CLASSIFICATION OF MAMMOGRAPHIC IMAGES BY *OPENVINO*: A PROPOSAL OF USE TO ENHANCE MORE EFFECTIVITY IN CANCER DIAGNOSIS

Horacio Emidio de Lucca Junior[1,2] and Arnaldo Rodrigues Santos Jr[2]

[1]Centro Educacional da Fundação Salvador Arena, CEFSA,
São Bernardo do Campo, SP, Brazil
[2]Centro de Ciências Naturais e Humanas (CCNH),
Universidade Federal do ABC, São Bernardo do Campo, SP, Brazil

ABSTRACT

Diseases that are characterized by the disordered growth of cells that, in many cases, have the property of invading tissues and organs are commonly called cancer. Such cells divide quickly and the invasion can be very aggressive and uncontrolled, resulting in formation of malignant tumors Mammographic images from libraries of the American digital database DDSM were used in this research for digital improvement and characteristic analysis using the OpenVino computer program This work has as main objective to analyze mammography images of breast nodules and to propose a method of classification by shape and texture using computer programs that can maximize the accuracy in the correct diagnosis regarding the malignancy or not of a tumor. It is a tool that it can be useful as a contribution in the interpretation of the results to mastologists who identify such nodules through the analyzed radiological images.

KEYWORDS

Diagnostic imaging, Image processing, Computer-Aided Detection, Computer-Aided Diagnosis.

1. INTRODUCTION

The aid to early diagnosis of cancer has been an incessant search by researchers and companies that study analysis of computational images. They are dedicated to developing systems for this and perhaps making millions of lives less traumatic (EADIE et al., 2012) [1]. This research aims to contribute to a better understanding of mammographic images by evaluating the Intel *OpenVino* program, verifying its effective assistance in interpreting the results obtained. Studies indicate that, although breast cancer can affect people of all ages, the main risk factor is age because the rate of increase increases rapidly for patients up to 50 years old. After that age, the increase occurs more slowly (MARX, 2003) [2], but other risk factors such as those related to the woman's reproductive life, family history of breast cancer, in addition to the high density of breast tissue, are considered. Another situation that has also been found to be a risk factor is exposure to ionizing radiation, even at low doses, especially during puberty (INCA, 2019) [3].

On February 22, 2018, an interview by Lavínio Nilton Camarim, then president of the Regional Council of Medicine of São Paulo (Cremesp), was published in Exame Magazine (Editora Abril), reporting that in a survey conducted by them, it was found that 88% of newly graduated doctors

David C. Wyld et al. (Eds): MLIOB, SIPO, NET, DNLP, SOEA, AISCA - 2021
Pp.77-84, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.111206

did not know how to interpret mammography results. Diagnostic errors can also lead to false positives, where patients undergo unnecessary treatments. As the contour of the masses in mammographic examinations are not well defined as to the limits of the images, the use of techniques that are not capable of making precise segmentations can be effective. As seen in the works of Hussain et al (2014) [4], Cheikhrouhou, Djemal and Maaref (2011) [5] where the variations of the derivative signals at different points of interest in the contour of the masses were evaluated or in Rocha et al. (2016) [6], which used levels of diversity and patterns of LBP (local binary patterns) and gray level co-occurrence matrices (GLCM) for the extraction of texture characteristics.

Several computational techniques have been used in order to develop tools that assist in the interpretation of mammographic exams. These include identifying structures compatible with tumors, aiming at improving the rate of early detection of breast cancer (GIGER, 2000) [7]. Although we can see that this theme has been going on for some time, systems that help in the detection of threats, CAD (Computer-Aided Detection), and those that help in the diagnosis of diseases, CADx (Computer-Aided Diagnosis) systems, are already present in several diagnostic imaging centers. There occur especially in developed countries, such as the USA and European countries (TAYLOR et al., 2004; FENTON et al., 2007) [8] [9]. Such techniques are improved over time in order to have the greatest possible effectivity, therefore the objective of this work is to study a technique different from those specified here to analyze an accuracy of this method.

2. MATERIALS AND METHODS

In the area of computing, there are many challenges for collaborating with medical diagnostics using images, since the acquisition of these, going through the pre-processing and segmentation phases to, finally, classify them. The use of auxiliary programs for pattern recognition has increased exponentially in recent years, these recognitions are made better by machines than by humans (NOBESHI, 2016) [10], for professor researcher from USP (University of São Paulo), Dr. Alexandre Chiavegatto Filho*, "It was believed that the greatest transformations in medicine would occur with the use of robots in corridors or surgical centers, the great advance, however, are the systems that recognize patterns in illnesses and offer doctors elements that help them in making decision-making."*

After defining the language of use, in the Python case, the next step was the choice according to the program that helps in the interpretation of the images studied. Some computer visions are important for understanding the functioning of the program that contributes to the classification of images. Convolutional Neural Networks (CNN) are used mainly for image classification, while convolutional filters are used to extract characteristics from images. These are applied in several layers, and at the end of the training, the model learns to distinguish the most important characteristics. There is a need for a large volume of images for more effective classification. Therefore, it is important to transfer learning, that is, use the network that has already been trained and put a new layer on it. In this case, it would be training in mastering mammographic images. The program chosen was OpenCV, an image processing library developed by Intel. The choice of this program was mainly due to the fact that this library is available on Mac, Windows and Linux, works in C, C ++ and Python, and would be a free open source and easy to use and install.

2.1. Image Bank

For the analysis of mammographic images to be carried out effectivity, it was necessary to use an image bank that could corroborate the objectives of the work. Using a standard test database is important for researchers to compare results directly. The most common databases for the analysis of mammographic images are the database of the Mammographic Image Analysis Society (MIAS) and the Digital Database for Screening Mammography (DDSM). DDSM's main objective is to provide access to images to facilitate research in the development of computational algorithms that can assist in their screening, as well as assist in the diagnosis and development of didactic or even training material. Approximately 2500 studies are included in this database. After analyzing and studying the image banks available for the research, the use of the DDMS database was determined, mainly due to the quality, diversity of incidence and quantity of images. This project was approved by the UFABC Ethics Committee (Process 08/2020).

2.2. Using the program

The images used were in jpg format. First, the <cv.blur> algorithm was designed to simplify the texture and thus give greater projection to the research object, in this case, the breast nodules.

```
import cv2 as cv
import numpy as np
from matplotlib import pyplot as plt
img = cv.imread('Imagem1.jpg')
blur = cv.blur(img,(10,10))
plt.subplot(121),plt.imshow(img),plt.title('Original')
plt.xticks([]), plt.yticks([])
plt.subplot(122),plt.imshow(blur),plt.title('Blurred')
plt.xticks([]), plt.yticks([])
plt.show()
cv.imwrite('imagem1b.jpg',blur)
```

Figure 1 – commands for *image smoothing*

The following figure shows the original image (Imagem1.jpg) and cv.imwrite (imagem1b.jpg) after highlighting its texture

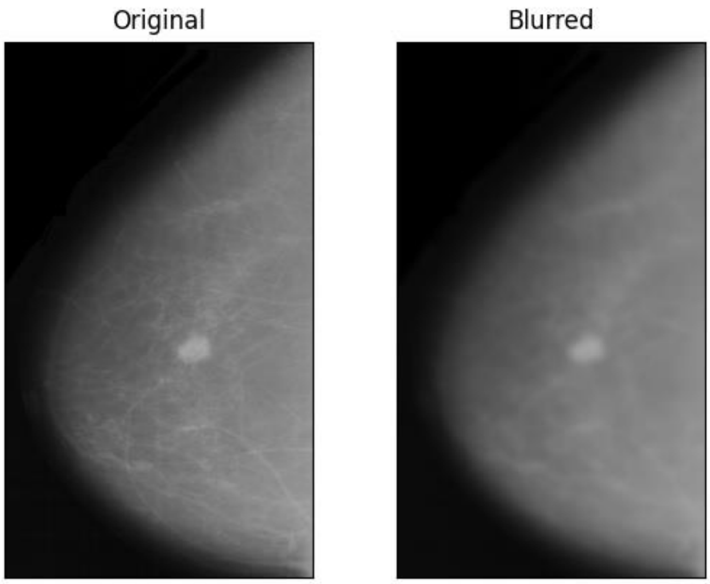

Figure 2 – blurred image obtained from the original image

This process is also called *image smoothing*, it removes high-frequency content normally occurs at the edge contour.

From the *image smoothing* or *blurred image*, the command that determined the segmentations on the edge of the identified tumors was used, observing that the input image is the one obtained by the previous process.

```
import numpy as np
import cv2 as cv
from matplotlib import pyplot as plt
img = cv.imread('imagem1b.jpg',0)
edges = cv.Canny(img,1,3)
plt.subplot(121),plt.imshow(img,cmap = 'gray')
plt.title('Original Image'), plt.xticks([]), plt.yticks([])
plt.subplot(122),plt.imshow(edges,cmap = 'gray')
plt.title('Edge Image'), plt.xticks([]), plt.yticks([])
plt.show()
cv.imwrite('imagem1contorno.jpg',edges)
```

Figure 3 – commands for segmentations on the edge

However, the edges = cv.Canny (img, x, y) parameters must be adjusted, because when x = 1 and y = 3, for example, many contours can be obtained, as shown in figure 4.

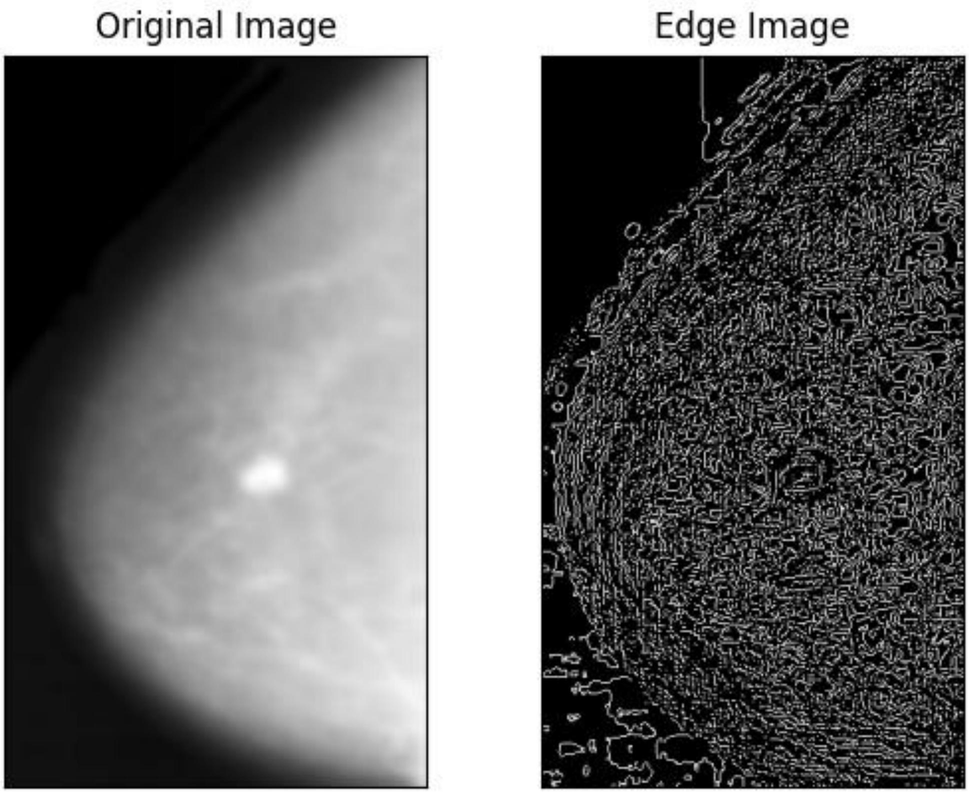

Figure 4 – segmentations on the edge image obtained from the original image (blurred image)

It is the appropriate adjustments of the x and y variants that will determine the proper identification of the nodules. The parameters x = 10 and y = 30 were then used.

```
import numpy as np
import cv2 as cv
from matplotlib import pyplot as plt
img = cv.imread('imagem1b.jpg',0)
edges = cv.Canny(img,10,30)
plt.subplot(121),plt.imshow(img,cmap = 'gray')
plt.title('Original Image'), plt.xticks([]), plt.yticks([])
plt.subplot(122),plt.imshow(edges,cmap = 'gray')
plt.title('Edge Image'), plt.xticks([]), plt.yticks([])
plt.show()
cv.imwrite('imagem1contorno.jpg',edges)
```

Figure 5 – commands for segmentations on the edge with x = 10 and y = 30

Obtaining the appropriate image, as shown in the figure 6, where it was possible to identify the nodule.

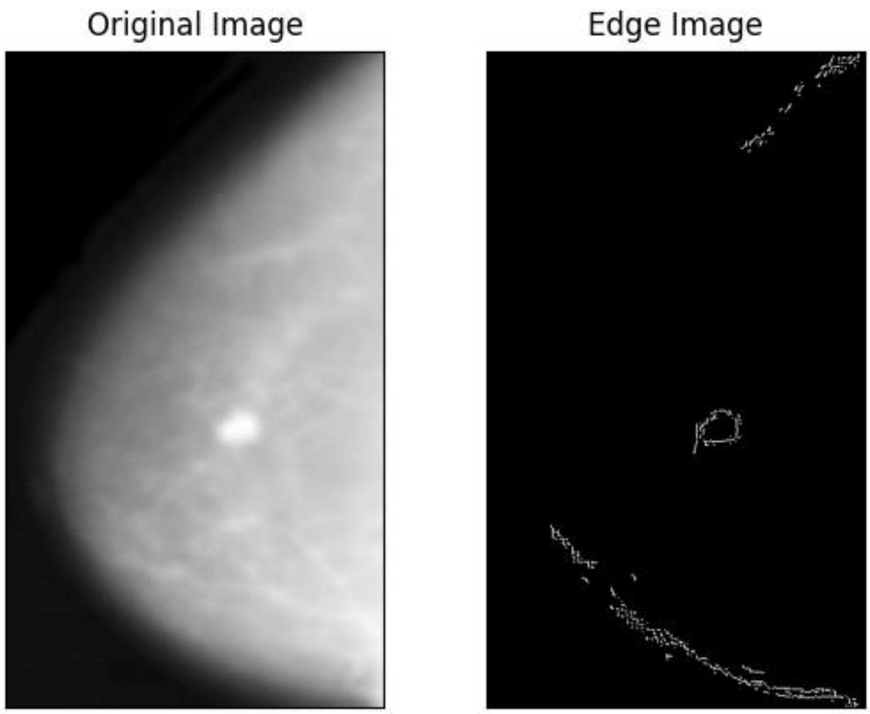

Figure 6 – segmentations on the edge image obtained from the original image (blurred image) with x = 10 and y = 30

After making several variations for x and y, it was concluded that the best results were obtained when 9 < x < 12 and 28 < y < 31.

2.3. Results

The mammographic images used were divided by incidence and side of the breast. The parameters x = 10 and y = 30 were then used for the correct identification of the nodule. For the craniocaudal incidence (CC) on the right side, 237 images were used, obtaining an 87.8% effective in the identification and classification of malignant nodules, since they were observed in 208 of the analyzed images.

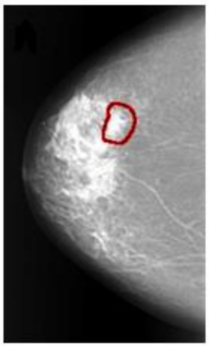

Figure 7 - nodule identified in the skull - caudal view on the right side – (DDMS)

For the left craniocaudal incidence (CC), 243 images were used with 86.8% effective (211 images) in the identification and classification of malignant nodules. From orthogonal views, there was also a classification for mediolateral-oblique view on the right side, with 186 images and 84.4% effective, observed in 157 images.

Figure 8 - nodule identified in right lateral mediolateral-oblique image (DDMS)

Finally, 213 mediolateral-oblique images on the left side were also used with 86.9% effective in the identification and classification of malignant nodules, that is, in 185 images.

The same images were analyzed using the parameters x = 11 and y = 29, but the correct results obtained percentages below with these parameters. For the craniocaudal incidence (CC) on the right side, 237 images were used, obtaining an 85.7% effective in the identification and classification of malignant nodules, since they were observed in 203 of the analyzed images. For the left craniocaudal incidence (CC), 243 images were used with 84.4% effective (205 images) in the identification and classification of malignant nodules. From orthogonal views, there was also a classification for mediolateral-oblique view on the right side, with 186 images and 83.3% effective, observed in 155 images, and 213 mediolateral-oblique images on the left side were also used with 85.4% effective in the identification and classification of malignant nodules, that is, in 182 images.

2.4. Discussion

Breast cancer is one of the most aggressive tumors and brings the highest incidence of death among women. Screening by means of mammograms is the main means of early detection for the diagnosis of malignant neoplasms of the breast, one of the main causes of death in different countries (XAVIER et al., 2016; CHOI et al., 2018; BOUJEMAA et al., 2019) [11] [12] [13].

In relation to the radiological technique used, the contrast between the tissue to be identified and the background is a major factor for the perception of breast lesions still in the early stages. The ratio, low electrical voltage or difference in electrical potential (kilovoltage - Kv) and high intensity of electrical current (milliamperage - mA) define the high contrast necessary to obtain the image of the breast. With the program used, it was possible to analyze several characteristics of the images. Additionally, some factors impacted the correct interpretation, such as heterogeneously dense breast tissue, which can hinder the detection of small nodules.

However, it was possible to observe isodense, oval, partially obscured, bilateral nodules, smaller than 1.0 cm, as well as identifications of suspicious microcalcifications, even thin and pleomorphic, with evolutionary forms. With greater ease, the program detected high density spiculated nodules, with diameters greater than 1.0 cm, characteristic of malignant neoplasia. The results obtained were satisfactory in relation to the effectiveness of the method. However, it is worth mentioning that the number of images analyzed was low for a better performance of the program so that they were properly classified in convolutional neural networks. Another factor that implies a more promising result is the fact that the images worked in the DDMS database are images that are more than 15 years old. Therefore, these images are not as sharp as the most recent ones coming from more modern mammographs.

3. CONCLUSIONS

With this work it was possible to conclude that the image training method used with the *OpenVino* program obtained promising results. An average of 86.6% effective was achieved in detecting malignant nodules from mammographic images. It is important to highlight that, in Brazil, a considerable part of the population does not have access to adequate medicine for the treatment, nor for the early diagnosis of the disease. Therefore, using a method that helps the correct interpretation of mammographic exams by doctors and radiologists, contributes more accurately to the correct diagnosis of breast cancer. The effectivity values and the success rate of the tumors are considered relatively good, however the use of more modern images and, consequently, with higher resolution, would give greater precision in the correct diagnosis.

REFERENCES

[1] EADIE, L. H.; TAYLOR, P.; GIBSON, A.P. A systematic review of computer-assisted diagnosis in diagnostic cancer imaging. European Journal of Radiology, v. 81, n.1), e70-e76, 2012.

[2] MARX, A. G.; FIGUEIRA, P. V. G. Fisioterapia no câncer de mama. Barueri, SP: Atlas, 2003.

[3] https://www.inca.gov.br/

[4] HUSSAIN, M.; KHAN, S.; MUHAMMAD, G.; AHMED, I.; BEBIS, G. Effective extraction of gabor features for false positive reduction and mass classification in mammography. Appl. Math, v. 8, n. 1L, p. 397-412, 2014.

[5] CHEIKHROUHOU, I.; DJEMAL, K.; MAAREF, H. Protuberance selection descriptor for breast cancer diagnosis. In: IEEE. 3rd European Workshop on Visual Information Processing (EUVIP) 2011. Paris, France, 2011. P. 280-285.

[6] ROCHA, S. V. da; JUNIOR, G. B.; SILVA, A. C.; PAIVA, A. C. de; GATTASS, M. Texture analysis of masses malignant in mammograms images using a combined approach of diversity index and local binary patterns distribution. Expert Systems with Applications, Elsevier, v. 66, p. 7-19, 2016.

[7] GIGER, M. L. Computer-aided diagnosis of breast lesions in medical images. Computing in Science & Engineering, AIP Publishing, v. 2, n. 5, p. 39-45, 2000.

[8] TAYLOR, P.; CHAMPNESS, J.; GIVEN-WILSON, R.; POTTS, H.; JOHNSTON, K. An evaluation of the impact of computer-based prompts on screen readers' interpretation of mammograms. The British journal of radiology, v. 77, n. 913, p. 21, 2004.

[9] FENTON, J. J.; TAPLIN, S. H.; CARNEY, P. A.; ABRAHAM, L.; SICKLES, E. A.; D'ORSI, C.; BERNS, E. A.; CUTTER, G.; HENDRICK, R. E.; BARLOW, W. E. et al. Influence of computer-aided detection on performance of screening mammography- New England Journal of Medicine, Mass Medical Soc, v. 356, n. 14, p. 1399-1409, 2007.

[10] NOBESCHI, A. Saúde: como a inteligência artificial pode ajudar nos diagnósticos. Available in: https://epoca.globo.com/saude/noticia/2016/12/saude-como-inteligencia-artificial-pode-ajudar-nos-diagnosticos.html. Access in 12/22/2019

[11] XAVIER, D.R.; OLIVEIRA, R.A.D.; MATOS, V.P.; VIACAVA, F.; CARVALHO, C.C. Cobertura de mamografias, alocação e uso de equipamentos nas Regiões de Saúde. Saúde Debate, v.40, n.110, 2016.

[12] CHOI, K.S.; YOON, M.; SONG, S.H.; SUH, M.; PARK, B.; JUNG, K.W.; JUN, J.K. Effect of mammography screening on stage at breast cancer diagnosis: results from the Korea National Cancer Screening Program. Scientific Reports, v.8, 8882, 2018.

[13] BOUJEMAA, S.; BOSMANS, H.; BENTAYEB, F. Mammography Dose Survey Using International Quality Standards. Journal of Medical Imaging and Radiation Sciences, v.50, n.4, p.529-535, 2019.

AUTHORS

Horacio Emidio de Lucca Junior – graduated in Mathematics (2001), graduated in Pedagogy (2015), postgraduate – 360h in Mathematics Education (2004) and Master in Applied Mathematics from Federal University of ABC (2016). Has experience in Mathematics, with emphasis on Analytical Geometry, Calculus, Financial Mathematics, Statistics and Probability. PhD student in Biotechnology at the Federal University of ABC (conclusion 2022)

Arnaldo Rodrigues Santos Junior - graduated in Biological Sciences (1993), Master's degree in Cell Biology (1996) and a PhD in Cell and Structural Biology (2001). Has experience in the area of Morphology, with emphasis in Cytology and Cell Biology. It works on the following topics: cell culture, cell growth and differentiation, biomaterials, morphological analysis, bioresorbable polymers and Biotechnology.