algorithms for image processing are organized as materials and then are discussed in Section 2; the two different frameworks considered for the classification of breast lesions are described in Section 3; experimental results are reported and discussed in Section 4, and finally, the conclusions are presented in Section 5.

2. Materials

In this section, details about patients, acquisition protocol, images and algorithms for image processing are given.

2.1. Patients

From January to November 2016, 16 patients underwent breast tomosynthesis examination. All the patients were women aged between 35 years and 65 years (average age 49.8 $\pm$ 9.2 years). In particular, some women underwent more than a single tomosynthesis examination, thus reaching a dataset composed of 39 breast exams.

2.2. Acquisition System: Protocol and Images

In DBT, several images of the compressed breast are obtained projecting X-ray from different emission sources [63]. Starting from these images, a reconstruction algorithm is performed to produce a three-dimensional view of the breast tissue, slice by slice, mutually parallel and suitably spaced.

The acquisition system used for this study is the 2nd generation DBT Giotto Tomo with the emission sources ranging in a wide scanning angle of 40°, as shown in Fig. 6. In particular, an innovative iterative reconstruction algorithm is used, which have been designed exclusively for tomosynthesis systems, thus it is not derived from CT or MR techniques.

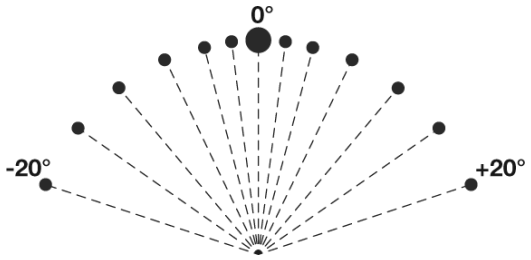


Figure 6: Emission angles distribution and doses by DBT Giotto Tomo (http://www.tomosynthesis-giotto.com/). The larger the dot, the greater the dose at the corresponding angle.

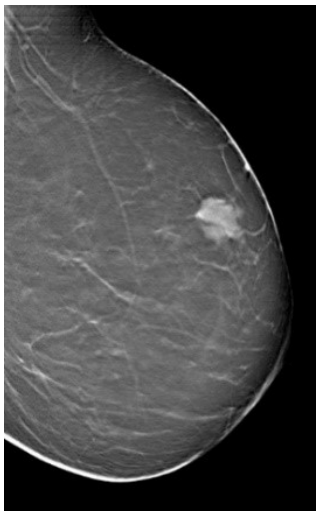


Figure 7: A slice extracted from a reconstructed image.

Despite the wide acquisition range, only 13 exposures with reduced optimised dose were used, as represented in Fig. 6, and the pixel size of 8 μm allowed to obtain high resolution images without binning.

An example of a slice from the reconstructed image is shown in Fig. 7.

2.3. Image Processing

The extraction of ROIs was performed following the success of previous empirical strategies [64] and it is described in the following paragraphs.

2.3.1. Image pre-processing

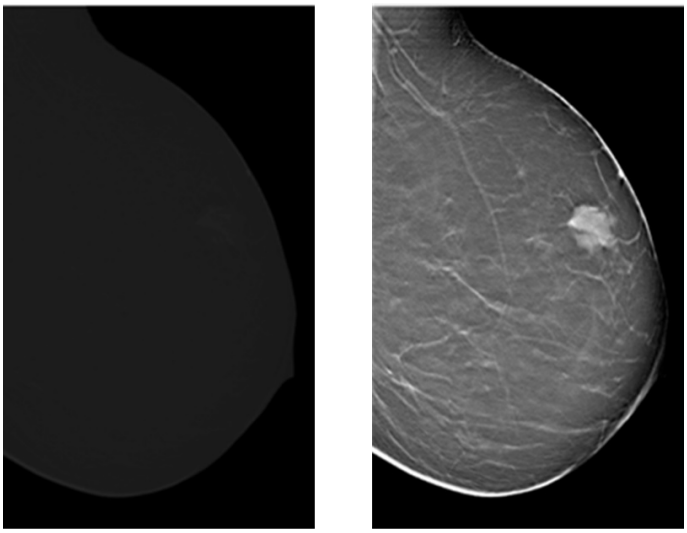
The first step of the pre-processing procedure started windowing and enhancing the contrast of the DICOM images [65]; it was performed following the algorithm reported in Alg. 1.

The proposed algorithm converted images by a linear transformation, considering the *Window Width* (w) and *Window Centre* (c) where x is the value of a pixel from the input DICOM image, y is the value of the corresponding pixel in the output image, y_min and y_max are the bounds of the range for pixel value in the final image (e.g. [0 - 255]). Fig. 8 shows the result of the windowing operation. The values of c and w were obtained empirically, based on the images acquired by the specific input device.

Algorithm 1 The proposed algorithm for image windowing.

$K \leftarrow \frac{c-0.5-(w-1)}{2}$
if $x \leq K$ **then**
  $y \leftarrow y_{min}$
else
  if $x > K$ **then**
    $y \leftarrow y_{max}$
  else
    $y \leftarrow \frac{x-(c-0.5)}{(w-1)+0.5} * (y_{max} - y_{min}) + y_{min}$
  end if
end if

After this preliminary operation, enhancement of the contrast, median filtering and border removal were sequentially performed to enhance the ROIs to be processed in the subsequent steps, as represented in Fig. 9. The enhancement of the contrast was applied on the image resulting from the previous windowing operation to map the intensity value of the pixels to new values such that 1% of the data was saturated in both low and high intensities. Then, a median filter was applied to remove the granular noise, also known as "salt & pepper", from the image using a 3x3 squared kernel [66]. Finally, the removal of the breast external border was performed; this last step was of fundamental importance in the pre-processing phase since the breast showed a white edge that could interfere in the subsequent segmentation phase. This last operation was performed expanding the background (the black part of the image) using a disk with 16-pixel radius.



(a) (b)

Figure 8: The Windowing result; (a) shows the original DICOM image; (b) shows the modified image (*Window Centre* was 3477, *Window Width* was 844).

2.3.2. Segmentation

After the enhancement of the acquired images, the extraction of ROIs containing suspicious lesion was performed in the segmentation step, which generally consists in several algorithms leading to the labelling of each pixel, according to the analysed characteristic [67, 68].

2.3.2.1 Image Binarisation

In grey levels tomosynthesis images, breast lesions appear brighter than the surrounding tissues. An adaptive thresholding algorithm was chosen to perform the binarisation task, according to a previous approach used to segment regions for better detecting benign or malignant masses in mammography [69] or retinal vessel extraction [70]. The proposed strategy was adaptive, in fact, the optimal value of the selected threshold allowed to minimise both the risk to remove meaningful regions and the searching subspace of the whole breast image. In details, the threshold was empirically determined by adding the mean value of the grey level values to their standard deviation multiplied by an empirical constant.

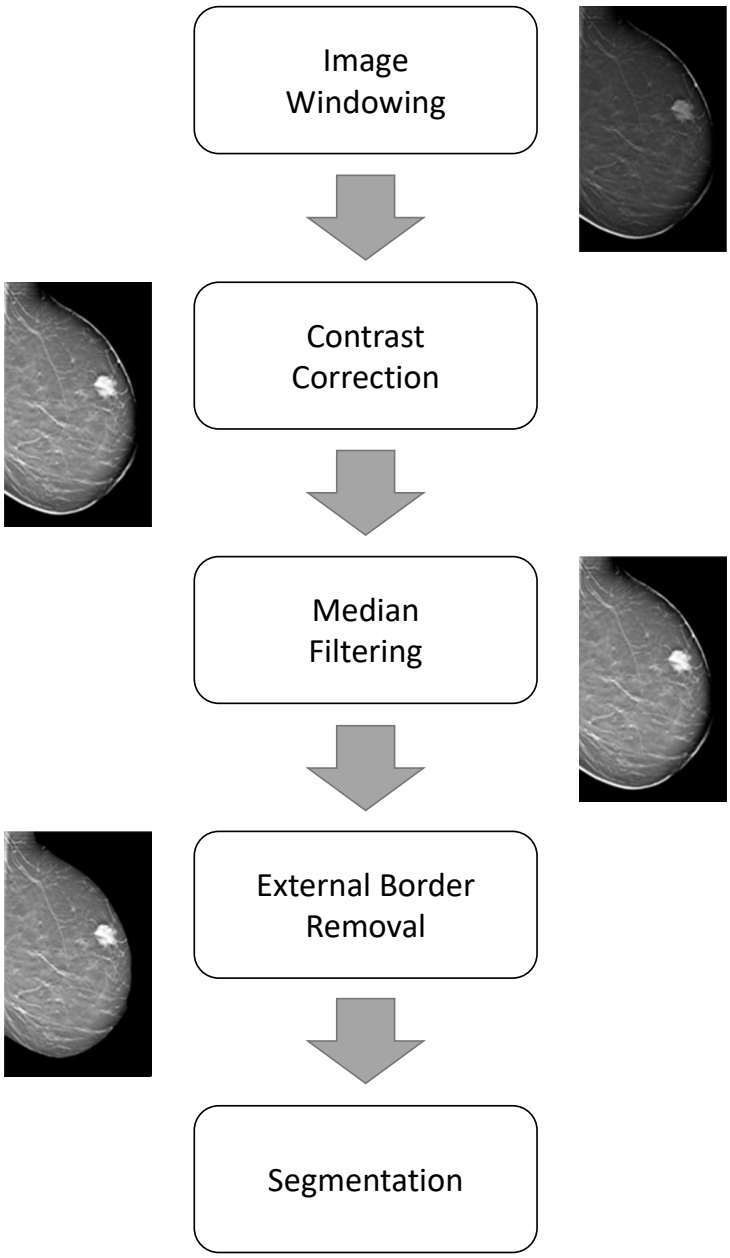


Figure 9: A schematic representation of the steps needed in the image pre-processing phase leading to the enhancement of the input images.

2.3.2.2 Hole Filling and Post-Processing

Finally, since the remaining regions included some blood vessels that could interfere in the subsequent classification step, a procedure based on the aspect ratio of the remaining areas was used to remove them [71, 30]. The approach used to select ROIs containing suspicious lesions is represented in Fig. 10.

In details, areas with aspect ratio below a threshold were removed, thus allowing the removal of all ROIs representing vessels and not suspicious areas. At the end of the segmentation phase, a dilatation of each segmented ROI was performed and the mean values of grey levels in both the initial and the extended ROIs were evaluated; if the difference between the obtained two values was higher than a threshold, or the grey levels of the two areas were significantly different, the segmented ROI was left as it was; otherwise, an active contour procedure was performed, considering the border coming from the binarisation process as the initial curve for the algorithm [72, 73].

2.3.3. Images Extraction

The goal of the previous segmentation step was to build an exhaustive and robust dataset of ROIs to design the classification workflow of the different supervised classifiers. In order to comply with the input size requirements of the considered pre-trained CNN models, a preliminary resizing of the ROIs was performed, obtaining images sized 227x227 [74, 54].

Finally, all the images were labelled according to the classification of the radiologists in the previous mentioned 4 classes, which are:

1. None: segmented ROI not containing any kind of lesion (Fig. 11(a));
2. Ori: segmented ROI containing an irregular opacity (Fig. 11(b));
3. Oro: segmented ROI containing a regular opacity (Fig. 11(c));
4. Ost: segmented ROI containing a stellar opacity (Fig. 11(d)).

3. Classification Frameworks

This section describes the two frameworks designed to classify the extracted ROIs containing candidate breast lesions: the first framework proposes the adoption of feed-forward ANNs, whereas the second one uses CNNs coupled with a non-neural classifier.

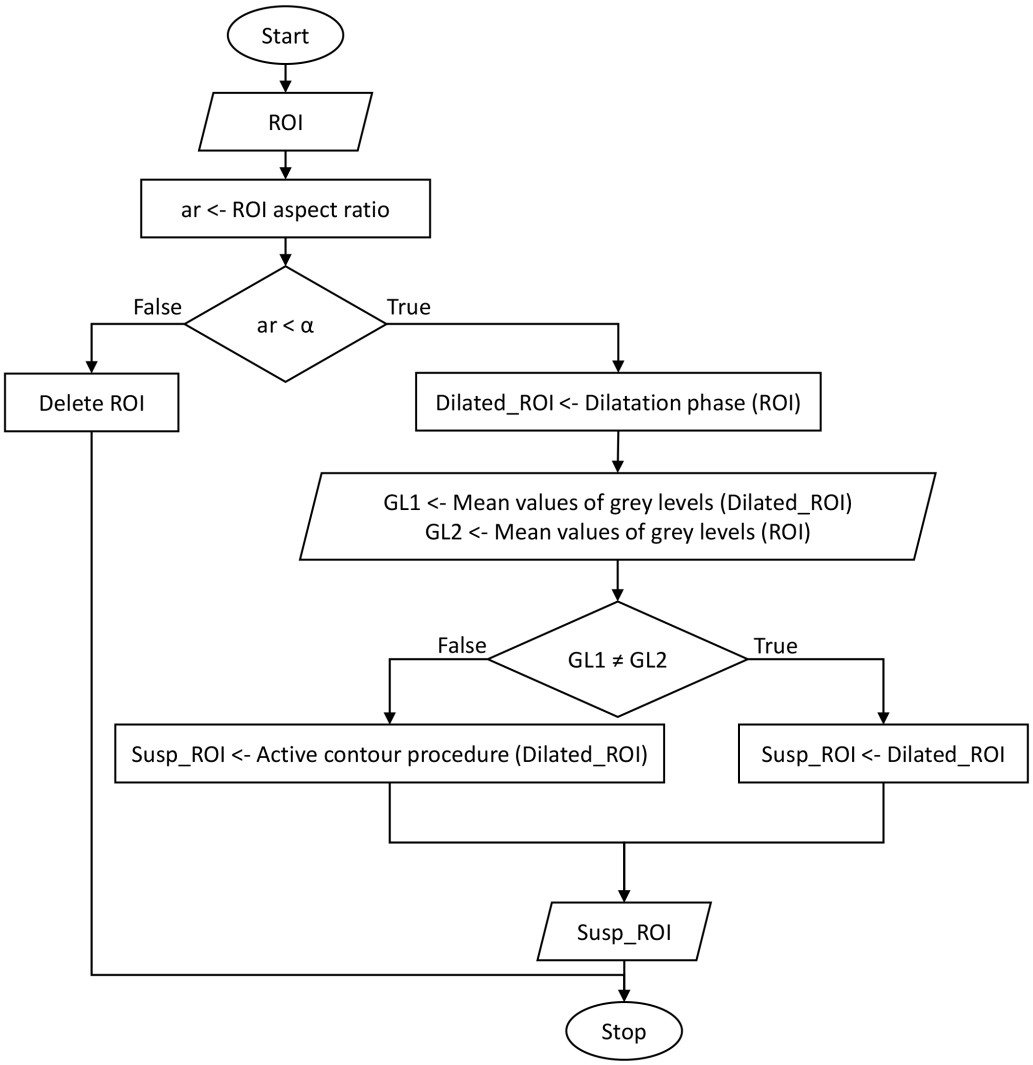
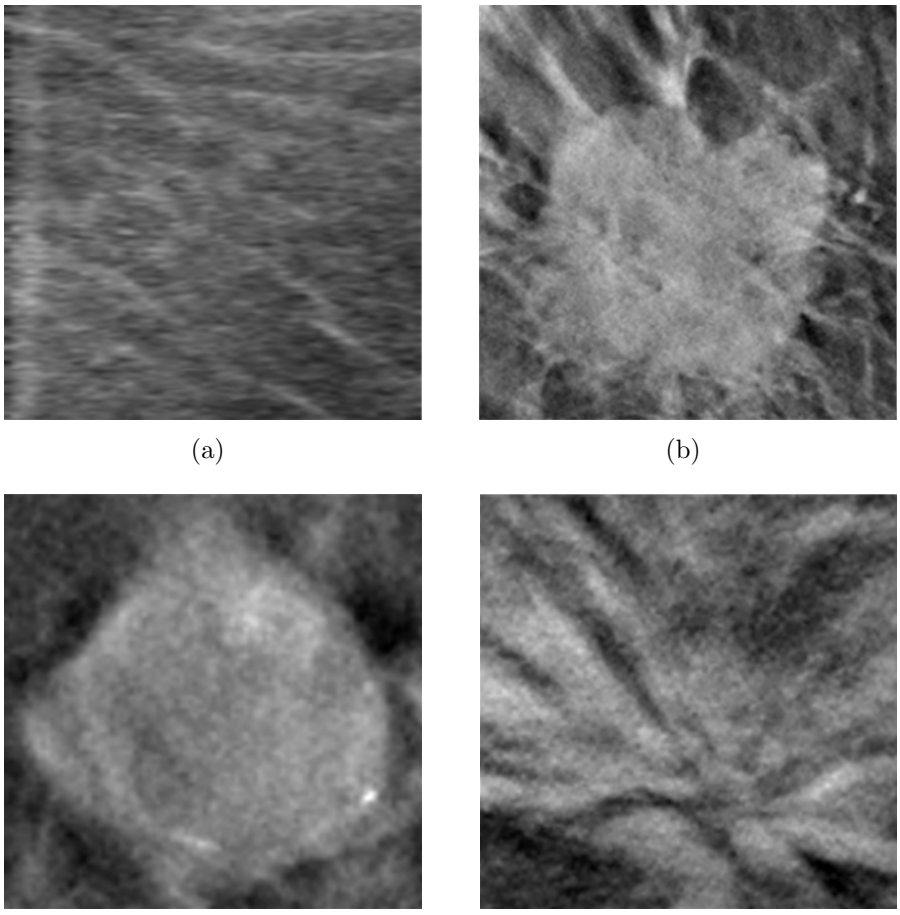


Figure 10: A block-diagram showing the algorithm for the ROIs extraction.



(a) (b)

(c) (d)

Figure 11: Images extracted after the segmentation phase: (a) ROI with no lesions; (b) ROI with irregular opacity; (c) ROI with regular opacity; (d) ROI with stellar opacity.

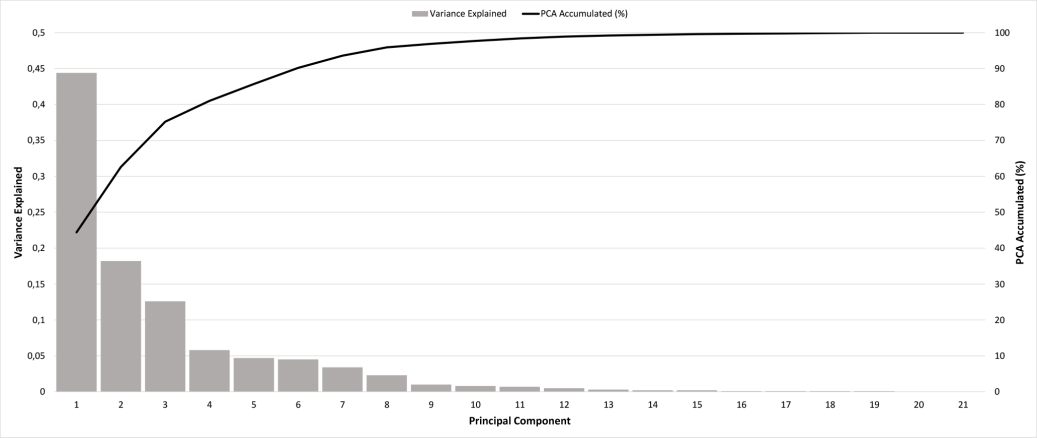


Figure 12: PCA explained variance of the principal components computed from the initial dataset. Bars show the variance for each component, whereas the black line indicate the cumulative variance for each principal component.

In particular, the workflow of the first framework is based on the extraction of some hand-crafted features from the ROIs, and then addresses the problem as a traditional ANN-based supervised pattern recognition. On the other side, the second approach considers CNNs as feature extractors with a non-neural classifier cascading to discriminate among the classes of the four lesions.

3.1. First Framework with Artificial Neural Networks

The ROIs extracted with the algorithm for segmentation were described using both morphological and textural features computed using the Grey Level Co-occurrence Matrix (GLCM), which have been used in similar classification problems [29, 75, 76, 77]. In particular, 26 different features were computed and the arithmetic mean of pixels' grey levels in the region was also considered.

In order to reduce the number of features to be considered for the subsequent classification, dimensionality reduction was performed using the eigendecomposition Principal Component Analysis (PCA) [78, 79, 80]; finally, a set of 13 features in the new subspace was selected to cover an overall variance of 99.3%, as could be seen in Fig. 12.

Moreover, before performing PCA, each feature was normalized to zero mean and unitary standard deviation, by using the z-score remapping [81].

In this framework, two different classifications based on ANNs were considered:

- **binary classification** discriminating between the Positive class, containing images of three classes of lesions, i.e. Oro, Ori and Ost, and the Negative class, with no-lesion images;
- **multi-class classifier** to discriminate among all the considered classes of images, i.e. Oro, Ori, Ost and no-lesion.

After the preliminary data preprocessing, the obtained dataset was constituted by: 63980 samples for None class, 391 samples for Ori class, 654 for Oro class and 480 samples for Ost class.

3.1.1. Binary Classification

The considered dataset was considerably unbalanced since about 64000 entries were labelled as Negative, whereas the Positive samples were 1525 only. Since imbalance between the classes generally leads to low classification performance, thus leading to the need to develop specific procedures for improving classification performance, the number of entries labelled as Negative was reduced to 1670, removing the most correlated entries, having a correlation factor higher than 0.9 [82, 17, 83].

After dataset rebalancing, an optimised Artificial Neural Network was designed. Since evolutionary strategies are suitable approaches for the optimisation of classifiers, the optimal topology for the considered classifier was found by means of an evolutionary approach based on a mono-objective Genetic Algorithm (GA) [30, 84, 43]. The used GA searches for the ANN topology reaching the highest mean accuracy on the test set [85, 86, 87], where each ANN in the GA population was trained, validated and tested considering a dataset split into 60/20/20 respectively for training, validation and test sets (with a number of random permutations fixed to 200).

The optimisation algorithm found a shallow ANN whose topology consisted of 143 neurons for the single hidden layer with the log-sigmoid transfer function (*logsig*) and one neuron in the output layer with the hyperbolic tangent sigmoid transfer function (*tansig*). The input dataset was the balanced PCA dataset obtained as reported in the previous paragraph and, at each iteration, random samplings were performed to maintain balanced the ratio between the two considered classes.

3.1.2. Multi-class Classification

After the binary classification approach, the Genetic Algorithm in Bevilacqua *et al.* was modified to perform multi-class classification [30]. In details, the previous version of the GA was modified to describe ANN topologies with the output layer set to *softmax*, allowing the discrimination of multiple classes, rather than two.

As for the previous approach, each ANN was trained, validated and tested considering both the same partitions of the dataset and the same number of iterations. In this case, the returned ANN optimal topology was composed of two hidden layers with 248 and 46 neurons respectively with hyperbolic tangent sigmoid transfer function (*tansig*) for both the hidden layers where the output layer had 4 neurons with *softmax* transfer function.

3.2. Second Framework with Convolutional Neural Networks

Unlike the traditional approach described in the previous section, which is based on the classification considering hand-crafted features computed by processing the ROI extracted using the segmentation algorithm, Convolutional Neural Networks can perform classification taking images as input.

As already reported in Sect. 1, CNN classifiers may be used in different ways; in this work, CNNs have been used as automatic Features Extractors considering their output as input for non-neural classifiers. In order to use CNNs as Feature Extractors for the subsequent learners, the following workflow was adopted:

1. loading the pre-trained model;
2. loading the dataset;
3. dataset preparation:
    a. class balancing;
    b. training and test sets creation, considering 2/3 of samples for the training set, and the remaining samples for the test set, maintaining balanced the number of samples for each class;
4. features extraction using batch approach due to memory constraints (32 images per batch);
5. features normalization using Z-Score algorithm, if needed;

6. training of the non-neural classifier;

7. test of the classifier.

In order to evaluate the variability of the obtained results and the mean performance of the classifiers, 50 iterations of training and test have been performed, as for the described traditional approach. In particular, the sequence from (3.) to (7.) represents a single iteration.

As the number of samples is not equally distributed among the considered classes, data augmentation is an essential step to both increase the number of samples of classes with few images and train the desired invariance and robustness properties [88]. This preliminary step in the augmentation of the input images by means of specific transformations, such as rotation, translation, skewing and distortions [89]. Since the number of Positive samples (Oro, Ori and Ost classes) was lower than the Negatives, a combination of flips and rotations (0°, 90°, 180°and 270° respectively) was used to augment the Positive dataset; in this way, for each Positive image sample, 8 images were obtained.

In the next section, more details on the pre-trained models evaluated in this work are provided.

3.2.1. Pre-trained Models and Learners

Several pre-trained models were considered and evaluated for features extraction using CNNs. For this aim, the classification layer and the related fully-connected layer, were removed in order to obtain a set of features describing images automatically computed. The following models were considered:

- **GoogLeNet** [90]: the main feature of this net is the *Inception Module* that helps to reduce the total number of parameters of the net. It also uses average pooling layers instead of full-connected layers at the end of the net and it keeps only the last full-connected layer before the classification one.

- **ResNet** [91]: this net is characterized by skip connections and makes an extensive use of batch normalization. Skip connections help to simplify the training phase thanks to several residual blocks that directly propagate information from a bottom layer to a distant higher one; the gradient back-propagation is facilitated too, by allowing the gradient

to reach lower layer without losing magnitude. More details on skip connections are reported in the related work.

- **AlexNet** [54]: its architecture is similar to LeNet [92] but it is deeper and bigger. Another difference compared to LeNet (that has every single convolutional layer followed by a pooling layer) is the multiple convolutional layer stacked on the top of each other before the pooling layer.

- **VGG-verydeep (VD)** [93]: it consisted of two nets with different number of weighted layers: 16 and 19, respectively. In this work, VGG-Net with 19 layers and 144 million parameters was considered. This network is composed of several convolutional layers with 3-by-3 receptive fields and different numbers of channels.

- **VGG-F, M and S** [94]: three CNN models, representative of the state of the art, were proposed in the related work. Each model represents a different accuracy/speed trade-off. In details, VGG-F is the Fast model and it is based on the architecture of Krizhevsky *et al.* [54]; VGG-M is the Medium model and it is very similar to CNN developed by Zeiler and Fergus [95]; finally, VGG-S is the Slow model and its architecture is based on the accurate network from Overfeat package [96].

The output of the CNN models (or rather, their final activations) was used to train a new classifier. In this case, several non-neural learners were considerd and evaluated, which are:

- Linear Support Vector Machine (Linear SVM) [97];
- K-Nearest Neighbor (KNN) [98];
- Naïve Bayes (NBA) [99];
- Decision Tree (DT) [100];
- Linear Discriminant Analysis (LDA) [101, 102].

The performances for all the described approaches are reported in the following section.

4. Experimental Results

In this section, all the results obtained in the classification sessions are reported, for both the introduced framework. In particular, a comparison between binary and multi-class classification approaches is presented in the first subsection, whereas the results considering all the CNN pre-trained models combined with the different learners are reported in the subsequent paragraph.

4.1. First Framework Performance

Regarding the binary classification, the results are reported in terms of Accuracy (Eq. 1), Specificity (Eq. 2) and Sensitivity (Eq. 3) on the test set, as mean values on the iterations performed according to what already discussed in Sect. 3.1.1. The number of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) samples are evaluated considering the confusion matrices (Tab. 1) obtained from the classification on all the test sets created by permuting the dataset, where Positive and Negative classes have been already defined.

Table 1: Confusion Matrix

		True Condition	
		Positive	***Negative***
Predicted Condition	***Positive***	*TP*	*FP*
	Negative	*FN*	*TN*

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Specificity = \frac{TN}{FP + TN} \tag{2}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{3}$$

As reported in Table 2, more than 80 % of Accuracy, Sensitivity and Specificity have been reached for the approach based on the binary classification. On the other side, the optimized ANN designed for the multi-class

classification reached a mean Accuracy of 74.84 %±4.89, where the Accuracy have been evaluated as the ratio between the number of instances correctly classified and the total number of instances.

Table 2: Results for binary classification

	Accuracy	**Specificity**	**Sensitivity**
Mean	*84.19 %*	*82.82 %*	*85.90 %*
Standard Deviation	*3.06*	*5.33*	*5.17*

Comparing the obtained values of Accuracy, it is evident that the binary classifier is more accurate and stable than the multi-class classifier. As expected, the multi-class approach showed worse performance than the binary case. This behaviour was predictable and reasonable considering the way in which ANNs were designed and optimised: the extraction process of the most discriminant features heavily influences the overall capabilities of the classifier. In this case, the features used as inputs showed good results to discriminate binary class samples (presence or absence of lesions), whereas these did not provide enough information to correctly describe all the different kinds of lesions in the multi-class approach.

4.2. Second Framework Performance

In this section, the results obtained following the approach discussed in Sect. 3.2, using Convolutional Neural Networks as features extractors and non-neural learners, are reported. In particular, some preliminary tests were firstly performed and evaluated to find the optimal working set to improve the overall performance of the classifiers.

4.2.1. Activations Normalization

Firstly, all the CNN models, coupled with each learner, were evaluated comparing the classification performance considering the normalisation of the activations obtained from the CNN which are then given as input to the learner. As it could be seen in Table 3, which reports the best results among the analysed learner, the normalization of the activations extracted from all the models does not significantly affect accuracy. However, a slight improvement can be observed in almost all cases.

Table 3: Results obtained from two tests performed with the KNN classifier, trained on different models activations. Accuracy was averaged over 50 iterations using GPU and augmented images. The only difference between tests was the activations normalisation.

	No - Normalisation		**Normalisation**	
Nets	***Accuracy (%)***	***Std***	***Accuracy (%)***	***Std***
GoogLeNet	*80.79*	*0.58*	*81.43*	*0.62*
ResNet	*89.70*	*0.54*	*90.19*	*0.36*
AlexNet	*90.11*	*0.52*	*90.12*	*0.49*
VGG-VD	*89.78*	*0.48*	*89.65*	*0.39*
VGG-F	*91.44*	*0.4*	*91.61*	*0.42*
VGG-M	*90.62*	*0.52*	*90.77*	*0.51*
VGG-S	*91.91*	*0.36*	*92.02*	*0.51*

4.2.2. Images Augmentation

As for the normalisation approach, all the CNN models were evaluated comparing the classification performance considering the dataset modified with the augmentation approach. Results are reported in Table 4, which shows the performance obtained considering the KNN classifier, as for the normalisation comparison. Unlike the previous test, the improvement achieved with the augmented images is clearly evident. The accuracy improvement is approximately 10 % for VGG nets (VD, F, M and S) and AlexNet while it is higher (about 15 %) for GoogLeNet and ResNet; the standard deviation is reduced also showing a more stable behaviour of the CNN models.

4.2.3. Learners Comparison

The CNN models performing better in the tests described in the previous section were then used with different learners trying to classify the different kinds of lesions. Except for GoogLeNet, all the CNN models reached accuracy near to 90 %. Final tests to evaluate the performance of non-neural classifiers were performed on a subset of the considered CNNs; in particular, VGG-F, VGG-S and VGG-S were used for final evaluations, as they reported the higher mean accuracy and minimum processing time in the previous tests.

In order to perform the final tests, both dataset augmentation and activations normalisation were performed. In details, after the augmentation of the dataset for the samples belonging to Oro, Ori and Ost classes, the number of samples for each class was equalled to the number of samples in Ori class,

Table 4: Results obtained from two tests performed with the KNN classifier, trained on different models activations. Accuracy was averaged over 50 iterations using GPU and normalization. The only difference between tests was the augmentation of images.

	No - Augmentation		**Augmentation**	
Nets	***Accuracy (%)***	***Std***	***Accuracy (%)***	***Std***
GoogLeNet	*66.16*	*1.82*	*81.43*	*0.62*
ResNet	*75.61*	*1.90*	*90.19*	*0.36*
AlexNet	*80.62*	*1.77*	*90.12*	*0.49*
VGG-VD	*78.73*	*1.93*	*89.65*	*0.39*
VGG-F	*81.49*	*1.50*	*91.61*	*0.42*
VGG-M	*80.93*	*1.51*	*90.77*	*0.51*
VGG-S	*82.60*	*1.55*	*92.02*	*0.51*

Table 5: Results of the selected pre-trained CNNs used as features extractor, training the learners with normalization and augmented images. Results are reported in terms of mean value of accuracy (± standard deviation) on the test set.

	KNN	***LDA***	***LINEAR SVM***	***NAÏVE BAYES***	***DECISION TREES***
VGG-F	*91.63 ± 0.41*	*64.57 ± 0.66*	*67.29 ± 2.02*	*43.82 ± 0.59*	*59.68 ± 1.07*
VGG-M	*90.74 ± 0.48*	*66.25 ± 0.60*	*69.50 ± 2.16*	*42.85 ± 0.57*	*57.03 ± 0.98*
VGG-S	*92.02 ± 0.48*	*65.24 ± 0.80*	*68.84 ± 1.89*	*44.89 ± 0.60*	*56.16 ± 0.93*

which was the lowest one, by randomly removing samples from the other classes at each iteration. In order to evaluate the reliability and stability of the results, the training and test of the CNN coupled with the classifiers was performed 50 times, permuting the input dataset at each iteration. The input dataset (at each iteration) was constituted by 12512 images, equally distributed among the 4 classes; the training set was the 66 % (2064 samples per class) of the whole dataset, whereas the remaining 34 % was used as test set (1064 samples per class).

The obtained results are reported in Table 5, which shows the mean accuracy (± standard deviation) obtained by each classifier working on the set of features extracted by each CNN of the considered subset.

As reported in Table 5, the Naïve Bayes classifier was not recommended in this classification as it shows the worst mean accuracy; Decision Trees lead to an improvement of the mean accuracy in comparison to Naïve Bayes,

Table 6: Sensitivity and Specificity for the lesions evaluated through 1-vs-all approach.

	Ori vs all		**Oro vs all**		**Ost vs all**	
	Sensitivity	***Specificity***	***Sensitivity***	***Specificity***	***Sensitivity***	***Specificity***
VGG-F	*98.67 %*	*97.07 %*	*96.01 %*	*95.98 %*	*97.24 %*	*96.93 %*
VGG-M	*98.14 %*	*96.64 %*	*95.00 %*	*95.76 %*	*97.18 %*	*96.62 %*
VGG-S	*98.36 %*	*97.13 %*	*95.61 %*	*96.33 %*	*96.67 %*	*97.25 %*

Table 7: Confusion Matrix of the best classifier (accuracy 93.26% with VGG-S and KNN)

		True Condition			
		Oro	***Ori***	***Ost***	***None***
Predicted Condition	***Oro***	*1017*	*10*	*9*	*91*
	Ori	*16*	*1051*	*3*	*51*
	Ost	*12*	*1*	*1033*	*54*
	None	*19*	*2*	*19*	*868*

but it was far from being considered as a reliable classifier, whereas both the linear classifiers (SVM and LDA) further increased the performance. Despite these results, mean accuracy was lower than the expected value.

On the other side, different outcomes were obtained from the KNN classifier, where, in addition to the accuracy reported in Table 5, detailed results about sensitivity and specificity are reported in Table 6. In this case, the mean performance shows high values of accuracy, specificity and sensitivity which are higher than 90 % independently from the CNN used.

Moreover, the drop in the overall accuracy, if compared to the values of sensitivity and specificity which are higher than 95 %, is justifiable and predictable analysing the confusion matrix reported in Table 7, which is related to the best classifier obtained during the final tests (accuracy 93.26 % with VGG-S and KNN). In fact, it shows a high number of misclassifications involving the "none" class, leading the classifier to have a high number of false positives impacting negatively on the overall performance.

5. Conclusions

In this work, two different frameworks to support radiologists in classifying breast tomosynthesis images are discussed and quantitatively compared. Both frameworks were based on supervised learning strategies and

used features extracted from the images: while the first framework is based on optimised Artificial Neural Network topologies for the classification of hand-crafted features extracted by an ad-hoc procedure, the second framework consider and compare different non-neural learners processing features automatically extracted by Convolutional Neural Network models.

In details, in case of traditional Artificial Neural Networks, morphological and textural features were extracted after the processing and segmentation of the input images; subsequently, the optimal topology for both binary and multi-class neural classifiers were designed through an evolutionary approach based on a mono-objective genetic algorithm.

On the other side, several CNN models were tested as feature extractors using the activation of each considered last layer as input for different learners. Regarding the training of CNNs coupled with learners, it was found that the activation normalisation is useful to obtain slight performance improvement, whereas images augmentation is necessary to improve the classification performance significantly, especially when the number of samples per class in unbalanced.

After the evaluation of different CNN pre-trained models and non-neural classifiers, VGG-F, VGG-M and VGG-S were considered for the final tests as they showed the best performance coupled with the KNN classifier.

The reported results show that the approach based on CNNs as features extractors is very useful and powerful for this kind of classification problem, where the extraction of meaningful features able to discriminate more than two classes could be a problematic approach. In fact, according to the presented literature, the use of convolutional classifier for the extraction of image descriptors is a very promising approach if compared to the design of effective strategies for the extraction of hand-crafted features for classification purposes.

Thanks to the innovation introduced by these models, and to the simplification in the design and implementation of automatic decision support systems, future works will examine the reliability of such systems as feature extractors to support clinical diagnosis also for other body districts.

Essential benefits will certainly be evident from the physicians' point of view, in terms of both time (hence cost for the national sanitary system) and diagnostic reliability. In addition, patients will also benefit from these systems, thanks to which innovative techniques, predictive of the pathological course, can be developed, thus exposing the patients to lower risks.

References

[1] J. Ferlay, I. Soerjomataram, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D. M. Parkin, D. Forman, F. Bray, Cancer incidence and mortality worldwide: sources, methods and major patterns in globocan 2012, International journal of cancer 136 (5).

[2] M. Movahedi, S. Haghighat, M. Khayamzadeh, A. Moradi, A. Ghanbari-Motlagh, H. Mirzaei, M. Esmail-Akbari, Survival rate of breast cancer based on geographical variation in iran, a national study, Iranian Red Crescent Medical Journal 14 (12) (2012) 798.

[3] E. Nooshinfar, A. Safaroghli-Azar, D. Bashash, M. E. Akbari, Melatonin, an inhibitory agent in breast cancer, Breast Cancer 24 (1) (2017) 42–51.

[4] P. Cortazar, L. Zhang, M. Untch, K. Mehta, J. P. Costantino, N. Wolmark, H. Bonnefoi, D. Cameron, L. Gianni, P. Valagussa, S. M. Swain, T. Prowell, S. Loibl, D. L. Wickerham, J. Bogaerts, J. Baselga, C. Perou, G. Blumenthal, J. Blohmer, E. P. Mamounas, J. Bergh, V. Semiglazov, R. Justice, H. Eidtmann, S. Paik, M. Piccart, R. Sridhara, P. A. Fasching, L. Slaets, S. Tang, B. Gerber, C. E. Geyer, R. Pazdur, N. Ditsch, P. Rastogi, W. Eiermann, G. von Minckwitz, Pathological complete response and long-term clinical benefit in breast cancer: the ctneobc pooled analysis, The Lancet 384 (9938) (2014) 164 – 172. doi:https://doi.org/10.1016/S0140-6736(13)62422-8.

[5] V. Bevilacqua, P. Pannarale, M. Abbrescia, C. Cava, A. Paradiso, S. Tommasi, Comparison of data-merging methods with svm attribute selection and classification in breast cancer gene expression, BMC bioinformatics 13 (7) (2012) S9.

[6] F. Menolascina, S. Tommasi, A. Paradiso, M. Cortellino, V. Bevilacqua, G. Mastronardi, Novel data mining techniques in acgh based breast cancer subtypes profiling: the biological perspective, in: Computational Intelligence and Bioinformatics and Computational Biology, 2007. CIBCB'07. IEEE Symposium on, IEEE, 2007, pp. 9–16.

[7] A. Vestito, F. F. Mangieri, G. Gatta, M. Moschetta, B. Turi, A. Ancona, Breast carcinoma in elderly women. our experience, Il giornale di chirurgia 32 (1) (2011) 411–416.

[8] D. Saslow, C. Boetes, W. Burke, S. Harms, M. O. Leach, C. D. Lehman, E. Morris, E. Pisano, M. Schnall, S. Sener, et al., American cancer society guidelines for breast screening with mri as an adjunct to mammography, CA: a cancer journal for clinicians 57 (2) (2007) 75–89.

[9] E. D. Pisano, C. A. Parham, Digital mammography, sestamibi breast scintigraphy, and positron emission tomography breast imaging, Radiologic Clinics of North America 38 (4) (2000) 861–869.

[10] H.-P. Chan, J. Wei, B. Sahiner, E. A. Rafferty, T. Wu, M. A. Roubidoux, R. H. Moore, D. B. Kopans, L. M. Hadjiiski, M. A. Helvie, Computer-aided detection system for breast masses on digital tomosynthesis mammograms: preliminary experience, Radiology 237 (3) (2005) 1075–1080.

[11] J. M. Boone, A. L. Kwan, K. Yang, G. W. Burkett, K. K. Lindfors, T. R. Nelson, Computed tomography for imaging the breast, Journal of mammary gland biology and neoplasia 11 (2) (2006) 103–111.

[12] T. Wu, J. Zhang, R. Moore, E. Rafferty, D. Kopans, W. Meleis, D. Kaeli, Digital tomosynthesis mammography using a parallel maximum-likelihood reconstruction method, in: Proc. SPIE, Vol. 5368, 2004, pp. 0277–786X.

[13] K. E. Korhonen, S. P. Weinstein, E. S. McDonald, E. F. Conant, Strategies to increase cancer detection: Review of true-positive and false-negative results at digital breast tomosynthesis screening, RadioGraphics 36 (7) (2016) 1954–1965.

[14] S. Astley, F. Gilbert, Computer-aided detection in mammography, Clinical Radiology 59 (5) (2004) 390 – 399. doi:https://doi.org/10.1016/j.crad.2003.11.017.

[15] K. Doi, Current status and future potential of computer-aided diagnosis in medical imaging, The British journal of radiology 78 (suppl_1) (2005) s3–s19.

[16] K. Doi, Computer-aided diagnosis in medical imaging: historical review, current status and future potential, Computerized medical imaging and graphics 31 (4-5) (2007) 198–211.

[17] M. A. Mazurowski, P. A. Habas, J. M. Zurada, J. Y. Lo, J. A. Baker, G. D. Tourassi, Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance, Neural networks 21 (2) (2008) 427–436.

[18] J. Hsieh, et al., Computed tomography: principles, design, artifacts, and recent advances, SPIE Bellingham, WA, 2009.

[19] A. Pizurica, W. Philips, I. Lemahieu, M. Acheroy, A versatile wavelet domain noise filtration technique for medical imaging, IEEE transactions on medical imaging 22 (3) (2003) 323–331.

[20] P. Bao, L. Zhang, Noise reduction for magnetic resonance images via adaptive multiscale products thresholding, IEEE transactions on medical imaging 22 (9) (2003) 1089–1099.

[21] A. Oliver, J. Freixenet, J. Marti, E. Pérez, J. Pont, E. R. Denton, R. Zwiggelaar, A review of automatic mass detection and segmentation in mammographic images, Medical image analysis 14 (2) (2010) 87–110.

[22] K. Nie, J.-H. Chen, J. Y. Hon, Y. Chu, O. Nalcioglu, M.-Y. Su, Quantitative analysis of lesion morphology and texture features for diagnostic prediction in breast mri, Academic radiology 15 (12) (2008) 1513–1525.

[23] L. A. Meinel, A. H. Stolpen, K. S. Berbaum, L. L. Fajardo, J. M. Reinhardt, Breast mri lesion classification: Improved performance of human readers with a backpropagation neural network computer-aided diagnosis (cad) system, Journal of magnetic resonance imaging 25 (1) (2007) 89–95.

[24] W. Chen, M. L. Giger, H. Li, U. Bick, G. M. Newstead, Volumetric texture analysis of breast lesions on contrast-enhanced magnetic resonance images, Magnetic resonance in medicine 58 (3) (2007) 562–571.

[25] M. A. Nogueira, P. H. Abreu, P. Martins, P. Machado, H. Duarte, J. Santos, Image descriptors in radiology images: a systematic review, Artificial Intelligence Review 47 (4) (2017) 531–559. doi:10.1007/s10462-016-9492-8.

[26] L. Van Der Maaten, E. Postma, J. Van den Herik, Dimensionality reduction: a comparative, J Mach Learn Res 10 (2009) 66–71.

[27] N. Kambhatla, T. K. Leen, Dimension reduction by local principal component analysis, Dimension 9 (7).

[28] W. Bao, Z. Huang, C.-A. Yuan, D.-S. Huang, Pupylation sites prediction with ensemble classification model, International Journal of Data Mining and Bioinformatics 18 (2) (2017) 91–104.

[29] V. Bevilacqua, N. Pietroleonardo, V. Triggiani, A. Brunetti, A. M. Di Palma, M. Rossini, L. Gesualdo, An innovative neural network framework to classify blood vessels and tubules based on haralick features evaluated in histological images of kidney biopsy, Neurocomputing 228 (2017) 143–153.

[30] V. Bevilacqua, A. Brunetti, M. Triggiani, D. Magaletti, M. Telegrafo, M. Moschetta, An optimized feed-forward artificial neural network topology to support radiologists in breast lesions classification, in: T. Friedrich, F. Neumann, A. M. Sutton (Eds.), Genetic and Evolutionary Computation Conference, GECCO 2016, Denver, CO, USA, July 20-24, 2016, Companion Material Proceedings, ACM, 2016, pp. 1385–1392. doi:10.1145/2908961.2931733.

[31] S. Sharma, P. Khanna, Computer-aided diagnosis of malignant mammograms using zernike moments and svm, Journal of Digital Imaging 28 (1) (2015) 77–90. doi:10.1007/s10278-014-9719-7.

[32] W. Bao, Y. Chen, D. Wang, Prediction of protein structure classes with flexible neural tree, Bio-medical materials and engineering 24 (6) (2014) 3797–3806.

[33] J. Dheeba, N. A. Singh, S. T. Selvi, Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach, Journal of biomedical informatics 49 (2014) 45–52.

[34] L. P. Cordella, C. De Stefano, F. Fontanella, A. S. di Freca, A weighted majority vote strategy using bayesian networks, in: International Conference on Image Analysis and Processing, Springer, 2013, pp. 219–228.

[35] I. Saritas, Prediction of breast cancer using artificial neural networks, Journal of Medical Systems 36 (5) (2012) 2901–2907.

[36] C. De Stefano, F. Fontanella, G. Folino, A. S. Di Freca, A bayesian approach for combining ensembles of gp classifiers, in: International Workshop on Multiple Classifier Systems, Springer, 2011, pp. 26–35.

[37] H.-L. Chen, B. Yang, J. Liu, D.-Y. Liu, A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis, Expert Systems with Applications 38 (7) (2011) 9014–9022.

[38] J. Jiang, P. Trundle, J. Ren, Medical image analysis with artificial neural networks, Computerized Medical Imaging and Graphics 34 (8) (2010) 617–631.

[39] R. Janghel, A. Shukla, R. Tiwari, R. Kala, Breast cancer diagnosis using artificial neural network models, in: Information Sciences and Interaction Sciences (ICIS), 2010 3rd International Conference on, IEEE, 2010, pp. 89–94.

[40] D. Newell, K. Nie, J.-H. Chen, C.-C. Hsu, H. J. Yu, O. Nalcioglu, M.-Y. Su, Selection of diagnostic features on breast mri to differentiate between malignant and benign lesions using computer-aided diagnosis: differences in lesions presenting as mass and non-mass-like enhancement, European Radiology 20 (4) (2010) 771–781. doi:10.1007/s00330-009-1616-y.

[41] Y. Rejani, S. T. Selvi, Early detection of breast cancer using svm classifier technique, arXiv preprint arXiv:0912.2314.

[42] M. F. Akay, Support vector machines combined with feature selection for breast cancer diagnosis, Expert systems with applications 36 (2) (2009) 3240–3247.

[43] V. Bevilacqua, G. Mastronardi, F. Menolascina, P. Pannarale, A. Pedone, A novel multi-objective genetic algorithm approach to artificial neural network topology optimisation: The breast cancer classification problem, in: Proceedings of the International Joint Conference on Neural Networks, IJCNN 2006, part of the IEEE World Congress on Computational Intelligence, WCCI 2006, Vancouver, BC, Canada, 16-21 July 2006, IEEE, 2006, pp. 1958–1965. doi:10.1109/IJCNN.2006.246940.

[44] H. A. Abbass, An evolutionary artificial neural networks approach for breast cancer diagnosis, Artificial intelligence in Medicine 25 (3) (2002) 265–281.

[45] D.-S. Huang, Systematic theory of neural networks for pattern recognition, Publishing House of Electronic Industry of China, Beijing 201.

[46] D.-s. Huang, Radial basis probabilistic neural networks: Model and application, International Journal of Pattern Recognition and Artificial Intelligence 13 (07) (1999) 1083–1101.

[47] D.-S. Huang, J.-X. Du, A constructive hybrid structure optimization methodology for radial basis probabilistic neural networks, IEEE Transactions on Neural Networks 19 (12) (2008) 2099–2115.

[48] M. A. Nielsen, Neural networks and deep learning (2015).

[49] K. G. Sheela, S. N. Deepa, Review on methods to fix number of hidden neurons in neural networks, Mathematical Problems in Engineering 2013.

[50] D. A. Winkler, T. C. Le, Performance of deep and shallow neural networks, the universal approximation theorem, activity cliffs, and qsar, Molecular informatics 36 (1-2).

[51] J. Schmidhuber, Deep learning in neural networks: An overview, Neural networks 61 (2015) 85–117.

[52] P. Benardos, G.-C. Vosniakos, Optimizing feedforward artificial neural network architecture, Engineering Applications of Artificial Intelligence 20 (3) (2007) 365–382.

[53] M. Oquab, L. Bottou, I. Laptev, J. Sivic, Learning and transferring mid-level image representations using convolutional neural networks, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.

[54] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in neural information processing systems, 2012, pp. 1097–1105.

[55] A. Vedaldi, K. Lenc, Matconvnet: Convolutional neural networks for matlab, in: Proceedings of the 23rd ACM international conference on Multimedia, ACM, 2015, pp. 689–692.

[56] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell, Caffe: Convolutional architecture for fast feature embedding, in: Proceedings of the 22nd ACM international conference on Multimedia, ACM, 2014, pp. 675–678.

[57] M. Oquab, L. Bottou, I. Laptev, J. Sivic, Learning and transferring mid-level image representations using convolutional neural networks, in: Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on, IEEE, 2014, pp. 1717–1724.

[58] R. K. Samala, H.-P. Chan, L. Hadjiiski, K. Cha, M. A. Helvie, Deep-learning convolution neural network for computer-aided detection of microcalcifications in digital breast tomosynthesis, SPIE medical imaging. International Society for Optics and Photonics (2016) 97850Y–97850Y.

[59] M. Kallenberg, K. Petersen, M. Nielsen, A. Y. Ng, P. Diao, C. Igel, C. M. Vachon, K. Holland, R. R. Winkel, N. Karssemeijer, et al., Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring, IEEE transactions on medical imaging 35 (5) (2016) 1322–1331.

[60] J. Karnowski, Alexnet + svm, online; accessed INSERT DATE (2015). URL `https://jeremykarnowski.files.wordpress.com/2015/07/alexnet2.png`

[61] I. Goodfellow, Y. Bengio, A. Courville, Y. Bengio, Deep learning, Vol. 1, MIT press Cambridge, 2016.

[62] V. Bevilacqua, D. Altini, M. Bruni, M. Riezzo, A. Brunetti, C. Loconsole, A. Guerriero, G. F. Trotta, R. Fasano, M. D. Pirchio, C. Tartaglia, E. Ventrella, M. Telegrafo, M. Moschetta, A supervised breast lesion images classification from tomosynthesis technique, in: D. Huang, K. Jo, J. C. Figueroa-García (Eds.), Intelligent Computing Theories and Application - 13th International Conference, ICIC 2017, Liverpool, UK, August 7-10, 2017, Proceedings, Part II, Vol. 10362 of

Lecture Notes in Computer Science, Springer, 2017, pp. 483–489. doi:10.1007/978-3-319-63312-1_42.

[63] L. T. Niklason, B. T. Christian, L. E. Niklason, D. B. Kopans, D. E. Castleberry, B. Opsahl-Ong, C. E. Landberg, P. J. Slanetz, A. A. Giardino, R. Moore, et al., Digital tomosynthesis in breast imaging., Radiology 205 (2) (1997) 399–406.

[64] V. Bevilacqua, G. Mastronardi, M. Marinelli, A neural network approach to medical image segmentation and three-dimensional reconstruction, in: D. Huang, K. Li, G. W. Irwin (Eds.), Intelligent Computing, International Conference on Intelligent Computing, ICIC 2006, Kunming, China, August 16-19, 2006. Proceedings, Part I, Vol. 4113 of Lecture Notes in Computer Science, Springer, 2006, pp. 22–31. doi:10.1007/11816157_3.

[65] T. M. Lehmann, C. Gonner, K. Spitzer, Survey: Interpolation methods in medical image processing, IEEE transactions on medical imaging 18 (11) (1999) 1049–1075.

[66] J. S. Lim, Two-dimensional signal and image processing, Englewood Cliffs, NJ, Prentice Hall, 1990, 710 p.

[67] X.-F. Wang, D.-S. Huang, H. Xu, An efficient local chan–vese model for image segmentation, Pattern Recognition 43 (3) (2010) 603–618.

[68] X.-F. Wang, D.-S. Huang, A novel multi-layer level set method for image segmentation, J. Univers. Comput. Sci 14 (14) (2008) 2428–2452.

[69] G. Kom, A. Tiedeu, M. Kom, Automated detection of masses in mammograms by local adaptive thresholding, Computers in Biology and Medicine 37 (1) (2007) 37–48.

[70] L. Carnimeo, V. Bevilacqua, L. Cariello, G. Mastronardi, Retinal vessel extraction by a combined neural network-wavelet enhancement method, in: D. Huang, K. Jo, H. Lee, H. Kang, V. Bevilacqua (Eds.), Emerging Intelligent Computing Technology and Applications. With Aspects of Artificial Intelligence, 5th International Conference on Intelligent Computing, ICIC 2009, Ulsan, South Korea, September 16-19, 2009, Proceedings, Vol. 5755 of Lecture Notes in Computer Science, Springer, 2009, pp. 1106–1116. doi:10.1007/978-3-642-04020-7_118.

[71] V. Bevilacqua, M. Triggiani, M. Dimatteo, G. Bellantuono, A. Brunetti, L. Carnimeo, F. Marino, M. Telegrafo, M. Moschetta, Computer assisted detection of breast lesions in magnetic resonance images, in: D. Huang, V. Bevilacqua, P. Premaratne (Eds.), Intelligent Computing Theories and Application - 12th International Conference, ICIC 2016, Lanzhou, China, August 2-5, 2016, Proceedings, Part I, Vol. 9771 of Lecture Notes in Computer Science, Springer, 2016, pp. 306–316. doi:10.1007/978-3-319-42291-6_30.

[72] M. Kass, A. Witkin, D. Terzopoulos, Snakes: Active contour models, International journal of computer vision 1 (4) (1988) 321–331.

[73] V. Bevilacqua, G. Mastronardi, A. Piazzolla, An evolutionary method for model-based automatic segmentation of lower abdomen ct images for radiotherapy planning, Applications of Evolutionary Computation (2010) 320–327.

[74] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al., Imagenet large scale visual recognition challenge, International Journal of Computer Vision 115 (3) (2015) 211–252.

[75] D. A. Clausi, An analysis of co-occurrence texture statistics as a function of grey level quantization, Canadian Journal of remote sensing 28 (1) (2002) 45–62.

[76] L.-K. Soh, C. Tsatsoulis, Texture analysis of sar sea ice imagery using gray level co-occurrence matrices, IEEE Transactions on geoscience and remote sensing 37 (2) (1999) 780–795.

[77] R. M. Haralick, K. Shanmugam, et al., Textural features for image classification, IEEE Transactions on systems, man, and cybernetics (6) (1973) 610–621.

[78] I. T. Jolliffe, Principal component analysis and factor analysis, in: Principal component analysis, Springer, 1986, pp. 115–128.

[79] Z.-Q. Zhao, D.-S. Huang, W. Jia, Palmprint recognition with 2dpca+ pca based on modular neural networks, Neurocomputing 71 (1) (2007) 448–454.

[80] D.-S. Huang, J.-X. Mi, A new constrained independent component analysis method, IEEE Transactions on Neural Networks 18 (5) (2007) 1532–1535.

[81] K. L. Priddy, P. E. Keller, Artificial neural networks: an introduction, Vol. 68, SPIE press, 2005.

[82] L. P. Cordella, C. De Stefano, F. Fontanella, C. Marrocco, A. S. di Freca, Combining single class features for improving performance of a two stage classifier, in: Pattern Recognition (ICPR), 2010 20th International Conference on, IEEE, 2010, pp. 4352–4355.

[83] N. Japkowicz, S. Stephen, The class imbalance problem: A systematic study, Intelligent data analysis 6 (5) (2002) 429–449.

[84] C. De Stefano, F. Fontanella, C. Marrocco, A. S. di Freca, A hybrid evolutionary algorithm for bayesian networks learning: An application to classifier combination, in: European Conference on the Applications of Evolutionary Computation, Springer, 2010, pp. 221–230.

[85] W.-B. Zhao, D.-S. Huang, J.-Y. Du, L.-M. Wang, Genetic optimization of radial basis probabilistic neural networks, International Journal of Pattern Recognition and Artificial Intelligence 18 (08) (2004) 1473–1499.

[86] D.-S. Huang, W. Jiang, A general cpl-ads methodology for fixing dynamic parameters in dual environments, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 42 (5) (2012) 1489–1500.

[87] W. Jiang, D.-S. Huang, S. Li, Random walk-based solution to triple level stochastic point location problem, IEEE transactions on cybernetics 46 (6) (2016) 1438–1451.

[88] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer, 2015, pp. 234–241.

[89] P. Y. Simard, D. Steinkraus, J. C. Platt, et al., Best practices for convolutional neural networks applied to visual document analysis., in: ICDAR, Vol. 3, 2003, pp. 958–962.

[90] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1–9.

[91] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[92] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE 86 (11) (1998) 2278–2324.

[93] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556.

[94] K. Chatfield, K. Simonyan, A. Vedaldi, A. Zisserman, Return of the devil in the details: Delving deep into convolutional nets, arXiv preprint arXiv:1405.3531.

[95] M. D. Zeiler, R. Fergus, Visualizing and understanding convolutional networks, in: European conference on computer vision, Springer, 2014, pp. 818–833.

[96] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, Y. LeCun, Overfeat: Integrated recognition, localization and detection using convolutional networks, arXiv preprint arXiv:1312.6229.

[97] C. J. Burges, A tutorial on support vector machines for pattern recognition, Data mining and knowledge discovery 2 (2) (1998) 121–167.

[98] K. Beyer, J. Goldstein, R. Ramakrishnan, U. Shaft, When is nearest neighbor meaningful?, in: International conference on database theory, Springer, 1999, pp. 217–235.

[99] P. Domingos, M. Pazzani, On the optimality of the simple bayesian classifier under zero-one loss, Machine learning 29 (2) (1997) 103–130.

[100] L. Rokach, O. Maimon, Data mining with decision trees: theory and applications, World scientific, 2014.

[101] R. A. Fisher, The use of multiple measurements in taxonomic problems, Annals of human genetics 7 (2) (1936) 179–188.

[102] B. Li, C.-H. Zheng, D.-S. Huang, Locally linear discriminant embedding: An efficient method for face recognition, Pattern Recognition 41 (12) (2008) 3813–3821.