LOCATION-AWARE CONVOLUTIONAL NEURAL NETWORKS BASED BREAST TUMOR DETECTION

Huafeng Hu[1,6], Frans Coenen[2], Fei Ma[3,5*], Jeyarajan Thiyagalingam[4*], Jionglong Su[3,5,6*]

[1]Department of Electrical and Electronic Engineering, University of Liverpool based in Xi'an Jiaotong-Liverpool University, Suzhou, China
[2]Department of Computer Science, University of Liverpool, Liverpool, UK
[3]Department of Mathematical Sciences, Xi'an Jiaotong-Liverpool University, Suzhou, China
[4] Scientific Computing Department, Science and Technologies Facilities Council, Harwell Campus, Oxford, UK
[5]Research Center for Precision Medicine, Xi'an Jiaotong-Liverpool University, Suzhou, China
[6]Neusoft Corporation, Shenyang, China
**Authors to whom correspondence should be addressed*
**fei.ma@xjtlu.edu.cn, t.jeyan@stfc.ac.uk, jionglong.su@xjtlu.edu.cn*

Keywords: BREAST TUMOR DETECTION, CONVOLUTIONAL NEURAL NETWORKS.

Abstract

Breast cancer is one of the most common types of cancer affecting the lives of millions. Early detection and localization of the breast cancer tissues are vital for prevention and cure. Recently, there have been a number of developments on this front, particularly in the direction of automated image analysis. Although they are instrumental in expediting the process, such approaches lack the localization information and hence still demand substantial involvement of clinicians to deliver conclusive results. In this paper, we propose a novel approach for detecting and localizing cancer tissues from mammograms. In particular, we rely on Convolutional Neural Networks for exploiting the spatial relationship of the cancer tissues for detection and localization. Our evaluations on real datasets show that the proposed method is able to classify normal and tumor tissues with the classification accuracy of 90.8%. Furthermore, our approach achieves the sensitivity of 86.1% in detection with 1.4 false positives per image on the localization. In comparison to the state-of-the-art approaches, our method offers an additional 1.1% sensitivity improvement, along with reduced two false positives per image.

1 Introduction

Breast cancer, caused by malignant tumors in human epithelial tissue, is a kind of cancer with high incidence rate. According to the statistics from the International Agency for Research on Cancer, 22.9% of invasive cancers are breast cancer and 13.7% of cancers for women are breast cancer [1]. As such, early detection and localization of cancer tissues from mammograms are two vital aspects towards prevention and cure of breast cancer [2]. Routine mammography is a common approach for breast cancer diagnosis adopted by clinicians. However, analysis of mammograms requires significant involvement from clinicians and the assessments can be subjective.

Over the recent years, a number of image-processing-based approaches have been adopted to expedite this process and to introduce the notion of unbiased analysis of data. Given the wide-spread adoptation of convolutional neural networks (CNNs) [3] in image-based data analysis, it is natural to exploit CNNs in mammography analysis as well.

However, from a deep learning perspective, the detection and localization of breast cancer tissues is a challenging problem that cannot be simply treated as a general classification problem. This is because the classification or the diagnosis of a tumor is only determined by a few small regions called the regions of interest (ROIs) [4]. Hence, for the diagnoses with deep learning, one must detect the ROIs first. Therefore, it is crucial for obtaining regions containing tumors for further diagnosis. However, this aspect is heavily overlooked in the literature. As such, despite the fact that existing CNN-based approaches offering a good classification performance, they lack the spatial localization information. Thus, the methods are considered to be semi-automatic and require manual ROI detection in practices, which demands substantial involvement from clinicians.

In this paper, we propose a novel technique for detecting and localizing the tumor tissues. In other words, given a mammogram, we not only classify the lesion types (tumors or none), we also provide the location of the tumor(s), if one found. The paper makes the following key contributions:

- Instead of building a CNN from scratch, we show how the existing pre-trained networks (such as VGG16) can be adopted for this purpose;
- We construct an enhanced VGG16 that is more suitable for analysing mammograms; and
- We propose an approach for detecting and localizing the tumor tissues in mammograms.

The rest of this paper is organized as follows. In Section 2, the different methods for breast cancer detection and diagnosis from the literature are reviewed. In Section 3, we

provide a detailed account of our methods, tools and functional blocks. In Section 4, we evaluate the proposed approach and present our results. Finally, we conclude the paper in Section 5.

2 Related Work

There are a number of methods that have been used to identify the relevant tissues. Most of these rely on techniques for image segmentation [5], particularly in the medical context. On this note, methods based on thresholding using global intensity information, for separating different artefacts in an image are the simplest and common precursor techniques for segmenting the ROI [6]. Wei et al. used watershed transformation with pectoral muscle segmentation on breast segmentation [7]. However, the pectoral muscle segmentation cannot be automatically removed because it increases false positive for further diagnosis. Raba et al. proposed a combined method of the Histogram approach and pectoral muscle suppression [8]. However the results showed that the reliability of the method was not acceptable in practice [8].

Bellotti et al. proposed a system for detection of tumors by using an edge-based selection algorithm with co-occurrence matrices [9] in conjunction with an artificial neural network to obtain a sensitivity of 80% in detection with 4.23 false positives per image.

Chan et al. developed a system for detecting two-dimensional mammograms through a process called tomosynthesis [10]. Their results indicated the sensitivity of the system was 85% with 3.73 false positives per image.

Deep learning methods, based on CNNs, have been used for image segmentation with more than 80% success rate [11]. For example, Girshick et al. successfully applied region-based CNN in objective detection [12]. Moreover, different modified CNNs have been used to achieve more accurate segmentation, including Ball's adaptive level set segmentation [13] and Jaffar's deep learning [14]. These approaches provide the overall outcome tumors without any spatial localization information.

In our work, we use CNNs but we enhance the approach of using CNNs specifically for deriving the spatial localization information along with additional performance in accuracy.

3 Methodology

Fig. 1 shows the overall processing pipeline of our approach. We first pre-process and enhancement the mammograms which become the input to the CNN classifier. This is then followed by a customised CNN model, which we designed by enhancing the existing pre-trained models. In particular, we use transfer learning to detect the region of tumors. The aim of using CNNs is to obtain the probability for each patch having tumor tissues, and a pre-determined threshold for selecting the region of tumors for each image. The base CNN used in our model are five different networks, i.e. AlexNet, GoogLeNet, Res50, VGG16, and VGG19 [15-18]. The final layer can classify the tissues into two classes: normal and tumor-linked. The model outputs a region for each case (tumor-linked tissue region), and an associated probability of the region containing tumor tissues. The region which has the highest probability of containing the tumor is automatically prominently highlighted.

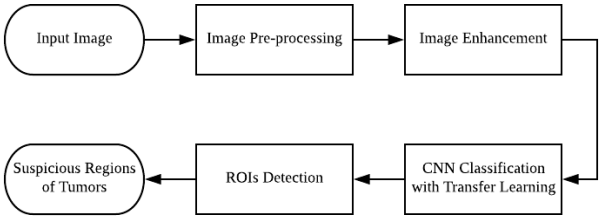


Fig. 1 The pipeline of our proposed method

3.1 Pre-process and Enhancement of the Input Images

One of the prevalent problems in mammograms is the low-contrast and high variation of contrast values in images. This step is to obtain refined and high-quality images through a series of transformations. Although an array of techniques, including histogram equalization, can be applied, the Contrast Limited Adaptive histogram equalization (CLAHE) method is very appropriate for improving the contrast variation in images, particularly in mammograms [19]. It differs from ordinary histogram equalization such that the adaptive method computes a number of histograms, with each histogram corresponding to different parts of the image. and uses them to redistribute the lightness values of the image. The exact sequences of steps embedded inside the pre-process stage are shown in Fig. 2

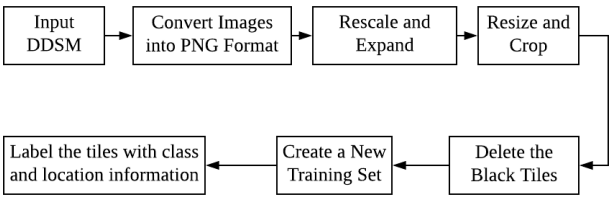


Fig. 2 The pre-process pipeline

We introduce each step in details as follows:

- We convert mammograms from public dataset format into a usable PNG format, and rescale the images to 8-bit colored images with three channels. The classification or the diagnosis of a tumor on a mammogram is only determined by the ROI. For example, a mammogram image typically has the resolution of 5000x4000 while the average tumor sizes identified on the ROIs can be as small as 20%-30% of the overall ROI. These small regions are hard to observe even in manual diagnosis. For this reason, we resize the images into the resolution of 1120x896 retaining the same aspect ratio. We then partition each image into 20

small patches, each with the size of 224x224. In the test set, we crop each image by a 224x224 sized window with the stride of 32-pixels yielding 638 patches.

- We remove all tiles where 90% of the pixels are black (undersaturated) to avoid misleading the classifier about the differences between non-tumor class and tumor class.
- We create a training set by sampling the patches around the mass.
- We augment the training set by including geometrically transformed tiles, such as rotations at various degrees 90 °, 180 °, 270 ° and horizontal / vertical flipping.
- Finally, using the ground truth, we label each as non-tumor or tumor, and with the localization information wherever appropriate.

(a) *(b)* *(c)* *(d)*

Fig. 3 (a) Original patch, (b) Black patch wich containing more than 90% of pixels as under-saturated pixels, (c) Patch by sampling, (d) Patch by sampling and enhancement

Fig. 3 shows an examples of different types of tiles from our dataset. 3(a) illustrates the original patch in the tumor class, which is difficult to extract the feature of the mass. 3(b) shows the black patch, which we delete during our pre-processing step. 3(c) shows the patch that is selected by the sampling process. 3(d) is derived from 3(c) through various enhancement techniques. It can be seen that local contrast and boundaries of the tumors are more distinct in 3(d). We then label the mammograms partition them into small into patches, which become the input to the CNN classifier.

3.2 Transfer Learning

Transfer learning is a technique in machine learning that pretrains a model crossing the original domain to decrease the need for abundant labeled data for training [20]. In our approach, the initial CNN model for training is a pretrained model using the ImageNet dataset with 1.2 million images. It is worth noting that ImageNet is the the largest and commonly used dataset for image classification and detection [21]. The pre-training step is useful for setting up our CNN model and to avoid the issues surrounding the lack of training data. Meanwhile, pretraining by transfer learning can also increase the learning speed as well as improving the performance of the architecture [20].

3.3 Convolutional Neural Networks (CNNs)

For detecting the regions of tumors, a feed-forward convolutional neural network is applied for the classification of mammogram patches. In this paper, we propose to use CNN. Primarily, CNN does not require explicit feature engineering. Secondly, it has much better performance on image classification task than other machine learning algorithms [23], such as multi-layer perceptron [24] and plain deep neural networks [25]. CNN is an advanced deep learning technique inspired by the unique structure of cerebral cortex neurons of cats [3]. Fig. 4 shows a simplified structure of CNN consisting two major parts: feature extraction and classification. The feature extraction part of a CNN consists of convolutional layers and subsampling layers. The classification part of CNN consists of fully connected layers.

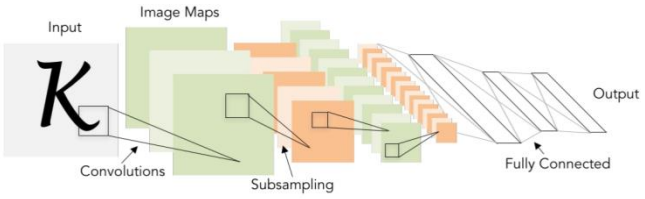


Fig. 4 A simplified structure of CNN [22]

The first part of the CNN consists of multiple layers collecting underlying features in input images. These layers extract a feature representation of the input image with the loss functions. The local or global subsampling layers (also known as pooling layers) combine the outputs for each neuron clusters [26]. The second part of the CNN, fully connected layers is interfaced to the first part [24]. Due to the shared weights of the CNN, the same filter is used for each pixel in the layer, which leads to lower memory footprint and better performance [27]. The detection is carried out using a modified CNN model with transfer learning for identification of ROIs. The initial CNN model for training is pretrained by the ImageNet dataset. In our experiment, we use a number of models, namely AlexNet, GoogLeNet, Res50, VGG16, and VGG19, as our base models. We use CNN to obtain the probability for each patch to have tumor segments and select the final regions of tumor for each image.

4 Experiments and Evaluation

In this section, we describe the dataset we use, the experimental settings and the evaluation of our results. We also give the experiment results of our model with different CNN architectures and compared it with other methods.

4.1 The Dataset

In our experiment we use a public dataset called digital database for screening mammography (DDSM) [28]. This public dataset consists of 2620 digitized film-screen screening mammograms with pixel-level ground truth annotation for tumors [28]. Each mammogram includes two standard projections, the craniocaudal (CC) view and the mediolateral oblique (MLO) view, along with localization information. The localization information stored in DDSM was supplied by specialists. We use the mammogram images from Lumisys scanner which has the highest resolution in DDSM as our whole dataset. The whole dataset has 666 images in the benign class and 657 images in the malignant class. In our experiment, we consider both benign tumor and malignant tumor as tumor class. The images are in LJPEG format and they are 16-bit grayscale images which only have one channel. In each experiment, we randomly split the

whole dataset into a training set, a validation set and a test set according to 80%, 10%, 10% ratio, respectively.

4.2 Evaluation Metric

We use classification accuracy (CA), Area under the ROC curve (AUC score), sensitivity in detection, and false positive per image (FPPI) as our evaluation metrics. These metrics are defined as follows.

4.2.1 Classification Accuracy (CA): is the ratio of the number of correct predictions to the total number of predictions using our CNN architecture. This is used to measure the classification performance of our model.

$$\mathrm{CA} = \frac{TN + TP}{TN + TP + FN + FP}$$

Where *TN* is the number of true negatives, *TP* is the number of true positives, *FN* is the number of false negatives, and *FP* is the number of false positives.

4.2.2 Area Under the Curve (AUC)-Receiver Operating Characteristics (ROC) Curve: is a performance measurement for classification problems at various threshold settings [24]. It consists of two components, ROC is a probability curve and AUC represents the degree or measure of separability. The separability shows the performance of the model for the ability of clearly distinguishing different classes. The ROC curve is obtained by plotting with true positive rate (TPR) against false positive rate (FPR). TPR and FPR in AUC-ROC Curve are defined as follows.

TPR is the number of correct positive predictions divided by the total number of actual positive events [29].

$$\mathrm{TPR} = \frac{TP}{TP + FN}$$

FPR is the number of negative events wrongly predicted as positive divided by the total number of actual negative events [29].

$$\mathrm{FPR} = \frac{FP}{FP + TN}$$

4.2.3 Sensitivity in Detection: the number of the correct selected region of tumors by our pipeline divided by the total number of tumors. It is used to measure the detection performance of our method.

$$\text{Sensitivity in Detection} = \frac{TP \text{ in Detection}}{(TP + FN) \text{ in Detection}}$$

4.2.4 False Positives Per Image (FPPI): the number of the wrong selected region of tumors by our pipeline divided by the total number of images [30]. It is used to measure the detection performance of our method.

$$\mathrm{FFPI} = \frac{\text{Wrong Selected Region}}{\text{Total Number of Images}}$$

4.3 Settings and Results

The whole training set of CNN architectures contains an average of 11310 patches for the non-tumor class and 18103 patches for the tumor class. To overcome the over-fitting problem, we add a dropout layer on the classification layer. All experiments in this study are conducted on a system with Intel i7-7700HQ quad-core processor, clocked at 2.80GHz, with 16GB RAM, and an NVIDIA GeForce GTX 1060 GPU with 6GB RAM.

To determine the optimal base model, we compared the performance of five different CNN models we outlined before: AlexNet, GoogLeNet, Res50, VGG16, and VGG19. We set the hyperparameters for each CNN model as given in Table 1.

Table 1 Hyperparameter settings

Optimizer	Learning Rate	Momentum	Batch Size	Dropout Rate
SGDM	0.00001	0.9	64	0.5

All of these models have been pretrained by the ImageNet dataset [14]. The final results on the validation set are shown in Table 2.

Table 2 The classification results of different models on the validation set

Model	CA on the Validation Set
AlexNet	86.5%
GoogLeNet	88.4%
Res50	90.1%
VGG16	90.8%
VGG19	89.7%

VGG16 achieves the highest classification accuracy of 90.8% on the validation set. The final results on the test set are shown in Table 3.

Table 3 The results of different models on the test set

Model	CA on the Test Set	AUC Score
AlexNet	77.5%	0.73
GoogLeNet	82.3%	0.77
Res50	83.7%	0.81
VGG16	84.6%	0.83
VGG19	83.9%	0.82

VGG16 also achieves the highest classification accuracy of 85.6% with an AUC score of 0.83 on the test set. The results show that, in our experiment, the VGG16 is the most appropriate base model. Therefore, we use the VGG16 as our final base model. With this, we test the classification

performance of the VGG16 model with different classification layers. The relevant hyperparameters are given in Table 1. In this paper, we use one-vs-one SVM with optimal linear kernel function. The final results are summarized in Table 4.

Table 4 The classification results for the VGG model with different classification layers on the test set

Type of Classification Layers	CA on the Test Set
VGG16-Softmax	84.5%
VGG16-Sigmoid	85.6%
VGG16-SVM	85.0%

It shows VGG16 with sigmoid function is the optimal classifier for our data. The final adjusted hyperparameters of our method are shown in Table 5. Furthermore, in the final method, binary cross entropy is used as the loss function. We trained the network in batches with the epoch size of 60.

Table 5 Adjusted hyperparameters of VGG16-Sigmoid

Optimizer	Learning Rate	Momentum	Batch Size	Dropout Rate
SGDM	0.0001	0.9	32	0.5

The final classification accuracy is 85.6%, and the ROC of our final model is shown as Fig. 5. It shows that our method has a powerful separability of distinguishing between the non-tumor class and the tumor class.

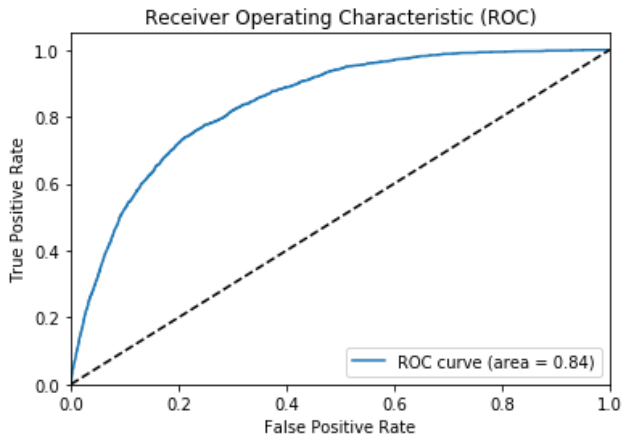


Fig. 5 The AUC-ROC curve of our method

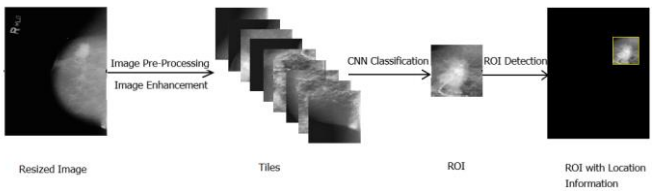


Fig. 6 An example of our proposed method

Fig. 6 shows a completed pipeline for one input mammogram with one output. However, the aim of our method is to select the tumor-linked patches as much as possible. Therefore, we change the criteria for selection. The method selects the patches with the highest probability of containing the tumor for each mammogram. In the test set, we crop each image into 638 patches. After that, the ones with highest probabilities of containing the tumors are selected as the suspicious patches for the regions of tumors. In order to test the model's ability to detect and accurately localize tumors with different number of regions, we evaluate the predictions on the test set using the sensitivity and FPPI. A detection result is considered correct if more than 50% area of the ground truth tumor fall inside the output window. The results with different number of regions are shown as Table 6.

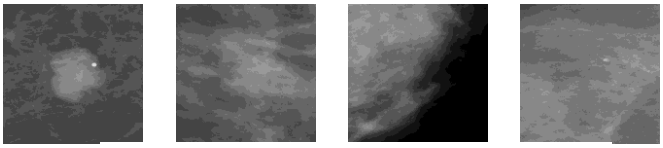
Table 6 The detection performance of different number of regions with the highest probabilities and two state-of-the-art methods

Method	Sensitivity	FFPI
One Most Appropriate Region Method	63.2%	0.4
Two Most Appropriate Regions Method	76.1%	0.9
Three Most Appropriate Regions Method	86.1%	1.4
Four Most Appropriate Regions Method	89.3%	2.1
Five Most Appropriate Regions Method	90.7%	3.0
Artificial Neural Network [9]	80%	4.0
Tomosynthesis Method [10]	85%	3.73

In our experiments, when we increase the number of selected regions, the sensitivity has the tendency to increase initially with the increasing number of tumors. However, the rate of the increase sensitivity is reduced with more than three regions. We acquire more accurate detection results as we select more regions if we use the sensitivity as the only evaluation criteria. However, when the number of regions used reaches three, the increase rate of sensitivity becomes negligible whereas the FPPI, which is another important evaluation, increases rapidly. The negative impact of our false positive rate leverages the positive impact of sensitivity improvement when three regions are selected. Therefore, when taking both evaluations into consideration, we shall select three regions as the suspicious ROIs to support further diagnosis.

In order to demonstrate the detection performance of our method, Fig. 7 shows a collection of correctly detected, false positive and missed tumors, that are cropped from the original mammograms by the localization information. The result shows that our proposed method is able to classify non-tumor and tumor tissue images with the classification accuracy of 90.8% on the validation set. The sensitivity of tumors in detection of our method is 86.1% with an acceptable FPPI of 1.4. Compared to the results in artificial neural network and tomosynthesis method [9, 10], our proposed approach outperforms state-of-the-art methods and offers at least additional 1.1% sensitivity improvement, along with reduced two false positives per image. Their results come directly from their papers, which were obtained in their

studies with their own testing datasets different to the ones used in this study. A disadvantage of our research is the slightly higher value of FPPI. Since information of normal lesions in the training set is not fully learned, some normal lesions could be wrongly detected as tumors.



(a) *(b)* *(c)* *(d)*

Fig. 7 (a) and (b) Correctly detected tumor, (c) False positive detection, (d) Missed tumor

5 Conclusions

In this paper, we proposed a novel method for breast tumor detection. The aim of our proposed method is twofold: to detect the cancer cells and to identify their location. More specifically, we provided a framework that not only detects the tumor cells, but also to provide a confidence on the probability of detection, and the exact location of these cells in mammograms.

The evaluations of our approach on the DDSM dataset show that our approach is able to detect 86.1% of the region of tumors with acceptable FPPI. It also achieves a good performance in terms of distinguishing between non-tumor and tumor patches.

Although our approach provides a reasonable performance, the overall performance can be improved through a number of approaches. In particular, we will be investigating using gradient boosting and decision trees to reduce the value of FFPI. Another avenue we will be investigating is the possibility of classifying tumors into three different classes, namely normal lesion, benign mass, and malignant mass.

6 Acknowledgements

This work is supported by Neusoft Corporation, item number SKLSAOP1702.

7 References

[1] Boyle, P., Levin, B., et al.: 'World cancer report 2008' (IARC Press, International Agency for Research on Cancer, 2008).

[2] Bekker, A.J., Shalhon, M., Green, H., et al.: 'Multi-view probabilistic classification of breast microcalcifications', IEEE Transactions on medical imaging, 2016, 35, (2), pp. 645-653.

[3] Matsugu, M., Mori, K., Mitari, Y., et al.: 'Subject independent facial expression recognition with robust face detection using a convolutional neural network', Neural Networks, 2003, 16, (5), pp. 555-559.

[4] Oliver, A., Freixenet, J., Marti, J., et al.: 'A review of automatic mass detection and segmentation in mammographic images', Medical image analysis, 2010, 14, (2), pp. 87-110.

[5] Lu, X., Dong, Y., Wang, K.: 'Automatic Mass Segmentation Method in mammograms based on improved VFC Snake model', in Hamid, R.: 'Trends in Image Processing, Computer Vision, and Pattern Recognition' (CSREA Press, 2014, 1st edn.), pp. 201-217.

[6] Zhou, S., Wang, J., Zhang, S., et al.: 'Active contour model based on local and global intensity information for medical image segmentation', Neurocomputing, 2016, 186, (1), pp. 107-118.

[7] Wei, K., Wang, G., Ding, Hui.: 'Segmentation of the breast region in mammograms using watershed transformation', 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Shanghai, China, September 2005, pp. 6500-6503.

[8] Raba, D., Oliver, A., Peracaula, M., et al.: 'Breast segmentation with pectoral muscle suppression on digital mammograms', Iberian Conference on Pattern Recognition and Image Analysis, Estoril, Portugal, June 2005, pp. 471-478.

[9] Bellotti, R., De Carlo, S., Tangaro, G., et al.: 'A completely automated CAD system for mass detection in a large mammographic database', Medical physics, 2006, 33, (8), pp. 3066-3075.

[10] Chan, H.-P., Wei, J., Zhang, Y., et al.: 'Computer-aided detection of masses in digital tomosynthesis mammography: Comparison of three approaches', Medical physics, 2008, 35, (9), pp. 4087-4095.

[11] LeCun, Y., Bengio, Y., Hinton, G.: 'Deep learning', Nature, 2015, 521, (7553), pp. 436-444.

[12] Girshick, R., Donahue, J., Darrell, T., et al.: 'Region-based convolutional networks for accurate object detection and segmentation', IEEE transactions on pattern analysis and machine intelligence, 2016, 38, (1), pp. 142-158.

[13] Ball, J.E., Bruce, L.M.: 'Digital mammographic computer aided diagnosis (CAD) using adaptive level set segmentation'. 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Lyon, France, August 2007, pp. 4973-4978.

[14] Jaffar, M.A.: 'Deep Learning based Computer Aided Diagnosis System for Breast Mammograms', International Journal of Advanced Computer Science and Applications, 2017, 8, (7), pp. 286-290.

[15] Krizhevsky, A., Sutskever, I., Hinton, G. E.: 'Imagenet classification with deep convolutional neural networks',

Advances in neural information processing systems, Nevada, United States, December 2012, pp. 1097-1105.

[16] Szegedy, C., Liu, W., Jia, Y.: 'Going deeper with convolutions', IEEE conference on computer vision and pattern recognition, Boston, United States, June 2015, pp. 1-9.

[17] He, K., Zhang, X., Ren, S., et al.: 'Deep Residual Learning for Image Recognition', IEEE conference on computer vision and pattern recognition, Las Vegas, United States, June 2016, pp. 770-778.

[18] Simonyan, K., Zisserman, A.: 'Very Deep Convolutional Networks for Large-Scale Image Recognition', arXiv preprint arXiv:1409.1556, 2014.

[19] Reza, A.M.: 'Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement', VLSI Signal Processing Systems for signal, image and video technology, 2004, 38, (1), pp. 35-44.

[20] Yosinski, J., Clune, J., Bengio, Y., et al.: 'How transferable are features in deep neural networks?'. Advances in Neural Information Processing Systems, Montreal, Canada, December 2014, pp. 3320-3328.

[21] Pan, S.J., Yang, Q.: 'A survey on transfer learning', IEEE Transactions on knowledge and data engineering, 2010, 22, (10), pp. 1345-1359.

[22] LeCun, Y., Bottou, L., Bengio, Y., et al.: 'Gradient-based learning applied to document recognition', Proceedings of the IEEE, 1998, 86, (11), pp. 2278-2324.

[23] Sharif, R.A., and Azizpour, H., Sullivan, J., et al.: 'CNN features off-the-shelf: an astounding baseline for recognition', Proceedings of the IEEE conference, Columbus, United States, June 2014, pp. 806-813.

[24] Driss, S.B., Soua, M., and Kachouri, R., et al.: 'A comparison study between MLP and Convolutional Neural Network models for character recognition', Real-Time Image and Video Processing, California, United States, April 2017, pp. 306-312.

[25] Abdel-Hamid, O., Deng, Li., Yu, D.: 'Exploring convolutional neural network structures and optimization techniques for speech recognition', 14th Annual Conference of the International Speech Communication Association, Lyon, France, August 2013, pp. 1173-1175.

[26] Ciresan, D.C., Meier, U., Masci, J., et al.: 'Flexible, high performance convolutional neural networks for image classification', IJCAI Proceedings-International Joint Conference on Artificial Intelligence, Barcelona, Spain, July 2011, pp. 1237-1241.

[27] Amari, S.: 'The handbook of brain theory and neural networks', (MIT press, 2003).

[28] 'DDSM: Digital Database for Screening Mammography', http://marathon.csee.usf.edu/Mammography/Database.html, accessed 6 August 2017.

[29] Metz, C.: 'Basic principles of ROC analysis', Seminars in Nuclear Medicine, 1978, 8, (4), pp. 283-298.

[30] Dollár, P., Wojek, C., Schiele, B., et al. 'Pedestrian detection: A benchmark', IEEE Conference on Computer Vision and Pattern Recognition, Miami, United States, June 2009, pp. 304-311.