

DIAGNOSTIC SUPPORT SYSTEM FOR THE DETECTION OF MICROCALCIFICATIONS IN DIGITAL MAMMOGRAMS

Weverson Machado de Oliveira (UCAM)

weversonmachado@yahoo.com.br

Italo O. Matias (UCAM)

italo@ucam-campos.br

Thiago Ribeiro de Azeredo (UCAM)

thiagoribeiro@gmail.com



Currently the breast cancer is one of the diseases that kill more women aged over forty years in Brazil. Although breast cancer is a cancer with early diagnosis and appropriate treatment, the mortality rate is still very high. This is due in large part by the late diagnosis of cancer where it is already in a more advanced stage and also by the misuse of the information provided by mammography examinations by doctors. In the early stages, even without the presence of lumps, tissue changes cause the accumulation of calcium and can be identified by mammography as a carbon pool and these can be benign or even indicate a cancer in its very early stages. This paper presents a software, which was built using techniques of Artificial Intelligence (AI) and Artificial Neural Networks (Multilayer Perceptron) in conjunction with the pattern detection algorithm AdaBoost, used by OpenCV library, for the detection of microcalcifications level BIRADS 2 images in digital mammography. This software is intended as aid to the radiologist in the detection of microcalcifications from digital images. During the tests come up with better results than the 94.5% accuracy. Once it is done to detect the radiologist to classify them.

Palavras-chaves: Microcalcifications, AdaBoost, OpenCV, Neural Networks

1.3

2 Introduction

For decades various research related to breast cancer, which has a very high mortality rates (Mavroforakis, et al., 2005), see references with the goal to save or prolong human life. With every effort performed by science, supported by technology, new tools are being developed as an aid in the prevention and detection of breast cancer.

Faced with this situation of high mortality caused by breast cancer this work aims to develop software to serve as an aid to the radiologist in the detection of microcalcifications. Using the combination of object detection algorithm AdaBoost and a neural network, it was possible the detection of microcalcifications (Level BIRADS 2) so that the system has obtained a hit ratio greater than 94.5%. It is now set to the radiologist detections made by the system.

3 Microcalcifications

Non palpable lesions, detected only by mammograms and ultrasound, microcalcifications may constitute a first sign of a mammary carcinoma. The diagnosis of lesions of this kind can produce very positive effects for the treatment and prognosis of same.

Microcalcifications in the breast were applets described in 1951 by Leborgne and malignant lesions can occur on or benign.

A microcalcification consists of calcium hydroxyapatite or calcium triphosphate resulting from increased activity in cell secretory ductal-lobe complex, rather than the result of necrotic cellular debris mineralization.

In Figure 1 we can check for the presence of microcalcifications level BI-RADS 2 already processed by the software for this project.

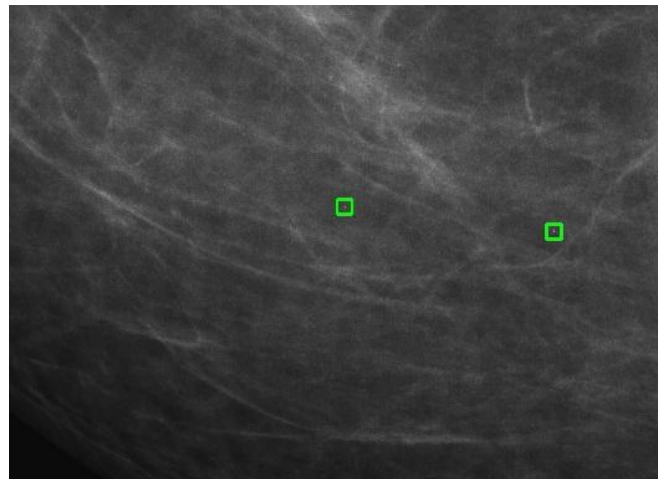


Figure 1: Microcalcifications level 2 BI-RADS. Source: own.

Seeking to estimate the level of malignancy of breast microcalcifications, the American College of Radiologists created the classifier BI-RADS (Breast Imaging Reporting And Data System), which categorized the lesions at several levels. In table 1 we display the classifier BI-RADS of clusters of microcalcifications.

Level BI-RADS	Description
1	Negative (nothing found)
2	Benign
3	Probably benign (2% probability to be cancer)
4	This can be A, B or C. The breakdown is less likely A, B and C the intermediate likely reaching 20% probability of being a cancer.
5	High likelihood of being cancer.
6	Used only when it was already confirmed the diagnosis.

Table 1: Classifier BI-RADS - Source: American College of Radiology - <http://www.acr.org/>

For evaluation of mammography and prescription of award, the radiologist says: size, number, shape, density and distribution of microcalcifications [Ministério da Saúde, INCA – Instituto Nacional do Câncer, 2007]. Larger particles may represent kindness, therefore, the detection of microcalcifications is the subject of our work (we will cover only the level 2 BI-RADS in the system, but the principle of detection may be used for any level).

4 Methodology

The Computational Intelligence (CI) aims to allow the machine acquires the ability to perform tasks that are performed by humans. Most often, these tasks require an ability to reason, such as planning, strategy, optimization, image recognition and sounds, pattern recognition, memorization, among others. In general, the algorithms implement a learning process, where they acquire knowledge for decision making.

The objective of this study was to generate a software that was capable of detecting microcalcifications in digital mammography images as a way to aid the radiologist in detecting and for this the system followed a few steps:

At first were obtained several images of digital mammography equipment, obtained these images after they were passed to a specialist in radiology where they were medical report to an understanding of how the training system would be, or what system would look in the picture. Soon after came to making the software that would detect these objects, and for their training were used positive images, where the objects found and negative images, or images that did not contain the object of research.

During this step we used the OpenCV library where she adaBoost implements the algorithm for detecting objects. After this initial training phase the system was detecting microcalcifications, but was detecting false positives.

To try to eliminate these false positives was implemented as an auxiliary multilayer artificial neural network where all the microcalcifications detected by the system go through this second stage in order to try to eliminate these false positives.

After the integration of object detection algorithm (microcalcifications) with neural network and the system was able to detect microcalcifications with an accuracy above 94.5%.

5 Artificial Neural Networks

Artificial Neural Networks are computational models based on the nervous system of living beings and try to emulate his smart way of processing information. According to Haykin

[Haykin, 2001] A neural network is a massively parallel distributed processor made up of simple processing units, which have a natural propensity for storing experiential knowledge and making it available for use.

Also according to Ludwig Jr. and Montgomery [Montgomery Jr. and Ludwig, 2007], the greatest virtue of neural networks is the ability to learn from input data with or without a teacher.

In the area of computer artificial neural networks are probably the oldest technique of Artificial Intelligence in use. This technique can be defined as the interconnection of processing units called artificial neurons that may contain large numbers of interconnections called artificial synapses.

5.1 The Biological Neuron

The human nervous system consists of a network of neurons, the basic unit of the brain, responsible for the phenomena known as thought, emotion and cognition, beyond the implementation of sensor motor and Autonomic functions-[WASSERMAN, 1989].

According to Silva and Spatti [Spatti and Silva, 2010] The information processing in the human brain is governed by biological elements operating in parallel, aiming to produce appropriate actions for each of its features, such as thinking and memorizing .

The elementary cell of our cerebral nervous system is the neuron (Figure 2). So if you know, our brain has about 10 billion of these neurons and each of these neurons is capable of performing up to 10,000 synapses with other neurons.

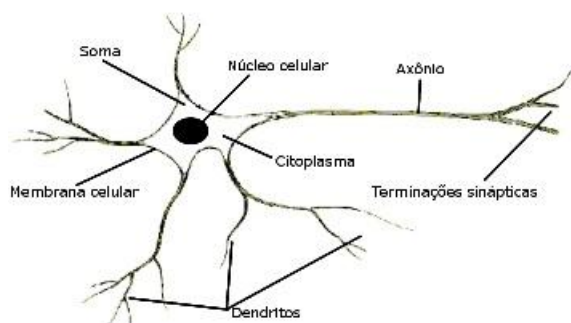


Figure2: Biological Neuron. Source: own.

The biological neuron is composed of three main parts that are the post-synaptic cell, the body and the axon.

The cell body has the function to process the signals from the post-synaptic potential generating activation can be raised along its Axon. The post-synaptic continuously capture transmitted by other stimuli, neurons have the function of Axon, which consists of a single tubular extension and that can reach up to a few meters, is pass the stimulus to other cells. Through the synapses neurons join together functionally, forming neural networks [BEALEI, 1990].

5.2 The Artificial Neuron

The units of an artificial neural network processors are the artificial neurons that are simplified neuron structures as well.

Such structures were inspired from the analysis of the generation and propagation of electrical impulses through the cell membrane of neurons [Hodgkin & Huxley, 1952]. Artificial neural networks of neurons perform simple functions such as receiving one or more input signals and give a single output signal, taking into consideration their activation function, which can serve as outbound network or as input to other neurons. In Figure 3 there is the representation of an artificial neuron.

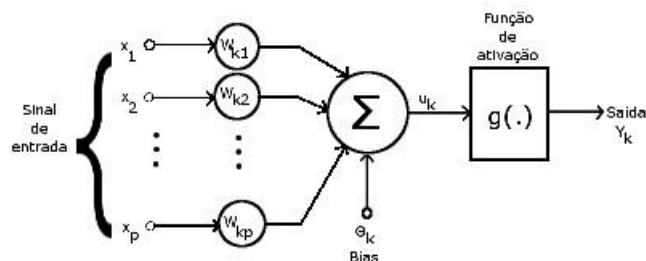


Figure 3: Representation of an artificial neuron. Source: own

When the input, x , are entered into the neuron, they are automatically multiplied by their respective weights synaptic contacts, w , generating this way the weighted entries.

$$W_{k1} x_1, W_{k2} x_2, \dots, W_{kp} x_p$$

These entries multiplied by their respective weights synaptic contacts are summed yielding as a result the output of sum (Σ) u_k .

$$u_k = \sum_{i=0}^n W_i x_i$$

The output of the sum (u_k) is then presented the transfer function. The transfer function is intended to limit the output of the neuron within a reasonable range of values.

$$Y_k = g(u_k)$$

5.3 Classification of Neural Networks

There are many types of artificial neural networks and various ways to classify them. One of the very important classification modes of neural networks is about learning the same way that can be supervised and unsupervised.

In supervised learning are presented the network input set and their corresponding output. During this process the input values are compared with the values of the desired output and then weights synaptic contacts and bias are adjusted to minimize the difference between the desired output and the output produced by the network.

In unsupervised learning or self supervised the network receives information of desired outputs; the network itself must self-organize based on the particularities that exist between components elements of samples submitted the same. This self-organizing occurs through the process of cooperation and competition between neurons in the network.

Another way to classify the artificial neural networks is how its features which may be continuous, discrete, deterministic or stochastic and as to its structure that may be multilayered networks (in the case of this work) – multilayer feedforward network, whose following information flow in a single direction or recursive networks – current network, where there is the feedback of the whole network.

6 OpenCV

OpenCV (Opens Source Computer Vision) is a cross-platform library for image processing created by Intel in 2000, and maintained by Willow Garage. It is available for various programming languages, including C, C++, and Python.

The OpenCV library has over 500 functions. His goal was to make the image processing and computer vision more accessible to developers and users in areas such as human-machine interaction in real time, and robotics.

When a program that uses this library is loaded into memory, it automatically invokes a DLL (Dynamic Linked Library) to detect the processor type and loads a specific DLL to the same; This library is optimized for Intel processors.

The OpenCV library is divided into five groups: Image Processing (cv), Graphical Interface (highgui), Data Structure and Algebraic Operations (cxcore), Input and Output Video (cvaux) and Machine Learning (ml).

The version of OpenCV used was 2.2 (december 2010) due to a characteristic function of detecting objects that are not found in previous version where it limits the maximum size in pixels of the object in question.

6.1 Object Detection with OpenCV

The first stage of the system was the detection of microcalcifications in digital mammography images and for this if you used processes present in OpenCV library [Bradski and Kaehler, 2004] which implements the Viola-Jones [Jones and Viola, 2003]. This method is based on *Haar* rectangular type characteristics that are equal to the difference in intensity between subregions. In Figure 4 are displayed the basic characteristics of type *Haar*.

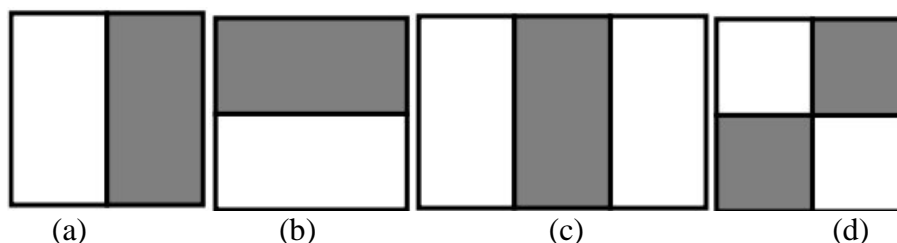


Figure 4: Characteristics of type Haar: (a) vertical Division, (b) horizontal Division, (c) two vertical divisions and (d) horizontal and vertical divisions. Source: [Jones and Viola, 2003].

The algorithm used when you choose the object he applies the characteristics detection in rectangular images finding a set of characteristics that define the desired object. Through the implementation of an algorithm called *AdaBoost* [Jones and Viola, 2003] which is a learning algorithms *boosting* is filtrate the set by decreasing the amount of generated features. With these features generates a classifier for that object.

In developing this classifier are required two types of images: positive and negative. In the affirmative are targets to detect objects that are negative images as on that does not contain the target object. With these images, negative and positive training can be accomplished.

7 Development

Software development is divided into two parts: the first is used to identify the microcalcifications and separate them on images with a green, square 24 x 24 pixels in the

second half was used a neural network to detect potential false positives, marked with yellow square, so that the radiologist to decide whether really is a false positive.

The equipment used for the development of this work was a machine with Intel Core I7, 8 GB RAM with Radeon HD 6770 video card with 2 GB of memory. The operating system used was Windows 7 Home Premium 64-bit. The development language is C/C++ and the IDE (Integrated Development Environment) was the code::blocks 10.05 version with GNU GCC compiler.

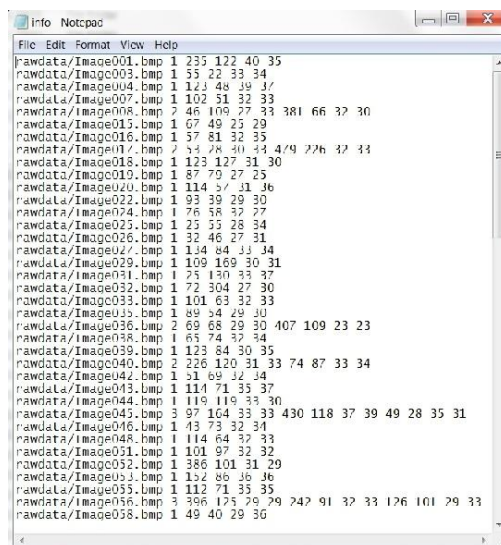
7.1 Using OpenCV in Haar classifier to detect objects

The steps used to generate objects detector and classifier (microcalcifications) were:

- Create the positive samples description file;
- Create the file of description of negative samples;
- Generate a vector with positive samples in a file of type *vec*;
- Train a classifier;
- Convert the generated file to an xml file;
- Use the input xml into the software so that it can detect the objects (microcalcifications).

For this work were used 118 positive images with resolution 24 x 24 pixels and 375 negative images.

The first step was to generate a text file with the locations of positive (microcalcifications) objects within images. For this was developed a small software which selects the picture containing the object to be detected and if it makes a rectangle for this object. The software in question generates a text file in which each line of the file is equal to one of these positive images. The first element of the line refers to the name of the image, the second the amount of positive objects present in this image and the rest are the x and y coordinates match the height and width of each object in the image. In our case was generated a file called info.txt as Figure 5.



```
Info Notepad
File Edit Format View Help
rawdata/Image001.bmp 1 235 122 40 35
rawdata/Image003.bmp 1 55 22 33 34
rawdata/Image004.bmp 1 123 48 39 37
rawdata/Image007.bmp 1 102 51 32 33
rawdata/Image008.bmp 2 46 109 27 33 381 66 37 30
rawdata/Image015.bmp 1 67 49 25 29
rawdata/Image016.bmp 1 57 81 32 35
rawdata/Image017.bmp 2 54 78 40 43 479 276 37 33
rawdata/Image018.bmp 1 123 127 31 30
rawdata/Image019.bmp 1 87 79 27 25
rawdata/Image021.bmp 1 114 57 41 36
rawdata/Image022.bmp 1 93 39 29 30
rawdata/Image024.bmp 1 76 58 37 27
rawdata/Image025.bmp 1 25 55 28 34
rawdata/Image026.bmp 1 32 46 27 31
rawdata/Image027.bmp 1 134 84 33 34
rawdata/Image029.bmp 1 109 169 30 31
rawdata/Image031.bmp 1 25 130 33 37
rawdata/Image032.bmp 1 72 304 27 30
rawdata/Image033.bmp 1 101 63 32 33
rawdata/Image035.bmp 1 89 54 29 30
rawdata/Image036.bmp 2 69 68 29 30 407 109 23 23
rawdata/Image038.bmp 1 65 74 37 34
rawdata/Image039.bmp 1 123 84 30 35
rawdata/Image040.bmp 2 226 120 31 33 74 87 33 34
rawdata/Image042.bmp 1 51 69 32 34
rawdata/Image043.bmp 1 114 71 35 37
rawdata/Image044.bmp 1 119 119 33 30
rawdata/Image045.bmp 3 97 164 33 33 430 118 37 39 49 28 35 31
rawdata/Image046.bmp 1 43 73 32 31
rawdata/Image048.bmp 1 114 64 37 33
rawdata/Image051.bmp 1 101 97 32 32
rawdata/Image052.bmp 1 386 101 31 29
rawdata/Image053.bmp 1 152 85 36 36
rawdata/Image055.bmp 1 112 71 35 35
rawdata/Image056.bmp 3 306 175 29 29 242 91 37 33 126 101 29 33
rawdata/Image058.bmp 1 49 40 29 36
```

Figure 5: File info.txt containing information of positive images. Source: own.

Then, so that the negative images can be read by OpenCV library, generates a text file the in which each row corresponds to one of the negative images. In our case bg.txt as Figure 6.

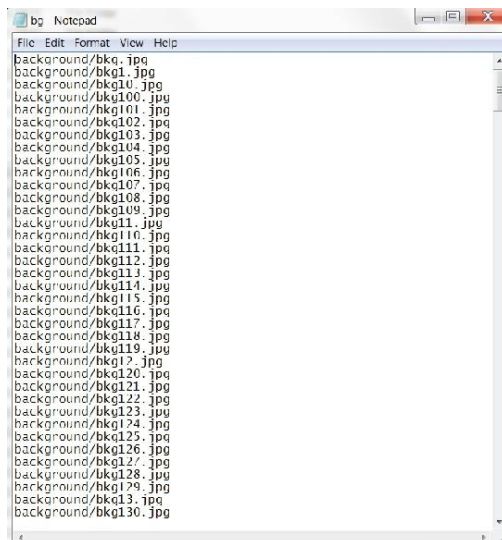


Figure 6: File bg.txt containing information of the negative images. Source: own.

The next step was to generate a vector (*vec*) with positive images. To generate this file *.vec* used an application called OpenCV library in *opencv-createsamples*. For the training was used the application of OpenCV called *haartraining* that provides as *opencv-saida* a training file. This training file is converted to an xml file so that it can be read by the program that will detect the microcalcifications.

After generating this file, *resultado.xml*, it will be passed to our program called "Haar.exe" so that it can detect the microcalcifications level 2 BI-RADS. Below in Figure 7 we display two screens of results of the program.

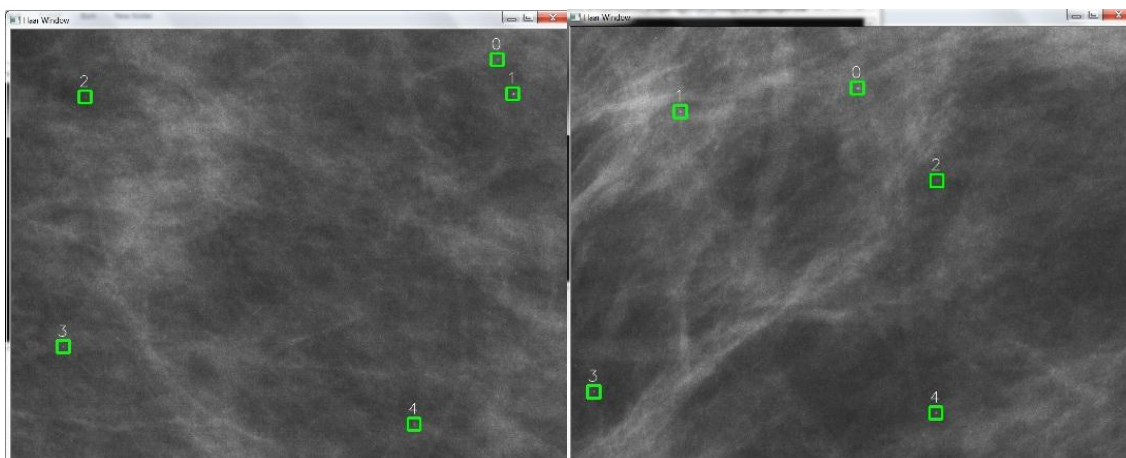


Figure 7: Output of the program Haar.exe. Source: own.

The Haar.exe software program has been tested on digital mammography and the result was:

- Total of microcalcifications are detected by the system (microcalcifications level BI-RADS 2): 94.50%

- Total of microcalcifications are not detected by the system: 5.49%

The system also has detected 10.98% of false positives.

To eliminate these false positives, the system was integrated with a neural network that is the subject of the next topic.

7.2 Using Neural Network to detect potential false positives

With the goal of detecting false positives found by the software "Haar.exe" we use a neural network for this purpose.

But before even using neural networks, the mammography image had to be crafted and how did this was to equalize your histogram like how to get the best image contrast.

According to Gonzales and Wood [Gonzales and Wood, 2010], the histogram of a digital image with intensity levels in the interval $[0, L-1]$ is a discrete function $h(r_k) = n_k$, where r_k is the k -th value intensity and n_k is the number of pixels of the image with intensity r_k .

Contrast is a qualitative measure and that is related to the distribution of gray tones of the image. In Figure 8 (a) we have a mammogram image before Equalization and in (b) the same image equalized.

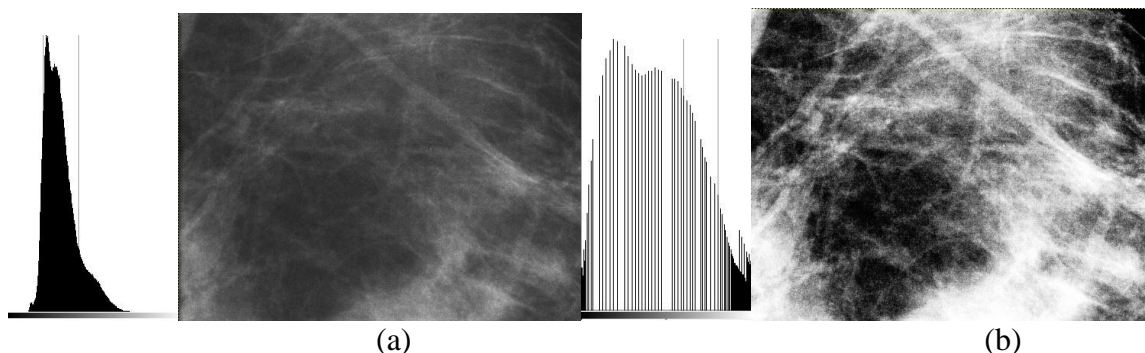


Figure 8: In (a) mammography image not equalized and (b) the same image equalized.
 Source: own.

The neural network used in this work was built has 22 neurons in the middle tier and a single neuron in the output layer. These 22 neurons receive part of the image that matches the selection made OpenCV which is a square of 24x24 pixels.

For neural network training were used 45 sets of input data, with 3000 training times and with a value of learning rate and with a rate of 0.000005 errors 0.000001.

The activation function used in this project was the sigmoid neural network:

$$f(u) = \frac{1}{1 + e^{-u}}$$

After training the neural network, marginal stat weights generated were entered into the final software. In Figure 9 you can see two screens where false positives were detected by the software.

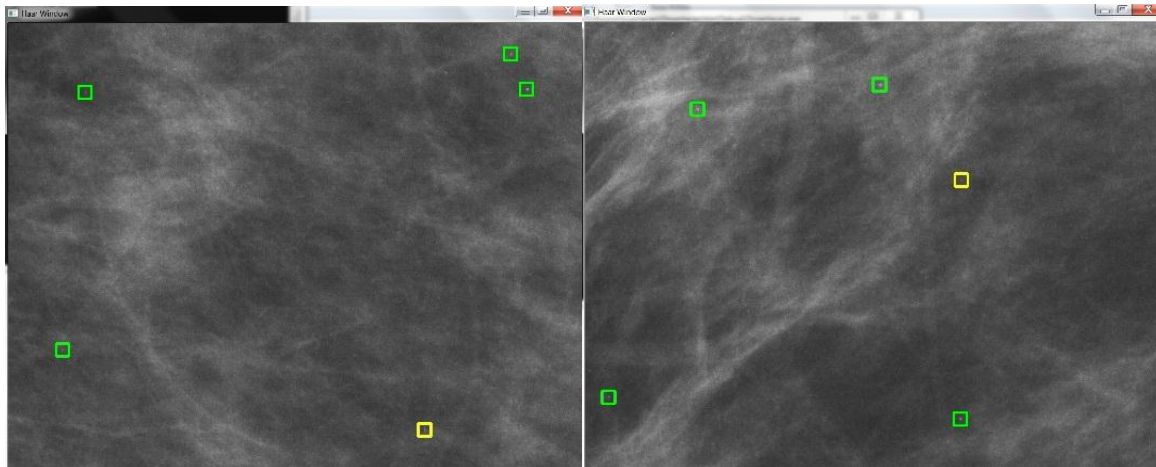


Figure 9: False positives detected by the neural network. Source: own.

After use the neural network software, he was able to detect 70% of false positives found previously, on the other hand, false positives were detected as 5.46% of microcalcifications positive.

8 Conclusion and future works

In this study we used artificial intelligence techniques, Artificial Neural Networks, in conjunction with algorithms provided by the OpenCV library, such as AdaBoost which makes the detection of objects in our case microcalcifications in digital mammography images.

Throughout this paper were made a lot of tests and configurations in the topology of neural networks and also the weights of the functions of the detector objects until they reached the best result possible. The system could detect the objects in question to hit up 94.5%.

The system is designed to detect the level BI-RADS 2 due to the lack of a higher picture with the other levels of BI-RADS.

This software can be useful as an aid to the radiologist detect possible microcalcifications in digital mammography images.

As future work we intend to adapt the system to detect microcalcifications in all levels of BI-RADS and transform the neural network on a neurofuzzy network for a better fit in the images by their histograms equalized.

References

- American College of Radiology,. Breast Imaging Reporting and Data System (BI-RADS). Breast Imaging Reporting and Data System, 1998.
- BEALE, R; JACKSON, T. Neural Computing : an Introduction. Bristol : Adam Hilger, 1990.
- Bradski, G.R. and Kaehler, A. (2004). Learning OpenCV Computer Vision with the OpenCV Library. O`Reilly, 2th edition.
- GONZALES, Rafael C.; WOODS, Richard E.. Processamento Digital de Imagens. 3. ed. São Paulo: Pearson Education do Brasil, 2010. 624 p.
- HAYKIN, Simon. Redes Neurais: Princípios e prática. 2. ed. Porto Alegre: Bookman, 2001. 900 p.

- Hodgkin, A.L., Huxley, A.F. (1952); A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology*, vol. 117, pp. 500-544.
- Jones, M. and Viola, P. (2003). Fast multi-view face detection. In MRL Technical Report TR2003-96, Cambridge.
- LUDWING JUNIOR, Oswaldo; MONTGOMERY, Eduardo. *Redes Neurais: Fundamentos e Aplicações com Programas em C*. Rio de Janeiro: Ciência Moderna Ltda, 2007. 125 p.
- Mavroforakis, M.; Georgiou, H.; Dimitropoulos, N.; Cavouras, D.; Theodoridis, S. (2005) Significance analysis of qualitative mammographic features, using linear classifiers, neural networks and support vector machines. *European Journal of Radiology*, abril, V. 54, ed. 1.
- Ministério da Saúde, Instituto Nacional do Câncer, Brasil. *Mamografia: da prática ao controle*: Universidade do Estado do Rio de Janeiro (UERJ) 2007.
- SILVA, Ivan Nunes da; SPATTI, Danilo Hernane; FLAUZINO, Rogério Andrade. *Redes Neurais Artificiais: curso prático*. São Paulo: Artliber Editora Ltda, 2010. 399 p.
- WASSERMAN, Philip D. *Neural Computing : Theory and Practice*. New York: Van Nostrand Reinhold, 1989.
- [Rosenblatt, 1958] Rosenblatt, F. (1958), The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, vol. 65 pp. 386-408.